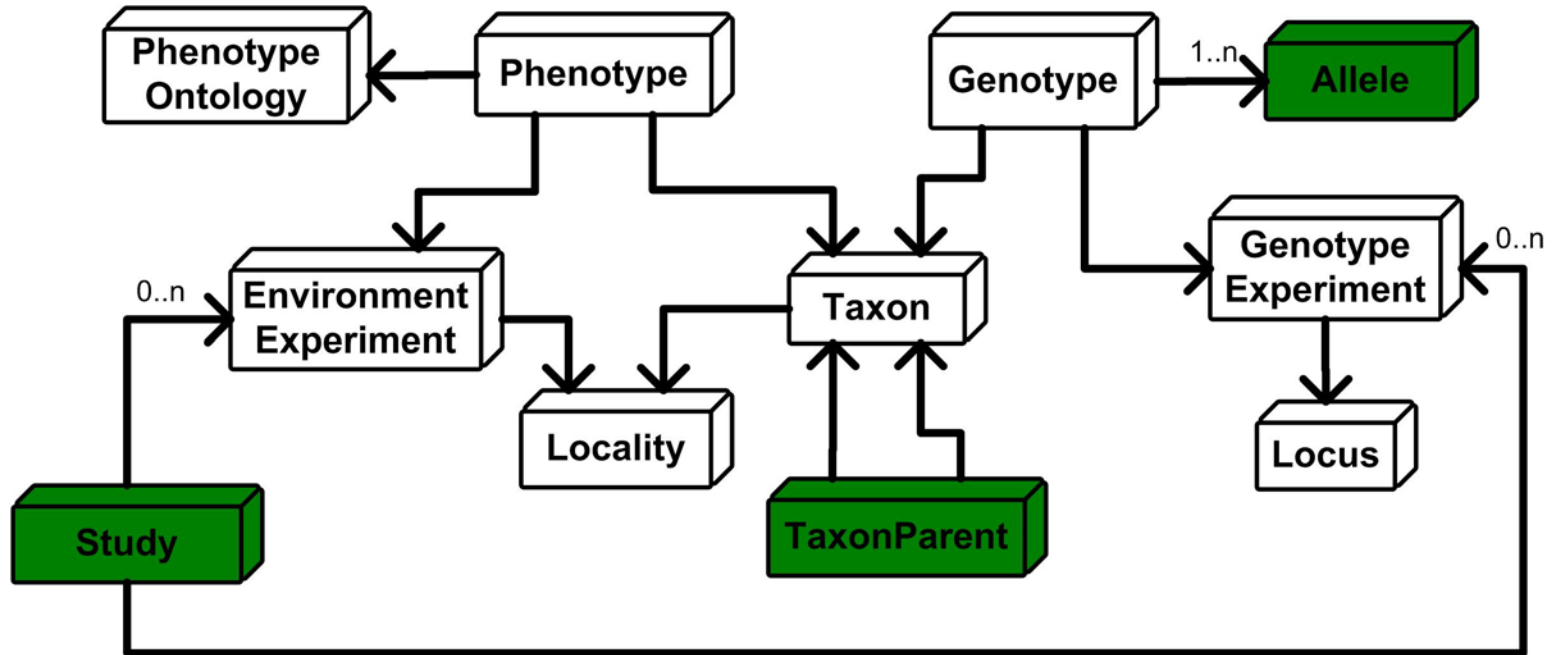


The Genomic Diversity and Phenotype Connection (GDPC)

Entities and their Relationships

Version 2.0 – draft March 7, 2007



Legend:

Item	New Items (version 2.0)
Item	Deleted Items
Item	Questions?
Item	Existing Items (version 1.1)

**Entity/Property descriptions for
The Genomic Diversity and Phenotype Connection (GDPC)
Version 2.0 - Draft March 7, 2007**

Entity/Property	Data Type	Ver.	Req'd	Description
Taxon	gov.usda.gdpc.Taxon			Entity representing a taxon. Can be defined with a different levels of granularity depending on instance being defined. The hierarchy follows starting with the highest level to the most specific. 1) accession, 2) source, 3) seed lot, 4) group of plants or individual plant. Many of the properties for Taxon come from the FAO/IPGRI Multi-Crop Passport Descriptors as indicated by the abbreviations (in parentheses) following some descriptions.
Accession	java.lang.String	1.1	no yes	Either a registered or other formal designation given to the accession. First letter uppercase. Multiple names separated with semicolon without space. i.e.: Rheinische Vorgebirgstrauben;Emma;Avlon (accename)
Accession Number	java.lang.String	2.0	no	Number serves as a unique identifier for accessions within a genebank collection. (accenumb)
Collector	java.lang.String	1.1	no	Collector of the accession Is this a person or institution?
Comments	java.lang.String	1.1	no	Comments
Data Source	java.lang.String	1.1	yes	Identifies origin of data
Genus	java.lang.String	1.1	no	Genus name for taxon. Initial uppercase letter required. (genus)
Germplasm Type	java.lang.String	1.1	no	Deprecated: Use Biological Status (sampstat) instead. Germplasm type of accession (i.e. Tripsacum, inbred)
ID	gov.usda.gdpc.Identifier	1.1	yes	Along with Data Source, this uniquely identifies instance of this class.
Locality	gov.usda.gdpc.Locality	1.1	no	Locality of this Taxon
Name	java.lang.String	1.1	yes N/A	This property can not be set directly. It is defined: Name = <Accession>:<Source>:<Seed Lot>:<Plant>.
Plant	java.lang.String	1.1	no	Arbitrary plant number (i.e. 1, 2, 3, pool).
Population	java.lang.String	1.1	no	Name of the population
Reference	java.lang.String	1.1	no	Publication in which this Taxon was first released.
Seed Lot	java.lang.String	1.1	no	An alphanumeric id that can be used by the source to access a specific stock and information about it. The combination of source and seedlot will be unique for a stock.
Source	java.lang.String	1.1	no	The individual or institution that PROVIDED the stock directly used to create a specific obs_unit [the person may no longer maintain or have it, i.e., if the entire stock was used in the experiment]
Species	java.lang.String	1.1	no	Specific epithet portion of the scientific name in lowercase letters. Following abbreviation is allowed: 'sp.' (species)

Entity/Property	Data Type	Ver.	Req'd	Description
Subspecies	java.lang.String	1.1	no	Subspecies
Institution Code	java.lang.String	2.0	no	Institution code for institution maintaining the accession. (instcode)
Collecting Number	java.lang.String	2.0	no	Collecting number, original number assigned by collector (collnumb)
Collecting Date	gov.usda.gdpc.util.Calendar	2.0	no	Collecting date. (colldate)
Species Authority	java.lang.String	2.0	no	Species authority (spauthor)
Subtaxa Authority	java.lang.String	2.0	no	Subtaxa (subspecies) authority (subtauthor)
Common Name	java.lang.String	2.0	no	Common crop name (cropname)
Acquisition Date	gov.usda.gdpc.util.Calendar	2.0	no	Acquisition date, date when accession entered collection (instcode above) (acqdate)
Breeding Code	java.lang.String	2.0	no	Breeding institution code (bredcode)
Remarks	java.lang.String	2.0	no	The remarks field is used to add notes or to elaborate on descriptors with value 99 or 999 (=Other). Prefix remarks with the field name they refer to and a colon (e.g. COLLSRC:roadside). Separate remarks referring to different fields are separated by semicolons without space. (remarks)
Biological Status	java.lang.Integer	2.0	no	Biological status of accession: The coding scheme proposed can be used at 3 different levels of detail: either by using the general codes such as 100, 200, 300, 400 or by using the more specific codes such as 110, 120 etc. 100) Wild 110) Natural 120) Semi-natural/wild 200) Weedy 300) Traditional cultivar/landrace 400) Breeding/research material 410) Breeder's line 411) Synthetic population 412) Hybrid 413) Founder stock/base population 414) Inbred line (parent of hybrid cultivar) 415) Segregating population 420) Mutant/genetic stock 500) Advanced/improved cultivar 999) Other (Elaborate in REMARKS field). Follows coding scheme in MCPD (sampstat).
Collecting Source	java.lang.Integer	2.0	no	Collecting/acquisition source, follows coding scheme in MCPD (collsrc)
Donor Code	java.lang.String	2.0	no	Donor institute code (donorcode)
Donor Number	java.lang.String	2.0	no	Number assigned to accession by donor (donornumb)
Safety Duplicates	java.lang.String	2.0	no	Location of safety duplicates (duplsite)
Storage	java.lang.String	2.0	no	Type of germplasm storage, follows coding scheme in MCPD (storage)
Selfing Number	java.lang.Integer	2.0	no	The number of generations of self pollination separating the parental cross from the stock.
Sibbing Number	java.lang.Integer	2.0	no	The number of generations of full-sib crossing separating the parental cross from the stock.
Taxon Parent	gov.usda.gdpc.TaxonParent			Entity representing one taxon parent.

Entity/Property	Data Type	Ver.	Req'd	Description
Parent	gov.usda.gdpc.Taxon	2.0	yes	Parent taxon
Child	gov.usda.gdpc.Taxon	2.0	yes	Child taxon
Recurrent	java.lang.Integer	2.0	yes	Integer value indicating number of times this parent repeatedly crossed. Default value is 0. Example: A = ((B x C) x B) x B. Parent B would have recurrent value of 2. Parent C would have recurrent value 0.
Role	java.lang.String	2.0	yes	Value indicating role of this parent (i.e. self, male, female).

Locus	gov.usda.gdpc.Locus			Entity representing a locus.
Chromosome Number	java.lang.Integer	1.1	no	Chromosome number where this locus resides.
Chromosome Name	java.lang.String	2.0	no	Chromosome name where this locus resides.
Comments	java.lang.String	1.1	no	Comments
Data Source	java.lang.String	1.1	yes	Identifies origin of data
Genetic Bin	java.lang.String	1.1	no	Position of the locus within a bin
Genetic Map	java.lang.String	1.1	no	Name of genetic map (map set)
Genetic Position	java.lang.Double java.math.BigDecimal	1.1	no	Position in cM along a chromosome
ID	gov.usda.gdpc.Identifier	1.1	yes	Along with Data Source, this uniquely identifies instance of this class.
Locus Type	java.lang.String	1.1	no	Type of this locus (i.e. Gene, Cytological).
Name	java.lang.String	1.1	yes	Name of this locus.
Physical Position	java.lang.Double java.math.BigDecimal	1.1	no	Physical position of this locus.

Genotype Experiment	gov.usda.gdpc.GenotypeExperiment			Entity representing an experiment used to collect Genotypic data. Since these GenotypeExperiments are so varied, we will continually add properties to accommodate. This will standardize the ones we have. The infrastructure allows undefined properties which could later be added to the standard list.
Align Program	java.lang.String	1.1	no	
Comments	java.lang.String	1.1	no	Comments
Data Source	java.lang.String	1.1	yes	Identifies origin of data
Date	gov.usda.gdpc.util.Calendar	1.1	no	date of this allele assay.
End Position	java.lang.String	1.1	no	
Human Validation	java.lang.String	1.1	no	
ID	gov.usda.gdpc.Identifier	1.1	yes	Along with Data Source, this uniquely identifies instance of this class.
Locus	gov.usda.gdpc.Locus	1.1	yes no	Locus associated with this experiment.
Name	java.lang.String	1.1	yes	Name of this experiment.
Polymorphism Type	java.lang.String	1.1	no yes	Polymorphism type (SNP, LENGTH, SEQALIGNMENT, SEQUENCE, RAWDATA, MOBILITY).
Primer 1	java.lang.String	1.1	no	Replaced with Primer List
Primer 2	java.lang.String	1.1	no	Replaced with Primer List

Entity/Property	Data Type	Ver.	Req'd	Description
Primer List	gov.usda.gdpc.PrimerList	2.0	no	List of primers used.
Producer	java.lang.String	1.1	no	Example: John Doe, NCSU
Source Experiment	gov.usda.gdpc.GenotypeExperiment	2.0	no	
Start Position	java.lang.String	1.1	no	
Reference Sequence	java.lang.String	2.0	no	Reference sequence used. Required or not???
Reference Taxon	gov.usda.gdpc.Taxon	2.0	no	Points to Taxon associated with the reference sequence.

Environment Experiment	gov.usda.gdpc.EnvironmentExperiment			Entity representing an experiment used to collect Phenotypic data.
Data Source	java.lang.String	1.1	yes	Identifies origin of data
Evaluation Site	java.lang.String	1.1	no	Is this same as Locality???
Harvest Date	gov.usda.gdpc.util.Calendar	1.1	no	Date crop harvested.
ID	gov.usda.gdpc.Identifier	1.1	yes	Along with Data Source, this uniquely identifies instance of this class.
Locality	gov.usda.gdpc.Locality	1.1	no	Locality object containing information about where this experiment was conducted.
Name	java.lang.String	1.1	yes	Name of this experiment.
Plant Date	gov.usda.gdpc.util.Calendar	1.1	no	Date crop planted.
Summary	java.lang.String	1.1	no	Summary of this experiment.
Coordinate X	java.math.BigDecimal	2.0	no	x coordinate
Coordinate Y	java.math.BigDecimal	2.0	no	y coordinate
Repetition	java.lang.String	2.0	no	Group of blocks
Block	java.lang.String	2.0	no	Group of plots
Plot	java.lang.String	2.0	no	Smallest measure observed. (i.e. row in field)

Study	gov.usda.gdpc.Study			Entity representing an study. This is a group of Environment Experiments and/or Genotype Experiments.
Name	java.lang.String	2.0	yes	Name of this study
Design	java.lang.String	2.0	no	Design of this study
Originator	java.lang.String	2.0	no	Originator of this study
Comment	java.lang.String	2.0	no	Comments
Environment Experiment List	gov.usda.gdpc.EnvironmentExperimentGroup	2.0	no	List of Environment Experiments in this Study.
Genotype Experiment List	gov.usda.gdpc.GenotypeExperimentGroup	2.0	no	List of Genotype Experiments in this Study.

Locality	gov.usda.gdpc.Locality			Entity representing a locality.
Altitude	java.lang.Double java.math.BigDecimal	1.1	no	Altitude of this locality expressed in meters above sea level. Negative values are allowed.
City	java.lang.String	1.1	no	City of this locality
Country	java.lang.String	1.1	no	Country of this locality
Data Source	java.lang.String	1.1	yes	Identifies origin of data
ID	gov.usda.gdpc.Identifier	1.1	yes	Along with Data Source, this uniquely identifies instance of this class.

Entity/Property	Data Type	Ver.	Req'd	Description
Latitude	java.lang.Double java.math.BigDecimal	1.1	no	Latitude should be recorded as a decimal number. North latitude should be positive values and South latitude should be negative values.
Longitude	java.lang.Double java.math.BigDecimal	1.1	no	Longitude should be recorded as a decimal number. West longitude should be positive values and East longitude should be negative values.
Name	java.lang.String	1.1	yes	Name of this locality
State/Province	java.lang.String	1.1	no	State or province of this locality.

Genotype	gov.usda.gdpc.Genotype			Entity representing set of alleles for a given Genotype Experiment and Taxon.
Allele List	gov.usda.gdpc.AlleleList (Allele)(String)	1.1	yes	List of alleles for associated Genotype Experiment and Taxon.
Data Source	java.lang.String	1.1	yes	Removed. Can derive this from its other properties.
Genotype Experiment	gov.usda.gdpc.GenotypeExperiment	1.1	yes	Associated Genotype Experiment.
Taxon	gov.usda.gdpc.Taxon	1.1	yes	Associated Taxon.

Allele	gov.usda.gdpc.Allele			Entity representing one allele value.
Num	java.lang.Integer	2.0	yes	number used to uniquely identify this allele from other alleles with the same genotype experiment and taxon.
Quality	java.lang.String	2.0	no	Quality scores of this allele.
Value	java.lang.String	2.0	yes	Value of this allele.
Data Source	java.lang.String	2.0	no	Removed. Can derive this from its other properties.
ID	gov.usda.gdpc.Identifier	2.0	no	Removed. This is not needed.

Phenotype	gov.usda.gdpc.Phenotype			Entity representing one trait value.
Data Source	java.lang.String	1.1	yes	Removed. Can derive this from its other properties.
Environment Experiment	gov.usda.gdpc.EnvironmentExperiment	1.1	yes	Environment experiment where this trait measured.
Ontology	gov.usda.gdpc.PhenotypeOntology	1.1	yes	Ontology describing trait being measured.
Taxon	gov.usda.gdpc.Taxon	1.1	yes	Taxon of plant being measured.
Value	java.lang.String	1.1	yes	Value of this trait.

Phenotype Ontology	gov.usda.gdpc.PhenotypeOntology			Entity representing one trait ontology. This is still a work in progress, as it will eventually represent a true ontology.
Data Source	java.lang.String	1.1	yes	Identifies origin of data
ID	gov.usda.gdpc.Identifier	1.1	yes	Along with Data Source, this uniquely identifies instance of this class.
Name	java.lang.String	1.1	yes	Name of this trait.
Statistic Type	java.lang.String	2.0	no	Statistic type used to measure this trait. (i.e. measure, mean, standard deviation, variance, mode, median, count)
Unit of Measure	java.lang.String	2.0	no	Unit of measure
Ontology	gov.usda.gdpc.Ontology	2.0+	no	This is for a formal ontology. Maybe we should include this in the future?