Ontology of plant genealogy

What does this set of words mean?

OG – ontology for plant genealogy

- It **IS** a description of the breeding history of ‘germplasm’
- It **IS NOT** a description of ‘germplasm’ phenotype
- It **IS NOT** a description of ‘germplasm” genotype

OG – a definition

- A formalised logical description of chapter 2 “**THE GENEALOGY MANAGEMENT SYSTEM”** of the ICIS manual in such a way that both biologists (plant breeders, plant geneticists) and database developers (bio-informaticists and programmers) can understand each other.

Annex 2.2.b: Generative, derivative and maintenance methods for GMS.

METHN	MTYPE	MGRP	MCODE	MNAME	MDESC	MREF	MPRGN	MFPRG	MATTR	GENEQ	MUID	LMID	MDATE
Methods for storing historical pedigrees with incomplete information													
1	GEN	S	UGM	UNKNOWN GENERATIVE METHOD	Unknown generative method for storing historic pedigrees for self pollinating species		0	0	0	0	1	2	19980610
2	GEN	S	UCM	UNKNOWN CONVENTIONAL METHOD OF CF	Unknown conventional method for storing historic pedigrees for cross fertilising species		0	0	0	0	1	0	19980610
3	GEN												
4	GEN												
5	GEN												
6	GEN												
7	GEN												
8	GEN												
9	GEN												
31	DER												
32	DER												
33	DER												
40	GEN												
60	MAN												
61	MAN												
62	MAN												

Annex 2.2.a: Cultivars, Populations, Breeding strategies, Generative, Derivative and Maintenance Methods for Self-fertilised, Cross-fertilised and Vegetatively Propagated crop species.

Breeding System	Cultivars and Populations	Breeding Strategies	Generative Methods	Derivative methods	Maintenance methods
Out Breeding Cross breeding Cross fertilising	Cultivars Open Pollinated Variety Synthetic CV Hybrid CV Obsolete CV Populations Landrace population Landrace cultivar Collection population Collection wild spp population Collection weedy spp population Genetic Stock Mechanical Mixture	Half Mass Selection Full Mass Selection Recurrent Half-sib Selection Recurrent Full-sib Selection Recurrent selection among Selfed Families Population Back Cross	Single cross cf Three-way cross cf Double cross cf Full diallel cross Full diallel cross bulked Half diallel cross Half diallel cross bulked Partial diallel cross Partial diallel cross bulked Interspecific cross cf Selected pollen cross cf Random pollen cross cf Narrow based tester, line cf Narrow based tester, pop cf Broad based tester, line cf Broad based tester, pop cf Polycross cf Random Mating cf Open pollination cf Random Half Sib Fam Selected Half Sib Fam Natural mutation cf Induced mutation population Somaclone cf Auto-polyploid cf Allo-polyploid cf Haploid cf Transgenic event cf	Induced mutation pop Double haploid pop Rouging cf Single plant selection cf Single seed descent cf Restorer selection Full mass selection Half mass selection Landrace population cf Landrace cultivar cf Collection population cf Collection wild spp population cf Collection weedy spp population cf 209(S), 506(O) Restorer Selection 210(S) Maintainer Selection	Seed increase open pollination Seed increase full diallel bulked Seed increase half diallel bulked Seed increase partial diallel bulked Seed increase sunset cross Hybrid formation cf Synthetic formation cf Breeders Seed cf Foundation seed cf Certified seed Cultivar release Import Seed
	Female Source	Same Single Plant	40(G) Selfing	Derivative Methods	Descent
	Single Plant	Plus all the methods describe below for selection to produce fixed lines.	60(G) Plant Identification 204(O) Single Plant Selection 208(S), 505(O) Single Seed Descent		
		Different Single Plant Pedigree Modified Bulk Selection Bulk Selection Single Seed Descent Double Haploid Back Cross Re-selection	101(S) Complex Top Cross 102(S),402(O) Three-way Cross 103(I),403(O) Double Cross 404(O) Full diallel cross 405(O) Full diallel cross bulked 406(O) Half diallel cross 407(O) Half diallel cross bulked 408(O) Partial diallel cross 409(O) Partial diallel cross bulked 424(O) Convergent Cross 425(O) Partial Plus Diallel Cross 107(S) Backcross 108(S) Forces Recessive 109(S) Interspecific Cross 416(O) Narrow based tester, line CF		
		Selected Bulk	104(S) Female Complex Top Cross 110(S) Selected Pollen Cross SF 416(O) Narrow based tester, line CF 418(O) Broad based tester, Line CF 417(O) Narrow Based Tester, POP CF		
		Bulk	104(S) Female Complex Top Cross 111(S) Random Pollen Cross SF 426(O) Random Half Sib Families 427(O) Selected Half Sib Families		
	Selected Bulk	Single Plant	105(S) Male Complex Top Cross 108(S) Forces Recessive 416(O) Narrow based tester line CF		
		Same Selected Bulk	420(O) Polycross CF 421(O) Random Mating CF	203(S) Purification 204(S), 503(O) Rouging 206(S) Selected Bulk 508(O) Full Mass Selection	
		Different Selected Bulk	106(S) Complex Cross 112(S) Open Pollenated SF 404(O) Full diallel cross 405(O) Full diallel cross bulked 406(O) Half diallel cross 407(O) Half diallel cross bulked		

Table 2.2.b: Schematic outline of a general model for crossing practices in plants

The bases for the classification contained in this model is that the source of the female gametes is determined by the source of the seed, which can be a single seed or seed bulked from a single plant, a selected set of plants or a random set of plants. The source of the pollen needs to be determined but can be from the same plant, another plant, a selected set of plants or a random set of plants. The genetic consequences of these crosses or selections depends on the reproductive behaviour of the species.

Table 2.2.1.a: Germplasm Information stored in GMS.

GERMPLASM_ID	Germplasm Identification Code. This is unique for any germplasm recorded. The general rule is: if you wouldn't mix the packets of seed that exist or did exist you give the GEN an entry in GMS.
METHOD NO	This is the method of genesis of germplasm and details are stored in the method table (**Annex 2.2.b**) The method indicates whether the germplasm was produced by a generative, derivative or preservative process, and then how the germplasm was produced. Details of the method which change with each application are stored as attribute information (**Section 2.2.4**)
GROUP_ID	Identification code for last instance of germplasm produced by a generative process from which the present instance was derived. GROUP_ID = GERMPLASM_ID if the current instance is produced by a generative process.
SOURCE_ID	Identification code for the immediate source of a derived or preserved germplasm. (Not valid for generated germplasm.)
GERMPLASM_USER_ID	Identification code of the user who submitted data on the genesis of the germplasm.
LOCATION_NO	Identifier of breeding or origin location.
GERMPLASM_DATE	Date of generation, derivation or preservation of the germplasm.
GERMPLASM_REFERENCE	Identifier of bibliographic reference from where the germplasm data was retrieved.

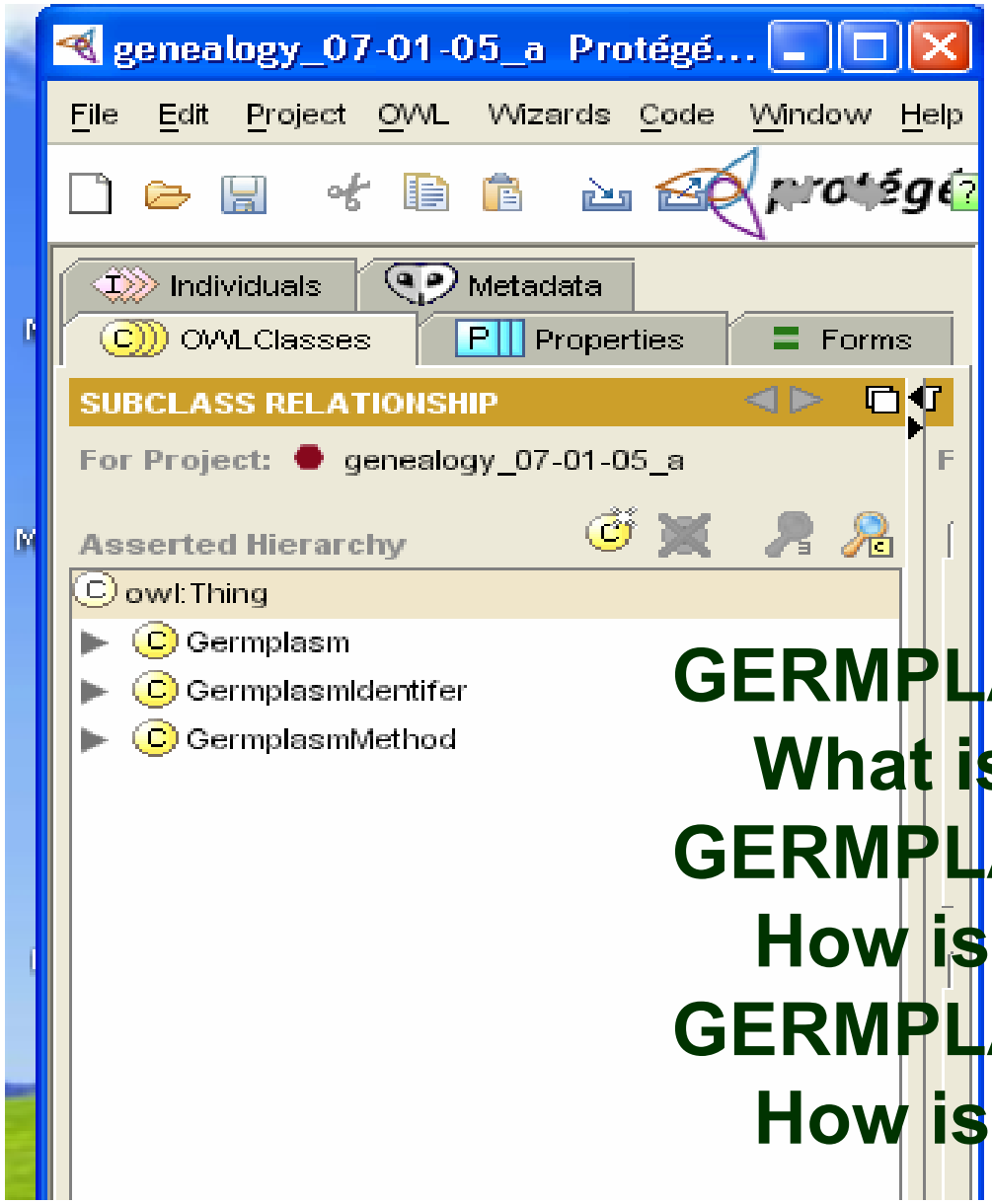
Table 2.2.a: Reproductive Systems in Crop Plants

Self Pollinated	Inbreeding, self pollinated or self fertilised spec Examples: Wheat, Rice, Barley
Cross Pollinated[2]	Examples: Cotton, Pigeonpea, Sorghum
Open Pollinated[1]	Outbreeding, cross pollinated or cross fertilised Examples: Maize, Pearl Millet, Cucurbits
Self Incompatible[2]	Self Incompatible species. Examples: Rye, White Clover, Papaya
Vegetatively Propagated[4]	Vegetatively or clonally propagated species. Examples: Potatoes, Casava, Yam, Taro, Sugar Pineapple, Strawberries
Apomictic[5]	Apomitic (produce seed by asexual means) spec Examples: Green Panic, Buffel grass

[1] Catered for in the ICIS Methods Table.
Probably catered for as breeding methods usually follow those for Self P crops.
[3] Probably catered for by methods suitable for Open Pollinated crops.
[4] Generative and derivative and most maintenance methods catered for as will be either self or cross fertilised in their mode of sexual reproduction.
[5] Probably catered for by methods suitable for Vegetatively Propagated cr

Ontology

- First bit
  - A taxonomy of concepts
    - hierarchical classification of definitions
- Second bit?
  - Defining relationships
    - Exclusive
    - Is a parent of
    - Is an off-spring of
    - ?
- A program that converts a human version (english?) to a logical framework



GERMPLASM

What is it?

GERMPLASMIDENTIFIER

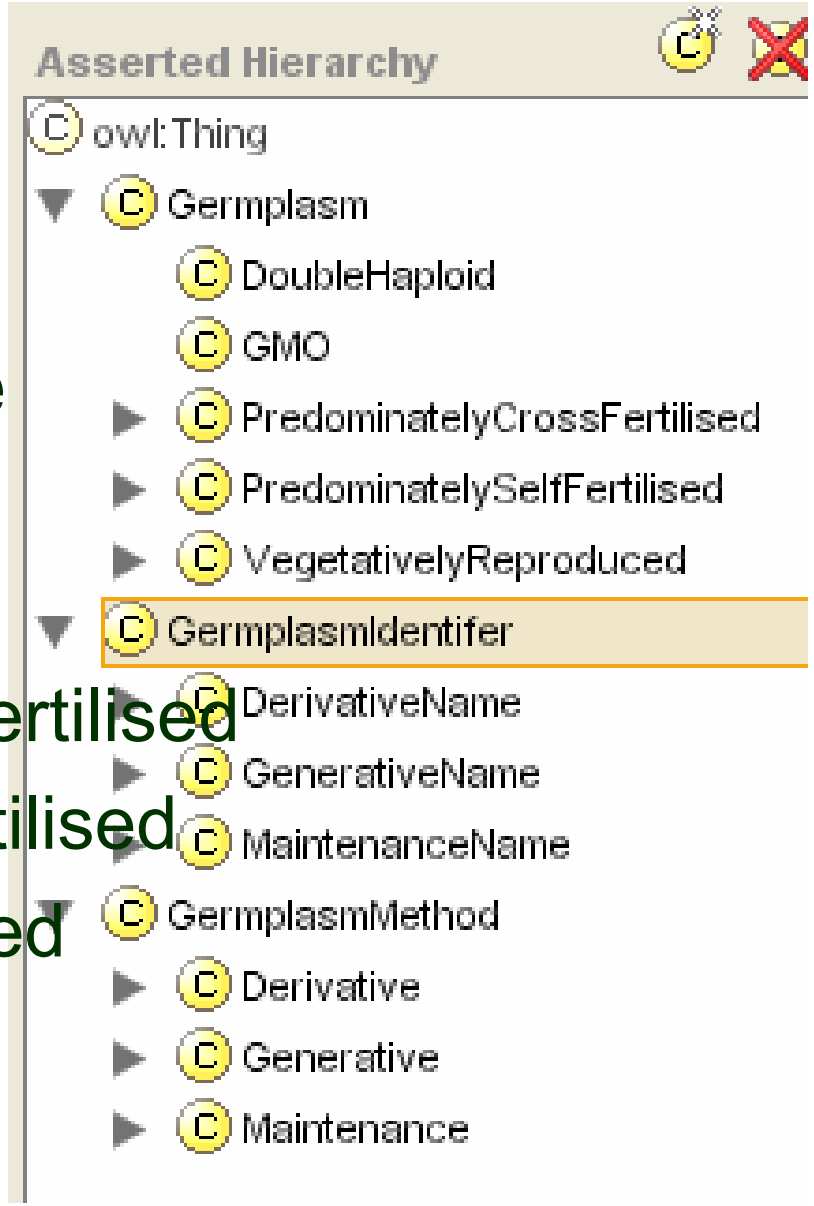
How is it named?

GERMPLASMMETHOD

How is it created?

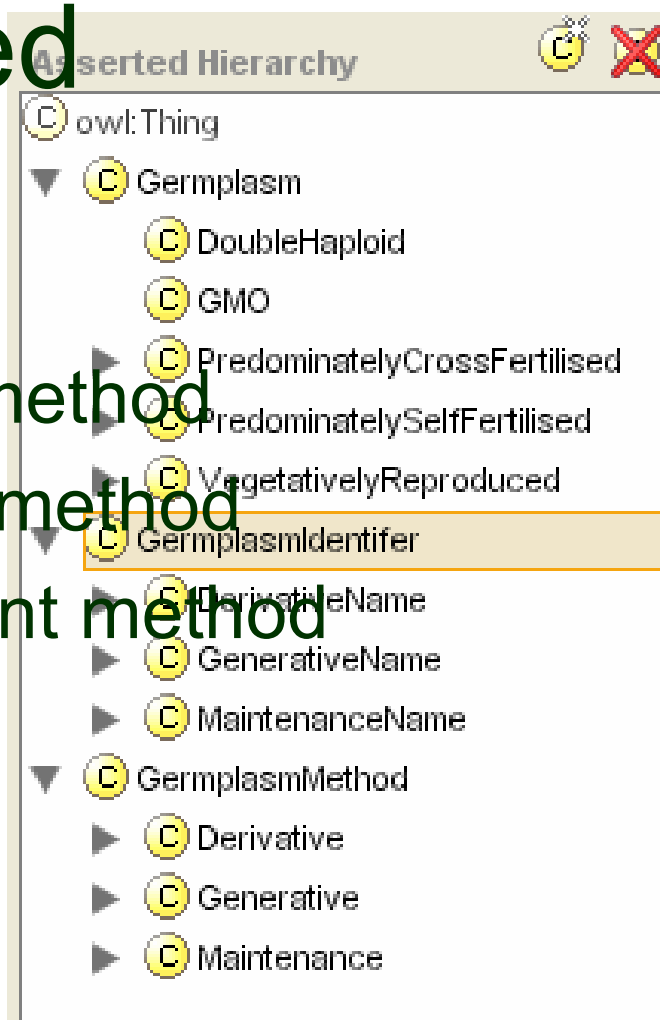
Germplasm is

- Propagules which are
  - double haploid
  - genetically modified
  - predominately cross fertilised
  - predominately self fertilised
  - vegetatively reproduced



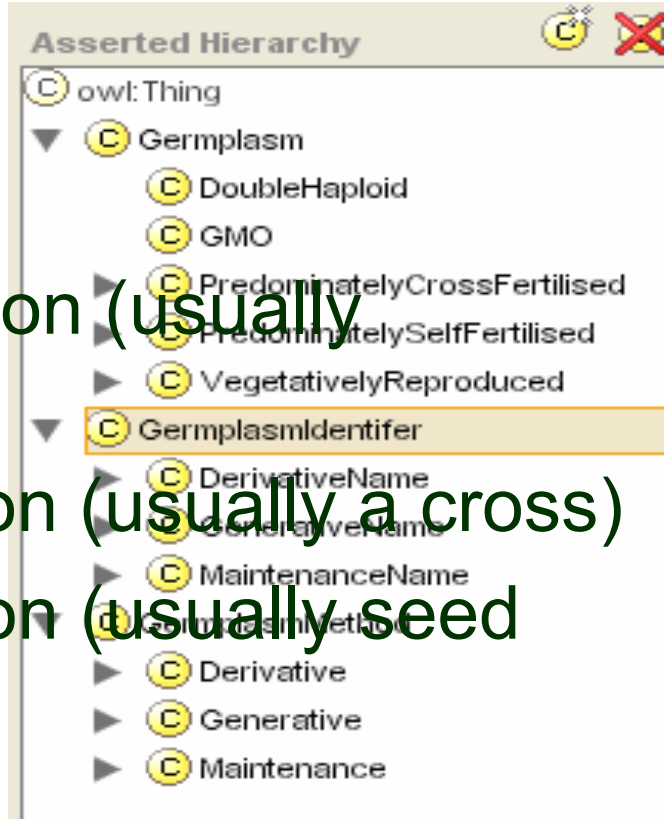
Germplasm is named

- Name of propagules
  - derived from a derivative method
  - derived from a generative method
  - derived from a management method



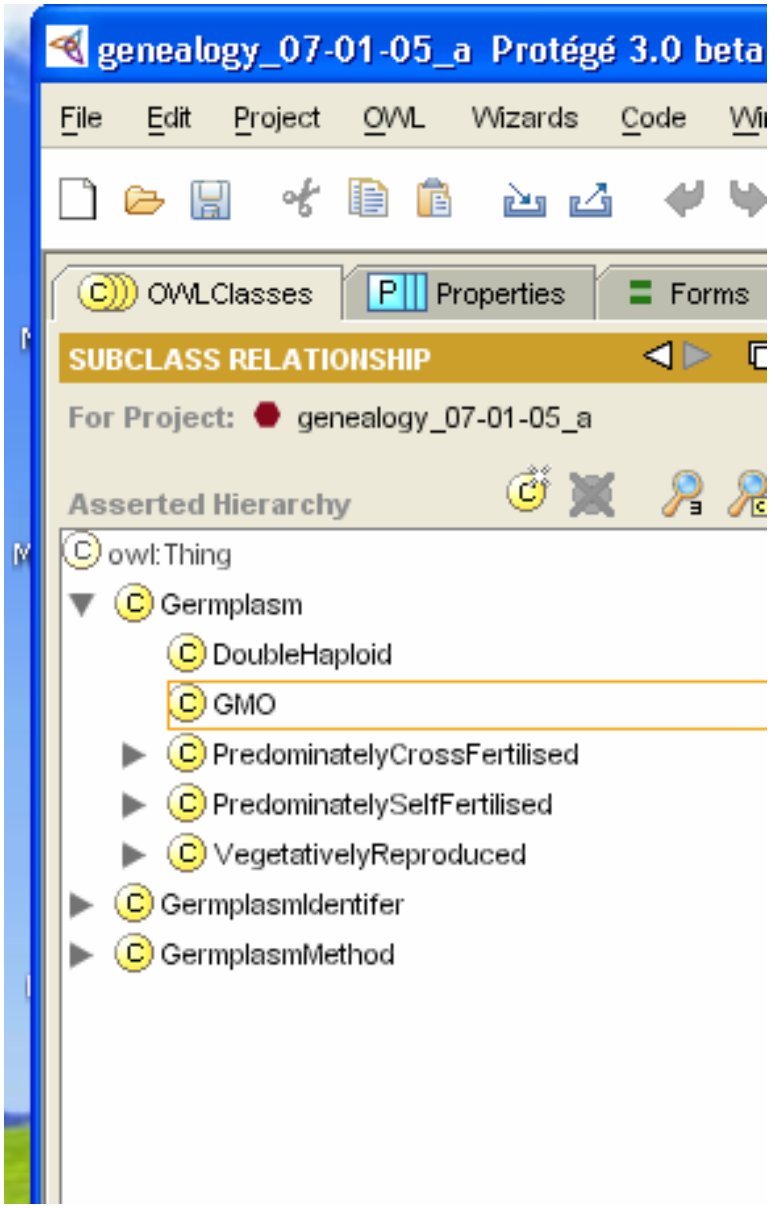
Germplasm is created

- Propagules produced
  - to decrease genetic variation (usually selection)
  - to increase genetic variation (usually a cross)
  - to maintain genetic variation (usually seed increase)



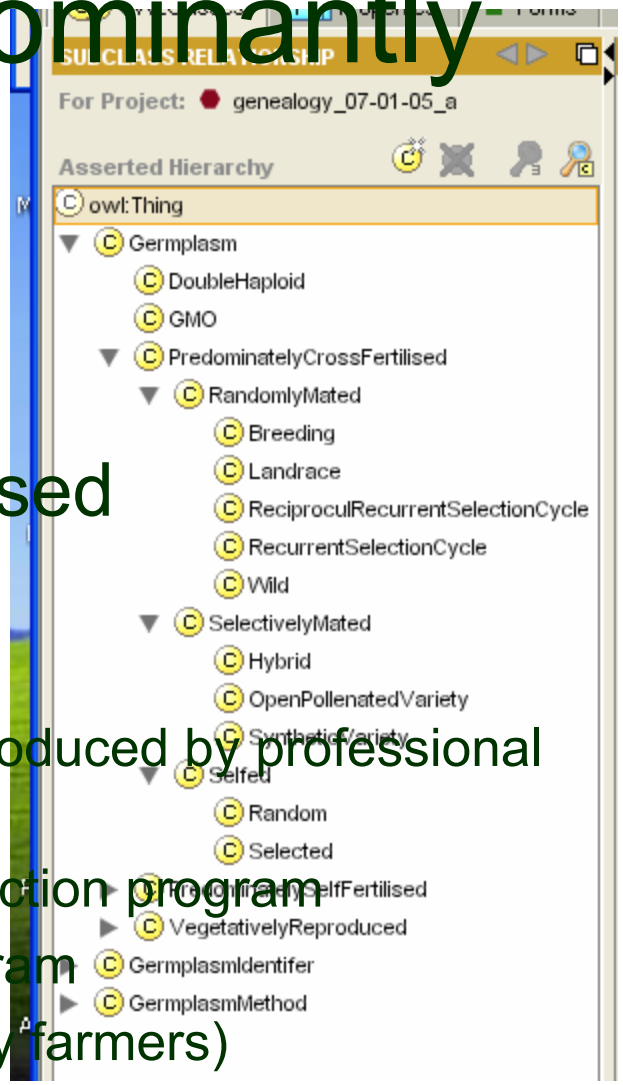
Germplasm is

- Propagules which are
  - Double haploid
  - Genetically modified



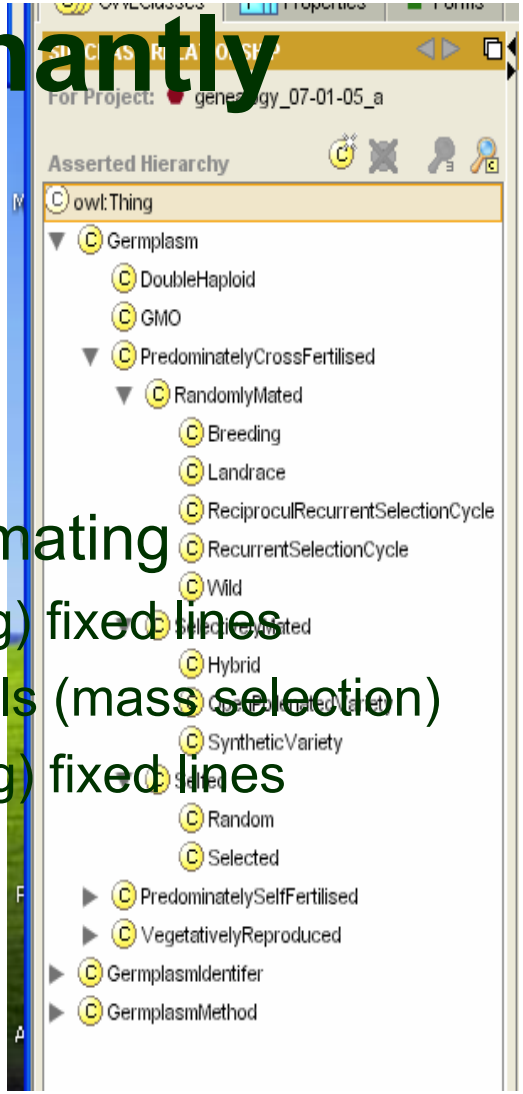
Germplasm is predominantly cross fertilised

- Propagules which are
  - predominately cross fertilised
    - and are randomly mated
      - in professional breeding
      - as a traditional variety not produced by professional breeding
      - in a reciprocal recurrent selection program
      - in a recurrent selection program
      - in the wild (not maintained by farmers)



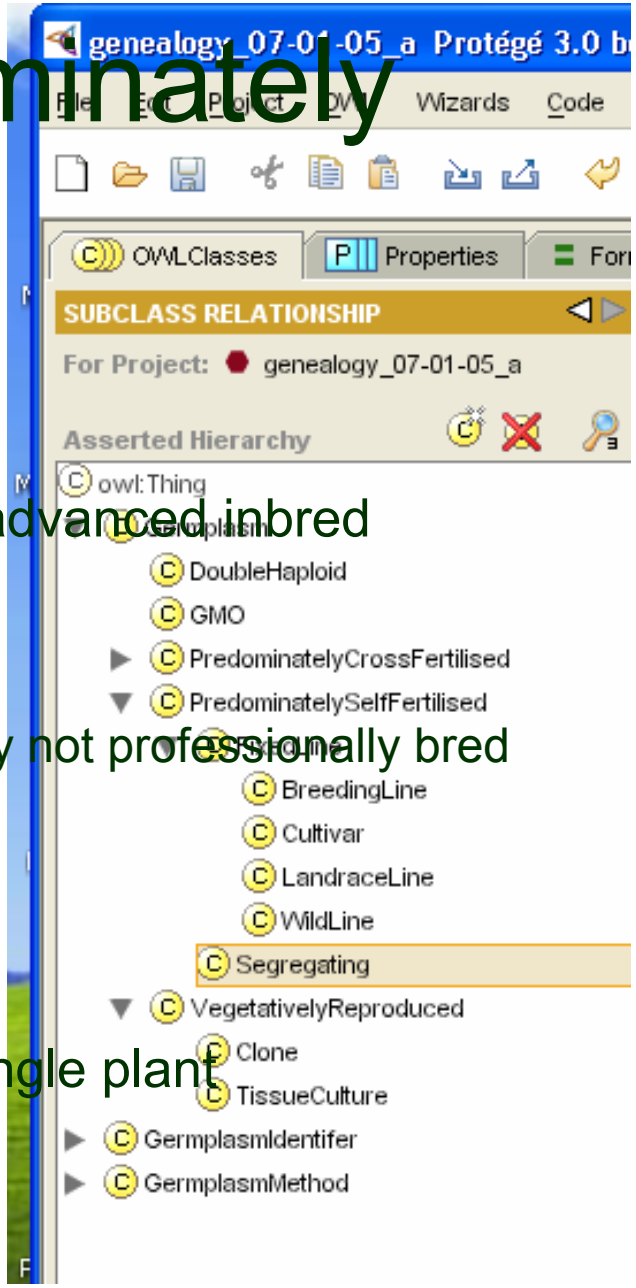
Germplasm is predominantly cross fertilised

- Propagules which are
  - predominantly cross fertilised
    - and are reproduced by selective mating
      - of a few selected (by progeny testing) fixed lines
      - of selected (by phenotype) individuals (mass selection)
      - of many selected (by progeny testing) fixed lines
    - and are reproduced by selfing
      - of randomly chosen individuals
      - of selected individuals



Germplasm is predominately self fertilised

- Propagules which are
  - predominately self fertilised
    - derived from a single plant from an advanced inbred population
      - and used for breeding
      - and released as a cultivar
      - and derived from a traditional variety not professionally bred
      - and collected off farm
    - are from an early inbred generation
  - vegetatively reproduced
    - from a single plant
    - from tissue culture derived from a single plant



Germplasm is created (derivative)

- Propagules produced
  - to decrease genetic variation (usually selection)
    - by collection (a landrace or wild population which may be an individual or a population)
    - by selection of
      - and bulking
        - selected plants
        - unselected plants
      - single plants
      - mutants
    - by doubling haploids

Germplasm is created (generative)

- Propagules produced
  - to increase genetic variation (usually a cross)
    - by producing haploids from heterozygotes
    - by mutation
      - caused by a mutagenic agent
      - recognized as a natural variant
      - by tissue culture methods
    - by crossing
      - individual plants
      - populations
      - individuals to populations
    - by genetically transforming a plant (the result of a conversion event)

Germplasm is created (management)

- Propagules produced
  - to maintain genetic variation (usually seed increase)
    - By forming a cultivar for release
    - for exporting to another organisation or project
    - by importation from elsewhere (the process of recording the accession)
    - by extraction of a fixed line from an advanced inbred population
    - by creating a population (usually by random mating)
    - by maintaining an existing population (usually by random mating)
    - by seed increase
    - by seed storage

Germplasm is named (derivative)

- Name of propagules
  - from a derivative method
    - is a breeder (or breeding program) name
    - is a selection history
      - of a (modified) pedigree program
      - of a recurrent selection program

Germplasm is named (generative)

- Name of propagules
  - from a generative method
    - due to genetic engineering event
    - due to a cross
      - is a sequential name (usually numeric) in order of crossing (overall, in a year or a cycle)
      - is a name
      - is a string representing parentage
        - by a non Purdy string
        - by a Purdy string
    - due to haploid doubling
    - due to mutation

Germplasm is named (management)

- Name of propagules
  - is an accession number
    - externally given
    - locally given
  - is a name for a commercial variety
    - Which is heterozygous
      - propogated asexually
        - is a homogenous collection of a single heterozygous genotype
        - is a traditional clonal variety
      - propagated by crossing
        - of a small number of inbred lines
        - by farmer selection
        - by mass selection
        - Many inbred lines
    - Which is homozygous
      - derived from an a single plant of a highly inbred population
      - is a traditional variety
      - is a deliberate mixture of fixed lines

- is the name of a (collection of) homozygous individual(s)
  - from a breeding population
  - collected
    - from a bulk of many plants
      - is a weed
      - not a weed
    - from a single plant
      - is a weed
      - not a weed
- is the name of a population (collection of heterogenous/heterozygous individuals)
  - used for breeding
  - is a recurrent selection cycle name
  - is a recognised genetic stock
  - is a deliberate mixture of populations
  - named after place of collection
  - collected off farm