MOLGENIS bioinformatics toolkit & XGAP eXtensible Genotype And Phenotype platform

GMOD meeting Europe
Cambridge, Sept 13, 2010

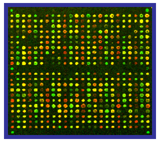
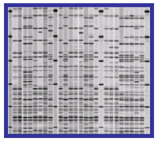
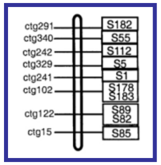
Morris A. Swertz, K Joeri van der Velde, Alexandros Kanterakis, Juha Muilu, Tomasz Adamusiak, Martijn Dijkstra, Gudmundur A. Thorisson, George Byelas, Danny Arends, Members of EU-GEN2PHEN, NL-NBIC, EU-CASIMIR, BBMRI-NL, EU-PANACEA, Anthony J. Brookes, Ritsert C. Jansen and Helen Parkinson

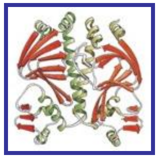
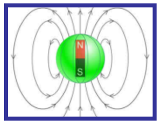
Outline

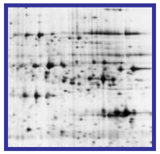
- MOLGENIS?
    - Flexible bioinformatics application toolkit
    - Demo: Model -> Generate -> Use
- XGAP?
    - eXtensible Genotype And Phenotype model
    - MOLGENIS generated xQTL & GWAS software
- Link to GMOD?
    - Chado? DAS? BioMART? Intermine? Gbrowse?

MOLGENIS
Flexible bioinformatics application toolkit for data management and interfacing

etc.



etc.

university of groningen

GEN2PHEN

LIFELINES

FIMM

BBMRI•NL

SYSGENET

CASIMIR

WAGENINGEN UR
For quality of life

EMBL-EBI

umcg

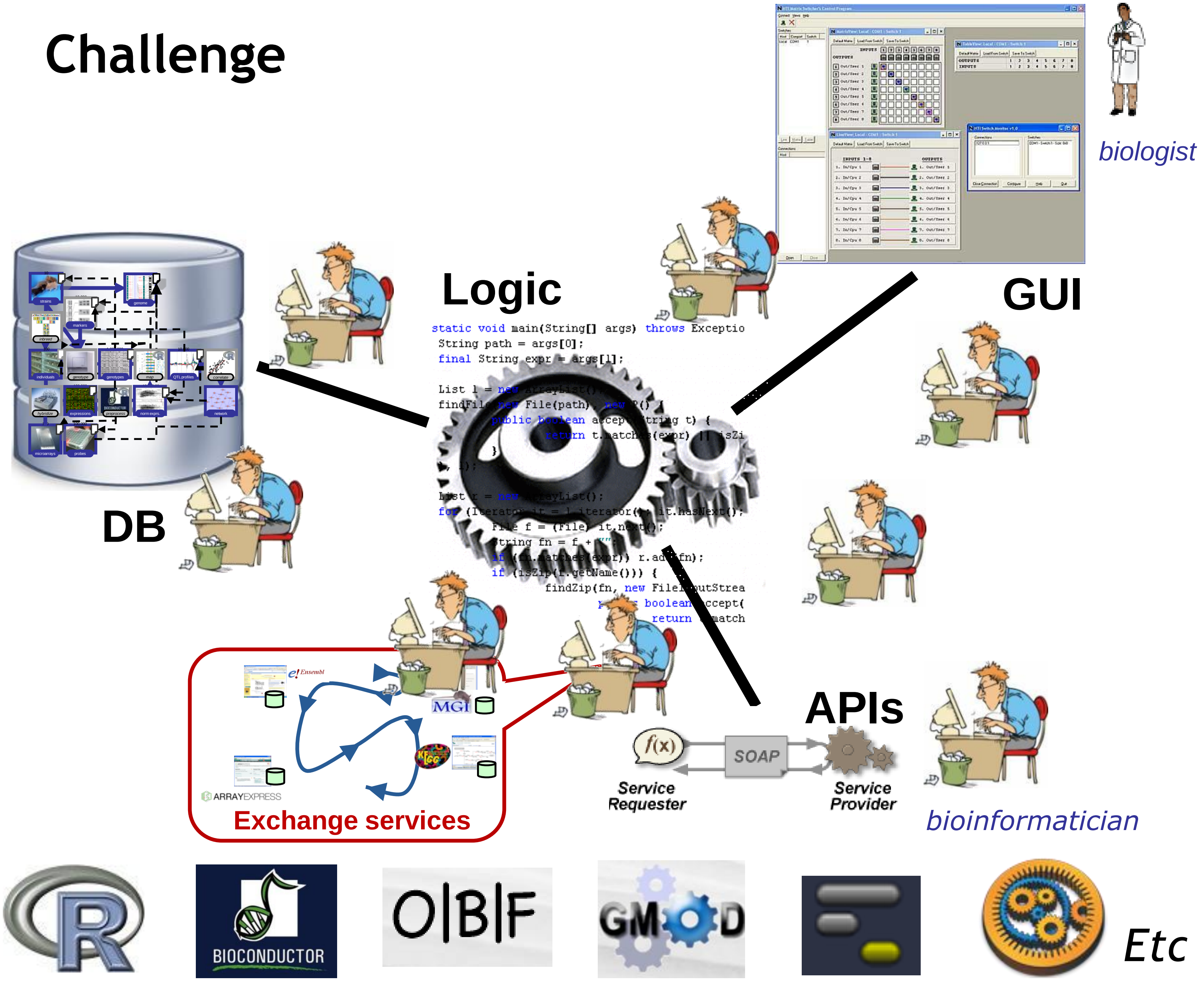
IOP

nbicassist

PANACEA

University of Leicester

Challenge



```
static void main(String[] args) throws Exceptio
 String path = args[0];
 final String expr = args[1];

 List l = new ArrayList();
 findFile(new File(path), new F() {
         public boolean accept(String t) {
                 return t.matches(expr) || isZi
         }
 }, l);

 List r = new ArrayList();
 for (Iterator it = l.iterator(); it.hasNext();
         File f = (File) it.next();
         String fn = f + "";
         if (fn.matches(expr)) r.add(fn);
         if (isZip(f.getName())) {
                 findZip(fn, new FileInputStrea
                         public boolean accept(
                                 return t.match
```

Challenge multiplied by project

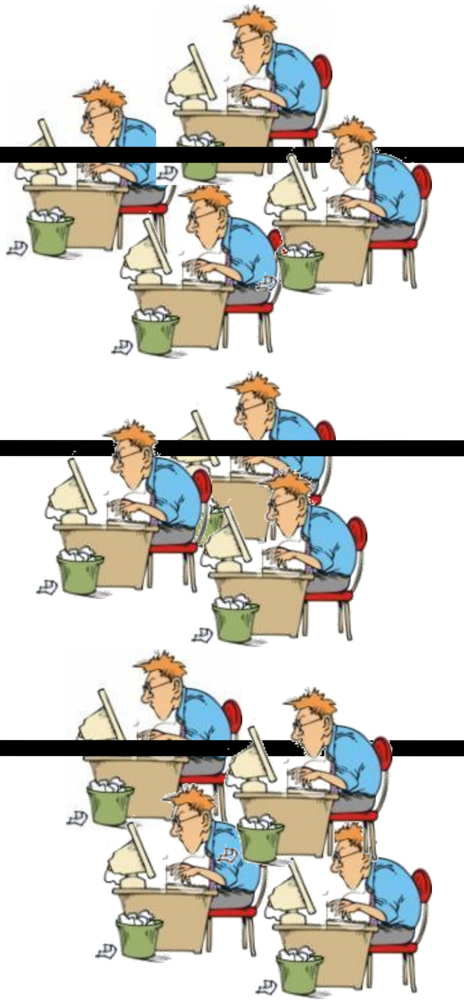
Biologist needs

NextGenSeq

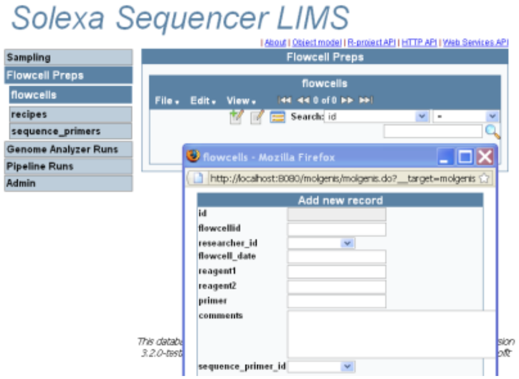
Mutation database

Moдel organısms

Work very hard



Use



Solexa Sequencer LIMS

| About | Object model | R-project API | HTTP API | Web Services API

Search: id ▾ = ▾

http://localhost:8080/molgenis/molgenis.do?__target=molgenis

Add new record

id

flowcellid

researcher_id ▾

flowcell_date

reagent1

reagent2

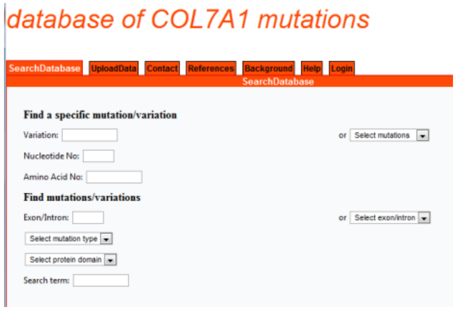
primer

comments

sequence_primer_id ▾

This databa
3.2.0-test

sion
oft



database of COL7A1 mutations

Find a specific mutation/variation

Variation: or Select mutations ▾

Nucleotide No:

Amino Acid No:

Find mutations/variations

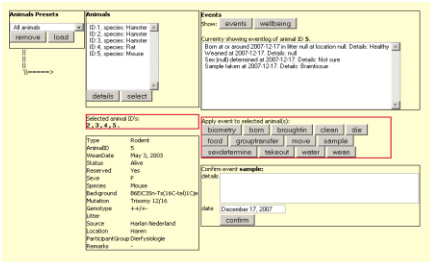
Exon/Intron: or Select exon/intron ▾

Select mutation type ▾

Select protein domain ▾

Search term:

Animal Observatory



ID:1, species: Hamster
ID:2, species: Hamster
ID:3, species: Hamster
ID:4, species: Rat
ID:5, species: Mouse

Currently showing eventlog of animal ID **5**.

Born at or around 2007-12-17 in litter null at location null. Details: Healthy
Weaned at 2007-12-17. Details: null
Sex (null) determined at 2007-12-17. Details: Not sure
Sample taken at 2007-12-17. Details: Braintissue

Selected animal ID's:
2 , 3 , 4 , 5 ,

Type	Rodent
AnimalID	5
WeanDate	May 3, 2003
Status	Alive
Reserved	Yes
Sexe	F
Species	Mouse
Background	B6EIC3Sn-Ts(16C-tel)1Cje
Mutation	Trisomy 12/16
Genotype	++/+-
Litter	
Source	Harlan Nederland
Location	Haren
ParticipantGroup	Dierfysiologie
Remarks	-

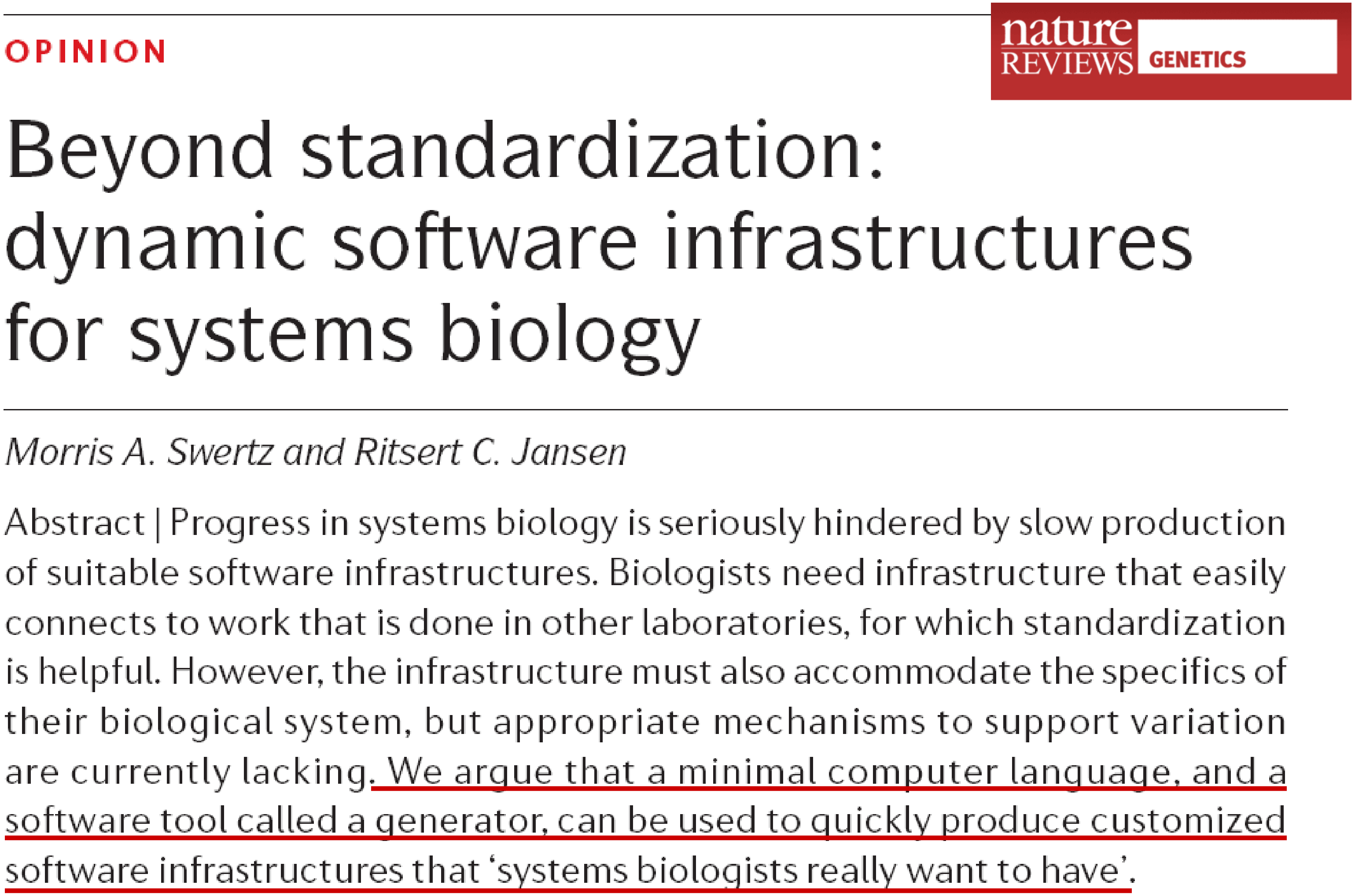
Apply event to selected animal(s):

Confirm event **sample:**

details

date December 17, 2007

Needed alternative method



OPINION

nature REVIEWS GENETICS

Beyond standardization: dynamic software infrastructures for systems biology

Morris A. Swertz and Ritsert C. Jansen

Abstract | Progress in systems biology is seriously hindered by slow production of suitable software infrastructures. Biologists need infrastructure that easily connects to work that is done in other laboratories, for which standardization is helpful. However, the infrastructure must also accommodate the specifics of their biological system, but appropriate mechanisms to support variation are currently lacking. We argue that a minimal computer language, and a software tool called a generator, can be used to quickly produce customized software infrastructures that 'systems biologists really want to have'.

http://www.molgenis.org
Swertz & Jansen (2007) *Nature Reviews Genetics* 8, 235-243
Swertz et al (2004) *Bioinformatics* 20(13), 2075-83

MOLGENIS

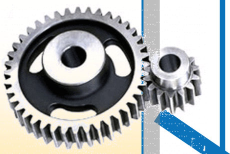
Model in DSL

NextGenSeq

Mutation database

Model organisms

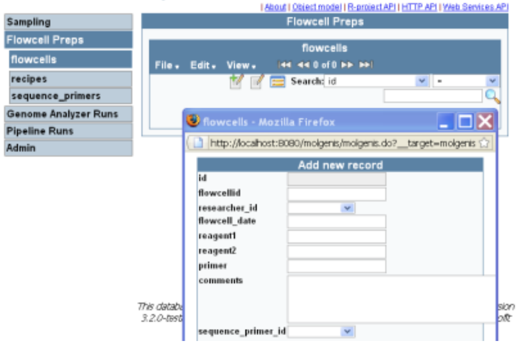
Generator



repeat often

Use generated software

Solexa Sequencer LIMS



| About | Object model | R-project API | HTTP API | Web Services API

- Sampling
- Flowcell Preps
  - flowcells
  - recipes
  - sequence_primers
- Genome Analyzer Runs
- Pipeline Runs
- Admin

Flowcell Preps

flowcells

File▾ Edit▾ View▾ |◀◀ ◀◀ 0 of 0 ▶▶ ▶▶|

Search: id ▾ = ▾

flowcells - Mozilla Firefox

http://localhost:8080/molgenis/molgenis.do?__target=molgenis

Add new record

id

flowcellid

researcher_id

flowcell_date

reagent1

reagent2

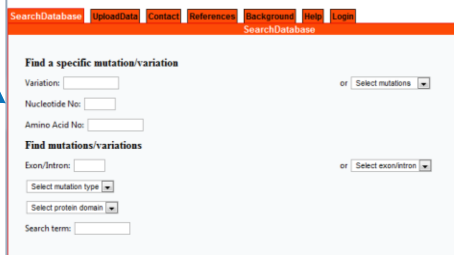
primer

comments

sequence_primer_id

This databa sion
3.2.0-test oft

database of COL7A1 mutations



SearchDatabase | UploadData | Contact | References | Background | Help | Login

SearchDatabase

Find a specific mutation/variation

Variation: or Select mutations ▾

Nucleotide No:

Amino Acid No:

Find mutations/variations

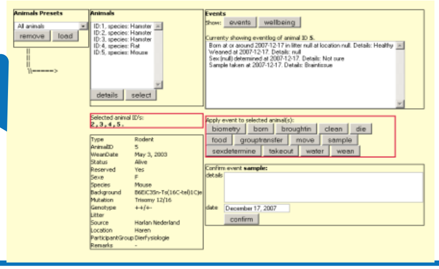
Exon/Intron: or Select exon/intron ▾

Select mutation type ▾

Select protein domain ▾

Search term:

Animal Observatory



Animals Presets

All animals ▾

remove load

Animals

ID:1, species: Hamster
ID:2, species: Hamster
ID:3, species: Hamster
ID:4, species: Rat
ID:5, species: Mouse

details select

Events

Show: events wellbeing

Currenty showing eventlog of animal ID **5**.

Born at or around 2007-12-17 in litter null at location null. Details: Healthy
Weaned at 2007-12-17. Details: null
Sex (null) determined at 2007-12-17. Details: Not sure
Sample taken at 2007-12-17. Details: Braintissue

Selected animal ID's:
2 , 3 , 4 , 5 ,

Type	Rodent
AnimalID	5
WeanDate	May 3, 2003
Status	Alive
Reserved	Yes
Sexe	F
Species	Mouse
Background	B6EiC3Sn-Ts(16C-tel)1Cje
Mutation	Trisomy 12/16
Genotype	++/+-
Litter	
Source	Harlan Nederland
Location	Haren
ParticipantGroup	Dierfysiologie
Remarks	-

Apply event to selected animal(s):

biometry born broughtin clean die
food grouptransfer move sample
sexdetermine takeout water wean

Confirm event **sample:**

details

date December 17, 2007

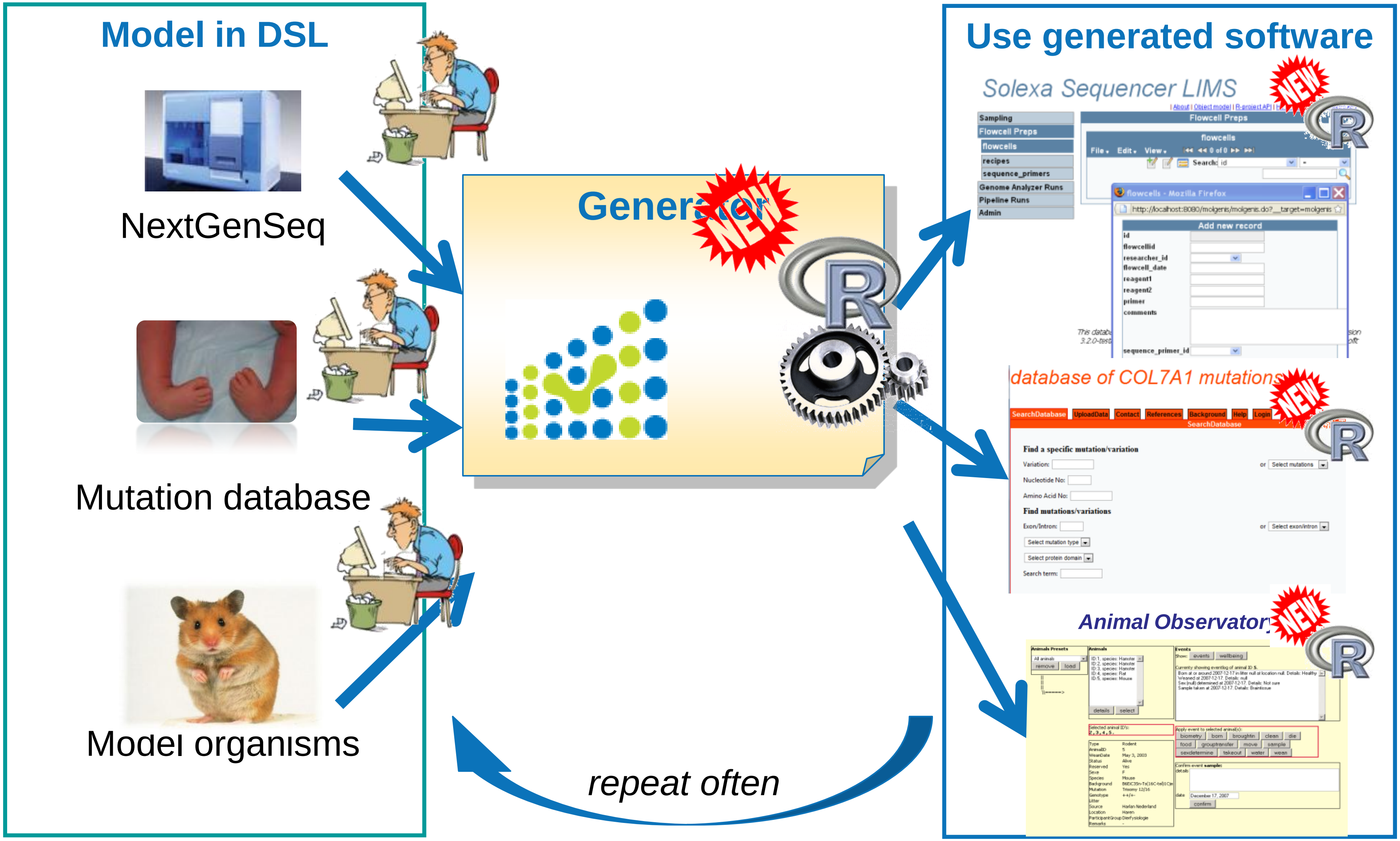
confirm

http://www.molgenis.org

Swertz & Jansen (2007) *Nature Reviews Genetics* 8, 235-243

Swertz et al (2004) *Bioinformatics* 20(13), 2075-83

MOLGENIS: Reuse in light of large variation



Model in DSL

NextGenSeq

Mutation database

Model organisms

Generator

repeat often

Use generated software

Solexa Sequencer LIMS

NEW
| About | Object model | R-project API |

Sampling
Flowcell Preps
flowcells
recipes
sequence_primers
Genome Analyzer Runs
Pipeline Runs
Admin

Flowcell Preps

flowcells

File▾ Edit▾ View▾ |◀◀ ◀◀ 0 of 0 ▶▶ ▶▶|

Search: id =

flowcells - Mozilla Firefox

http://localhost:8080/molgenis/molgenis.do?__target=molgenis

Add new record

id
flowcellid
researcher_id
flowcell_date
reagent1
reagent2
primer
comments
sequence_primer_id

database of COL7A1 mutations

SearchDatabase

Find a specific mutation/variation

Variation: or Select mutations

Nucleotide No:

Amino Acid No:

Find mutations/variations

Exon/Intron: or Select exon/intron

Select mutation type

Select protein domain

Search term:

Animal Observatory

Animals Presets

All animals

Animals

ID:1, species: Hamster
ID:2, species: Hamster
ID:3, species: Hamster
ID:4, species: Rat
ID:5, species: Mouse

Events

Show:

Currenty showing eventlog of animal ID **5**.

Born at or around 2007-12-17 in litter null at location null. Details: Healthy
Weaned at 2007-12-17. Details: null
Sex (null) determined at 2007-12-17. Details: Not sure
Sample taken at 2007-12-17. Details: Braintissue

Selected animal ID's:
2 , 3 , 4 , 5 ,

Type	Rodent
AnimalID	5
WeanDate	May 3, 2003
Status	Alive
Reserved	Yes
Sexe	F
Species	Mouse
Background	B6EiC3Sn-Ts(16C-tel)1Cje
Mutation	Trisomy 12/16
Genotype	++/+-
Litter	
Source	Harlan Nederland
Location	Haren
ParticipantGroup	Dierfysiologie
Remarks	-

Apply event to selected animal(s):

Confirm event **sample:**

details

date December 17, 2007

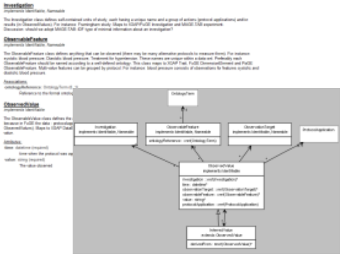
http://www.molgenis.org

Swertz & Jansen (2007) *Nature Reviews Genetics* 8, 235-243

Swertz et al (2004) *Bioinformatics* 20(13), 2075-83

Example output

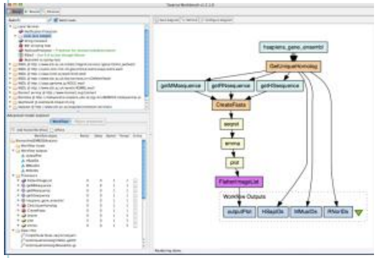
❶ UML documentation of your model



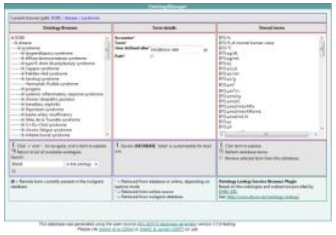
❹ Connect to statistics

```
find.investigation()
102 downloaded

obs<-find.observedvalue(
43,920 downloaded

#some calculation
add.inferredvalue(res)
36 added
```

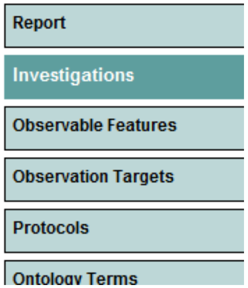
❺ Workflow ready web-services



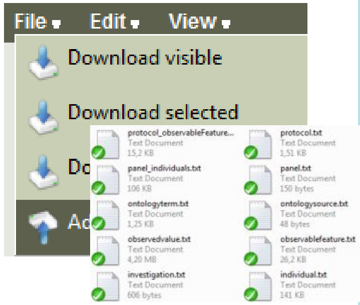
❻ plugin your own scripts (eg OntologyBrowser)



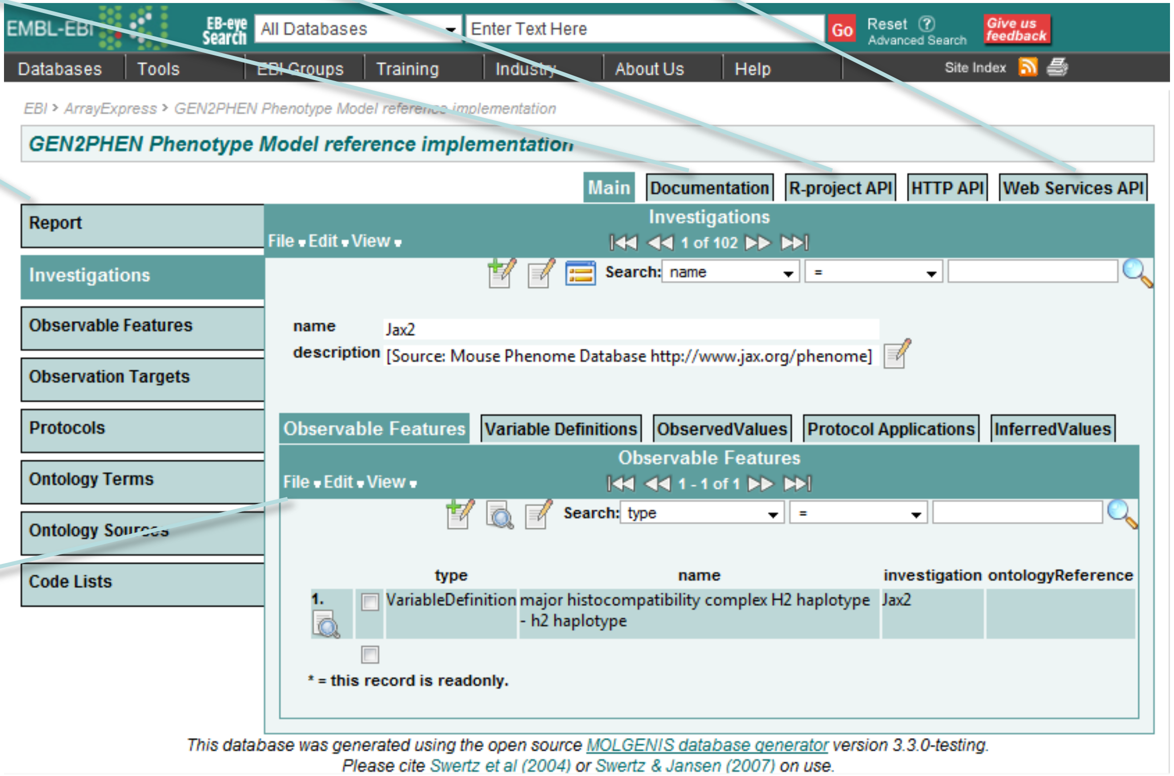
❷ Edit & trace your data



❸ Import/export to Excel



protocol_observableFeature... Text Document 15,2 KB
protocol.txt Text Document 1,51 KB
panel_individuals.txt Text Document 106 KB
panel.txt Text Document 150 bytes
ontologyterm.txt Text Document 1,25 KB
ontologysource.txt Text Document 48 bytes
observedvalue.txt Text Document 4,20 MB
observablefeature.txt Text Document 26,2 KB
investigation.txt Text Document 606 bytes
individual.txt Text Document 141 KB



EBI > ArrayExpress > GEN2PHEN Phenotype Model reference implementation

GEN2PHEN Phenotype Model reference implementation

Investigations

File ▾ Edit ▾ View ▾ |◀◀ ◀◀ 1 of 102 ▶▶ ▶▶|

Search: name =

name Jax2

description [Source: Mouse Phenome Database http://www.jax.org/phenome]

Observable Features

File ▾ Edit ▾ View ▾ |◀◀ ◀◀ 1 - 1 of 1 ▶▶ ▶▶|

Search: type =

		type	name	investigation	ontologyReference
1.	□	VariableDefinition	major histocompatibility complex H2 haplotype - h2 haplotype	Jax2	
	□				

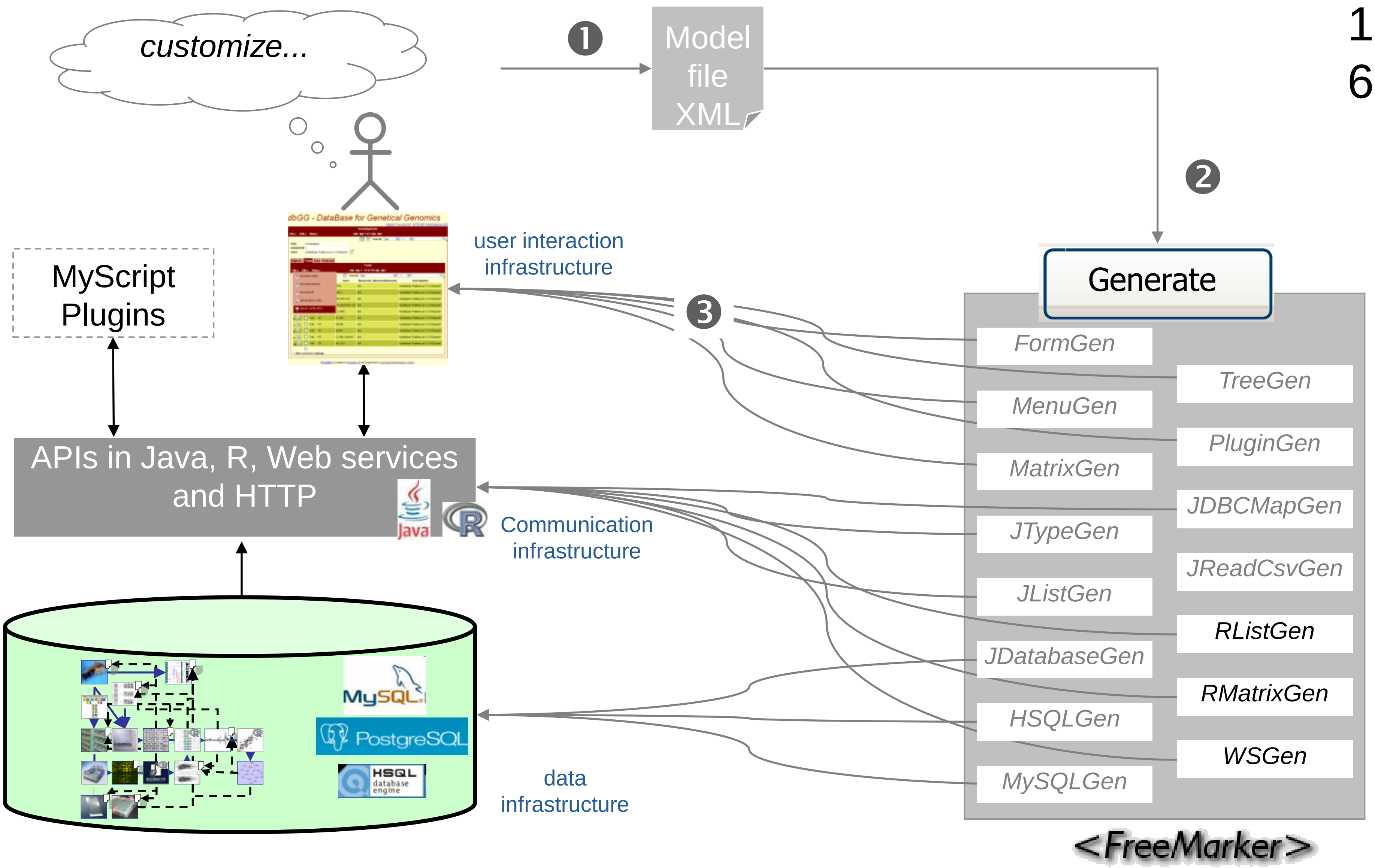
* = this record is readonly.

This database was generated using the open source MOLGENIS database generator version 3.3.0-testing. Please cite Swertz et al (2004) or Swertz & Jansen (2007) on use.

MOLGENIS demo

Model -> Generate -> Use

Model -> Generate -> Use



customize...

❶ Model file XML

❷ Generate

❸

user interaction infrastructure

MyScript Plugins

APIs in Java, R, Web services and HTTP

Communication infrastructure

data infrastructure

FormGen

TreeGen

MenuGen

PluginGen

MatrixGen

JDBCMapGen

JTypeGen

JReadCsvGen

JListGen

RListGen

JDatabaseGen

RMatrixGen

HSQLGen

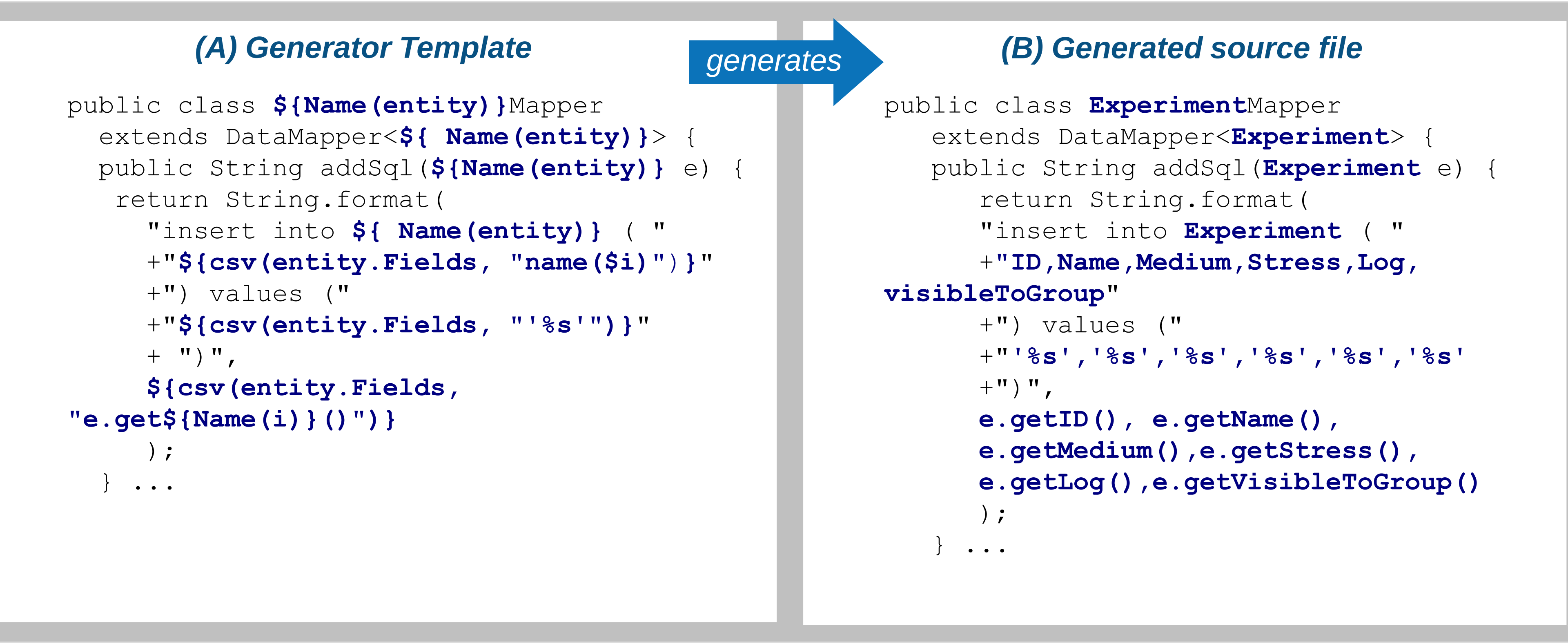
WSGen

MySQLGen

<FreeMarker>

A generator = template

e.g. ${Name(entity)} -> ExperimentMapper



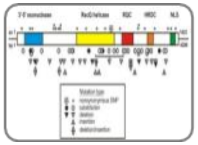
(A) Generator Template

```
public class ${Name(entity)}Mapper
  extends DataMapper<${ Name(entity)}> {
  public String addSql(${Name(entity)} e) {
   return String.format(
     "insert into ${ Name(entity)} ( "
     +"${csv(entity.Fields, "name($i)")}"
     +") values ("
     +"${csv(entity.Fields, "'%s'")}"
     + ")",
     ${csv(entity.Fields,
"e.get${Name(i)}()")}
     );
  } ...
```

(B) Generated source file

```
public class ExperimentMapper
   extends DataMapper<Experiment> {
   public String addSql(Experiment e) {
      return String.format(
      "insert into Experiment ( "
      +"ID,Name,Medium,Stress,Log,
visibleToGroup"
      +") values ("
      +"'%s','%s','%s','%s','%s','%s'
      +")",
      e.getID(), e.getName(),
      e.getMedium(),e.getStress(),
      e.getLog(),e.getVisibleToGroup()
      );
   } ...
```

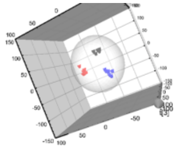
Usage examples in Life Sciences



Mutation

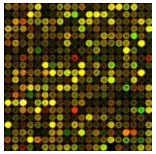
Phenotype

Sequencing LIMS



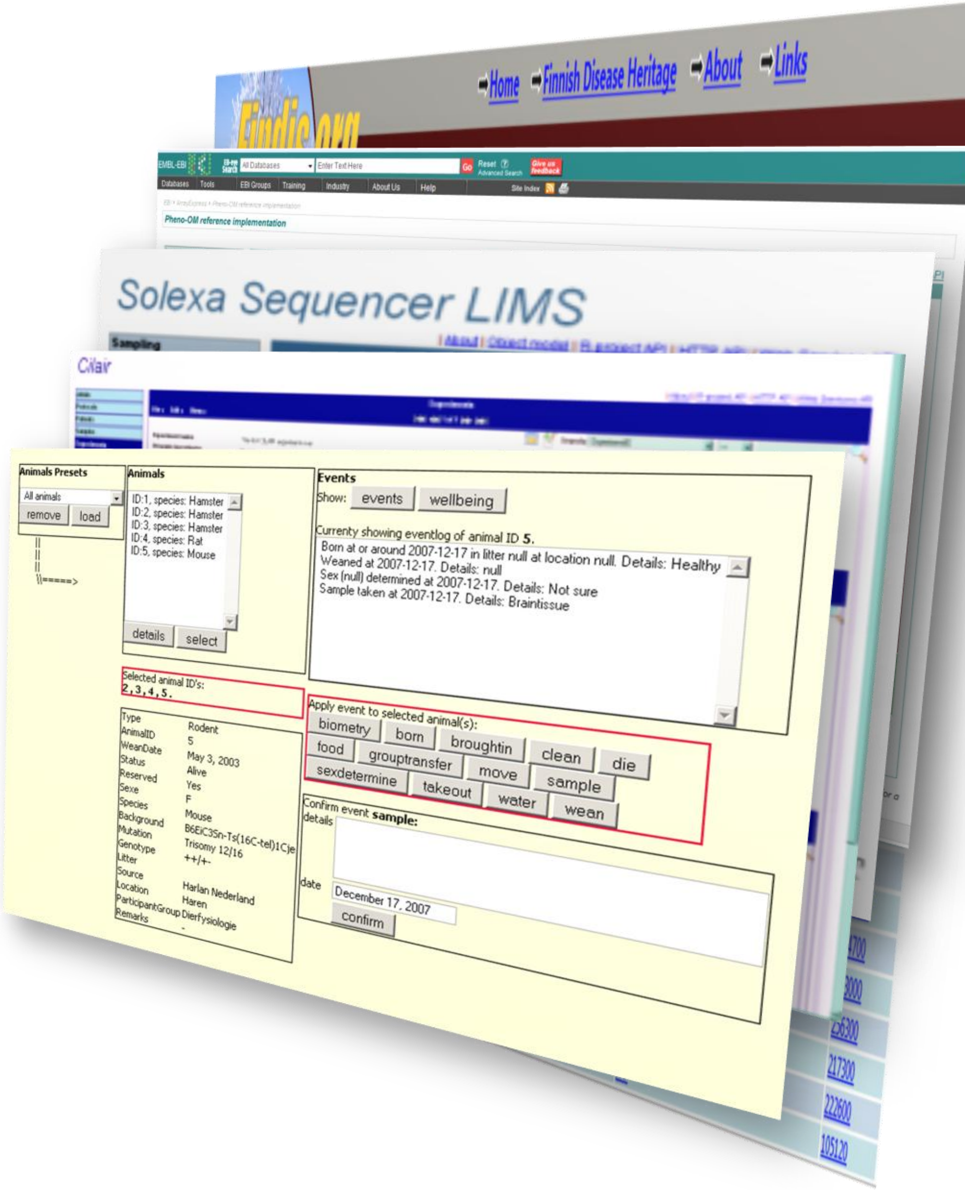
Proteo/Metabolomics

Animal LIMS



GWAS / GWLS

<add your project here>



Animals Presets

All animals

remove | load

Animals

ID:1, species: Hamster
ID:2, species: Hamster
ID:3, species: Hamster
ID:4, species: Rat
ID:5, species: Mouse

details | select

Selected animal ID's:
2,3,4,5.

Type	Rodent
AnimalID	5
WeanDate	May 3, 2003
Status	Alive
Reserved	Yes
Sexe	F
Species	Mouse
Background	B6EiC3Sn-Ts(16C-tel)1Cje
Mutation	Trisomy 12/16
Genotype	++/+-
Litter	
Source	Harlan Nederland
Location	Haren
ParticipantGroup	Dierfysiologie
Remarks	-

Events

Show: events | wellbeing

Currenty showing eventlog of animal ID **5.**

Born at or around 2007-12-17 in litter null at location null. Details: Healthy
Weaned at 2007-12-17. Details: null
Sex (null) determined at 2007-12-17. Details: Not sure
Sample taken at 2007-12-17. Details: Braintissue

Apply event to selected animal(s):

biometry | born | broughtin | clean | die | food | grouptransfer | move | sample | sexdetermine | takeout | water | wean

Confirm event **sample:**

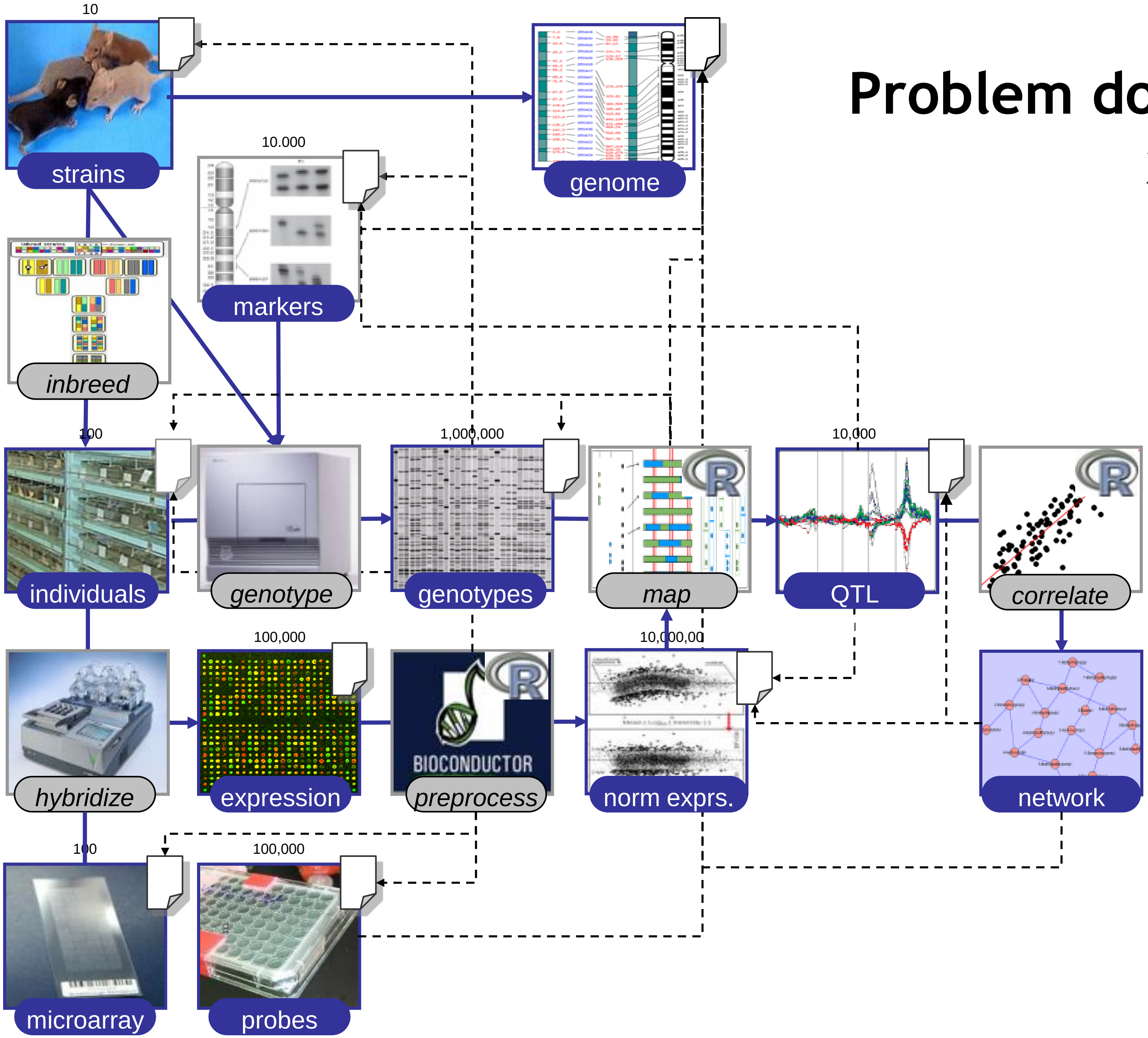
details

date December 17, 2007

confirm

XGAP
extensible genotype and phenotype data model for xQTL

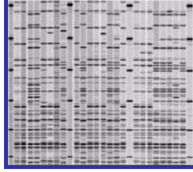
Problem domain: xQTLs GWAS



http://www.xgap.org

Swertz, van der Velde et al (2010) Genome Biology 9;11(3): R27.

Data in practice



Genotype data

Data in matrices

Subjects: PANELS

Traits: MARKERS

DATA ELEMENTS

TRAIT × SUBJECT

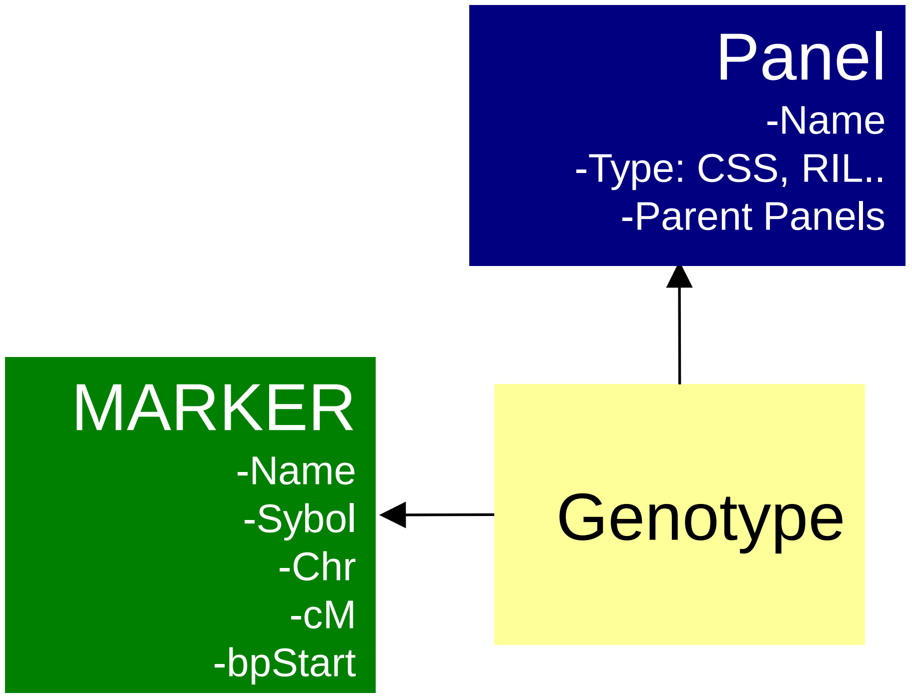
Annotations in practice

- Annotations in tables, e.g. Marker

MARKER
-Name
--Symbol
-Chr
-cM
-bpStart

Name	Symbol	Chr	cM	bpStart	mb
C1M1	I_1_pkP1050	1	-18.2603	168807	0.168807
C1M2	I_2_pkP1101	1	-17.2825	992188	0.992188
C1M3	I_3_pkP1103	1	-11.959	1884415	1.884415
C1M4	I_4_pkP1052	1	-6.1004	2818973	2.818973
C1M5	I_5_egPE107	1	-3.5488	3502476	3.502476
C1M6	I_6_egPF101	1	-1.4887	4338254	4.338254
C1M7	I_7_pkP1054	1	-0.6162	4845515	4.845515
C1M8	I_8_egPH102	1	0.4597	5893622	5.893622
C1M9	I_9_pkP1057	1	0.9366	6359867	6.359867
C1M10	I_10_pkP1116	1	2.1576	7589863	7.589863
C1M11	I_11_egPK103	1	2.4087	7894081	7.894081
C1M12	I_12_pkP1059	1	2.9456	8654360	8.65436
C1M13	I_13_pkP1122	1	3.7959	9569914	9.569914
C1M14	I_14_egPN104	1	4.7801	10259909	10.259909
C1M15	I_15_egPO105	1	6.0193	11085295	11.085295
C1M16	I_16_pkP1068	1	7.5226	11760182	11.760182

Model: try 1



Panel

-Name

-Type: CSS, RIL..

-Parent Panels

MARKER

-Name

-Sybol

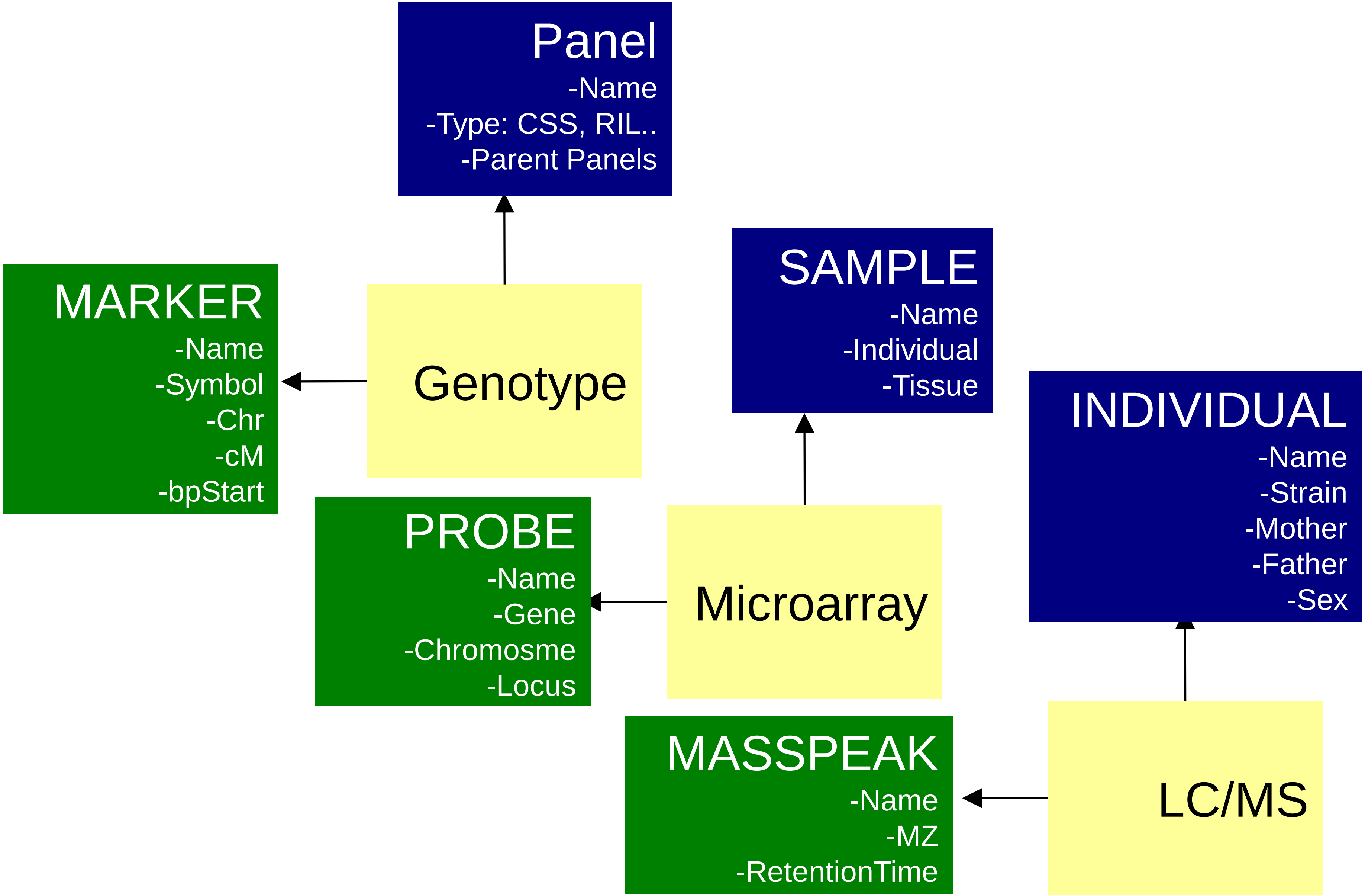
-Chr

-cM

-bpStart

Genotype

But...



Panel
-Name
-Type: CSS, RIL..
-Parent Panels

MARKER
-Name
-Symbol
-Chr
-cM
-bpStart

Genotype

SAMPLE
-Name
-Individual
-Tissue

INDIVIDUAL
-Name
-Strain
-Mother
-Father
-Sex

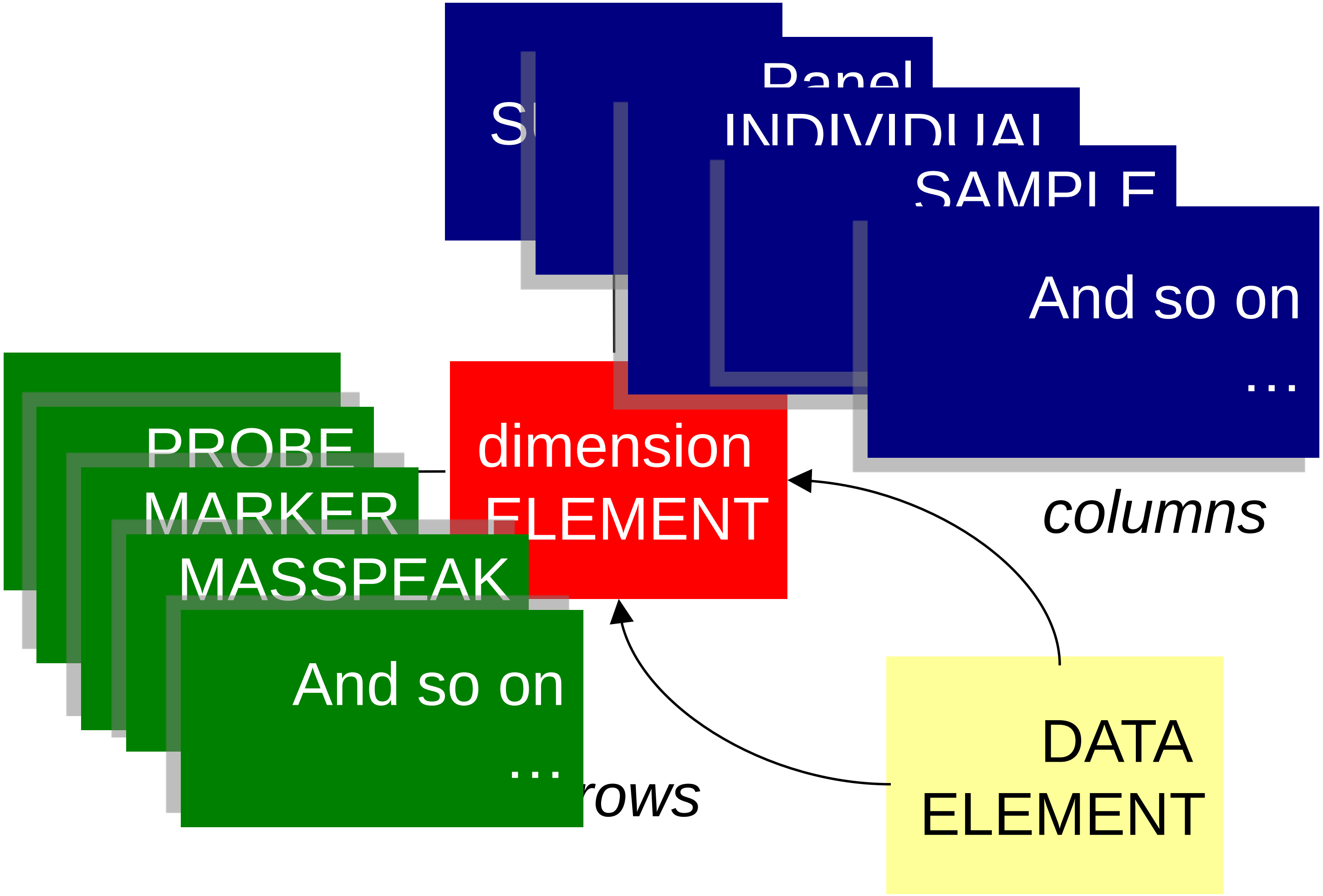
PROBE
-Name
-Gene
-Chromosme
-Locus

Microarray

MASSPEAK
-Name
-MZ
-RetentionTime

LC/MS

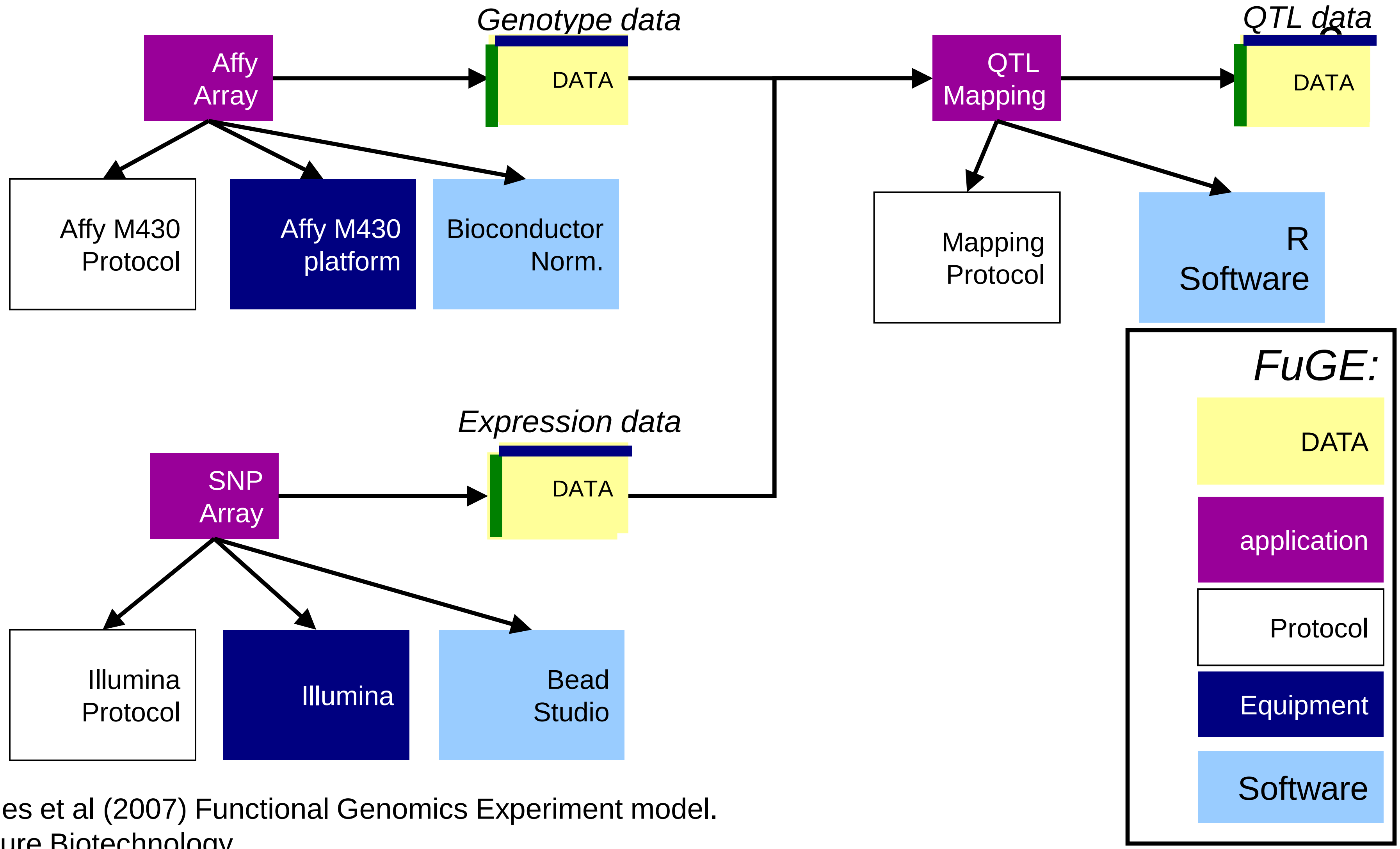
XGAP model: <any trait> X <any subject>



http://www.xgap.org

Swertz, van der Velde et al (2010) Genome Biology 9;11(3): R27.

Extending on FuGE



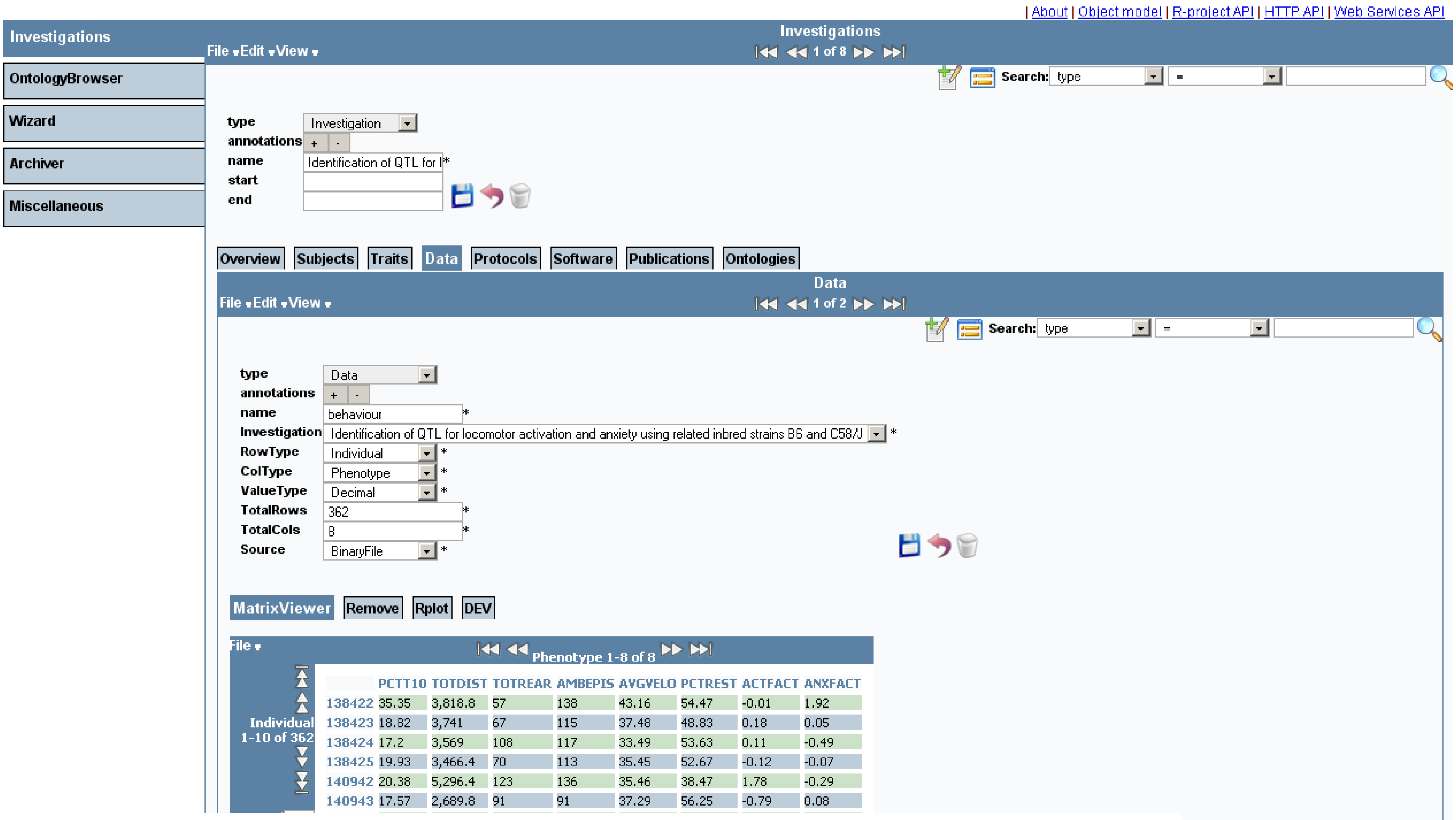
Jones et al (2007) Functional Genomics Experiment model. Nature Biotechnology

XGAP

extensible genotype and
phenotype software platform for xQTL

Generated: user interfaces

XGAP - eXtensible Genotype and Phenotype platform



Investigations

Search: type =

type	Investigation
annotations	+ -
name	Identification of QTL for l*
start	
end	

Data

Search: type =

type	Data
annotations	+ -
name	behaviour *
Investigation	Identification of QTL for locomotor activation and anxiety using related inbred strains B6 and C58/J *
RowType	Individual *
ColType	Phenotype *
ValueType	Decimal *
TotalRows	362 *
TotalCols	8 *
Source	BinaryFile *

Phenotype 1-8 of 8

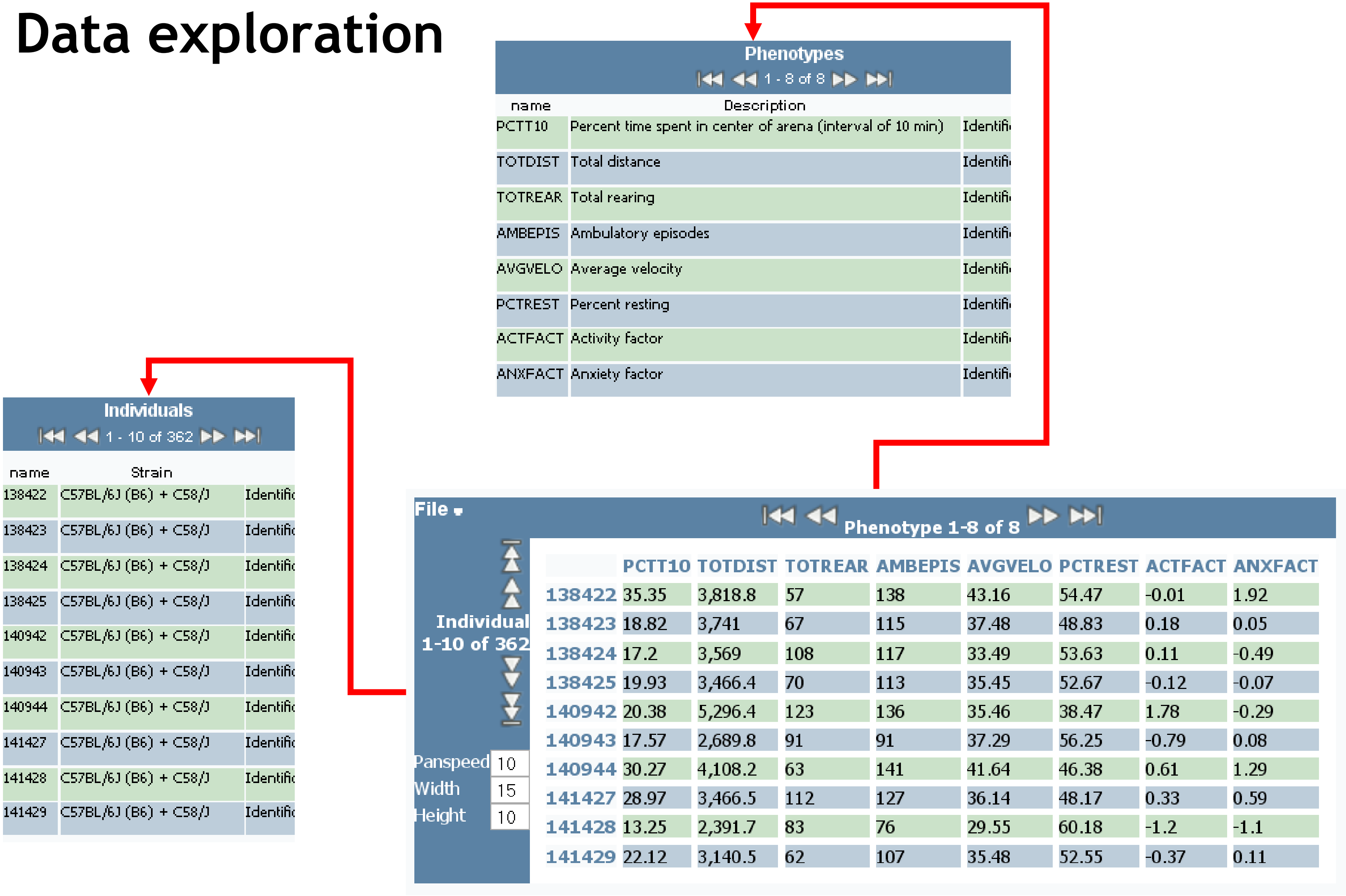
Individual 1-10 of 362

	PCTT10	TOTDIST	TOTREAR	AMBEPIS	AVGVELO	PCTREST	ACTFACT	ANXFACT
138422	35.35	3,818.8	57	138	43.16	54.47	-0.01	1.92
138423	18.82	3,741	67	115	37.48	48.83	0.18	0.05
138424	17.2	3,569	108	117	33.49	53.63	0.11	-0.49
138425	19.93	3,466.4	70	113	35.45	52.67	-0.12	-0.07
140942	20.38	5,296.4	123	136	35.46	38.47	1.78	-0.29
140943	17.57	2,689.8	91	91	37.29	56.25	-0.79	0.08

http://www.xgap.org

Swertz, van der Velde et al (2010) Genome Biology 9;11(3): R27.

Data exploration



Phenotypes

|◀◀ ◀◀ 1 - 8 of 8 ▶▶ ▶▶|

name	Description	
PCTT10	Percent time spent in center of arena (interval of 10 min)	Identifi
TOTDIST	Total distance	Identifi
TOTREAR	Total rearing	Identifi
AMBEPIS	Ambulatory episodes	Identifi
AVGVELO	Average velocity	Identifi
PCTREST	Percent resting	Identifi
ACTFACT	Activity factor	Identifi
ANXFACT	Anxiety factor	Identifi

Individuals

|◀◀ ◀◀ 1 - 10 of 362 ▶▶ ▶▶|

name	Strain	
138422	C57BL/6J (B6) + C58/J	Identific
138423	C57BL/6J (B6) + C58/J	Identific
138424	C57BL/6J (B6) + C58/J	Identific
138425	C57BL/6J (B6) + C58/J	Identific
140942	C57BL/6J (B6) + C58/J	Identific
140943	C57BL/6J (B6) + C58/J	Identific
140944	C57BL/6J (B6) + C58/J	Identific
141427	C57BL/6J (B6) + C58/J	Identific
141428	C57BL/6J (B6) + C58/J	Identific
141429	C57BL/6J (B6) + C58/J	Identific

File ▾ |◀◀ ◀◀ Phenotype 1-8 of 8 ▶▶ ▶▶|

Individual 1-10 of 362

Panspeed 10

Width 15

Height 10

	PCTT10	TOTDIST	TOTREAR	AMBEPIS	AVGVELO	PCTREST	ACTFACT	ANXFACT
138422	35.35	3,818.8	57	138	43.16	54.47	-0.01	1.92
138423	18.82	3,741	67	115	37.48	48.83	0.18	0.05
138424	17.2	3,569	108	117	33.49	53.63	0.11	-0.49
138425	19.93	3,466.4	70	113	35.45	52.67	-0.12	-0.07
140942	20.38	5,296.4	123	136	35.46	38.47	1.78	-0.29
140943	17.57	2,689.8	91	91	37.29	56.25	-0.79	0.08
140944	30.27	4,108.2	63	141	41.64	46.38	0.61	1.29
141427	28.97	3,466.5	112	127	36.14	48.17	0.33	0.59
141428	13.25	2,391.7	83	76	29.55	60.18	-1.2	-1.1
141429	22.12	3,140.5	62	107	35.48	52.55	-0.37	0.11

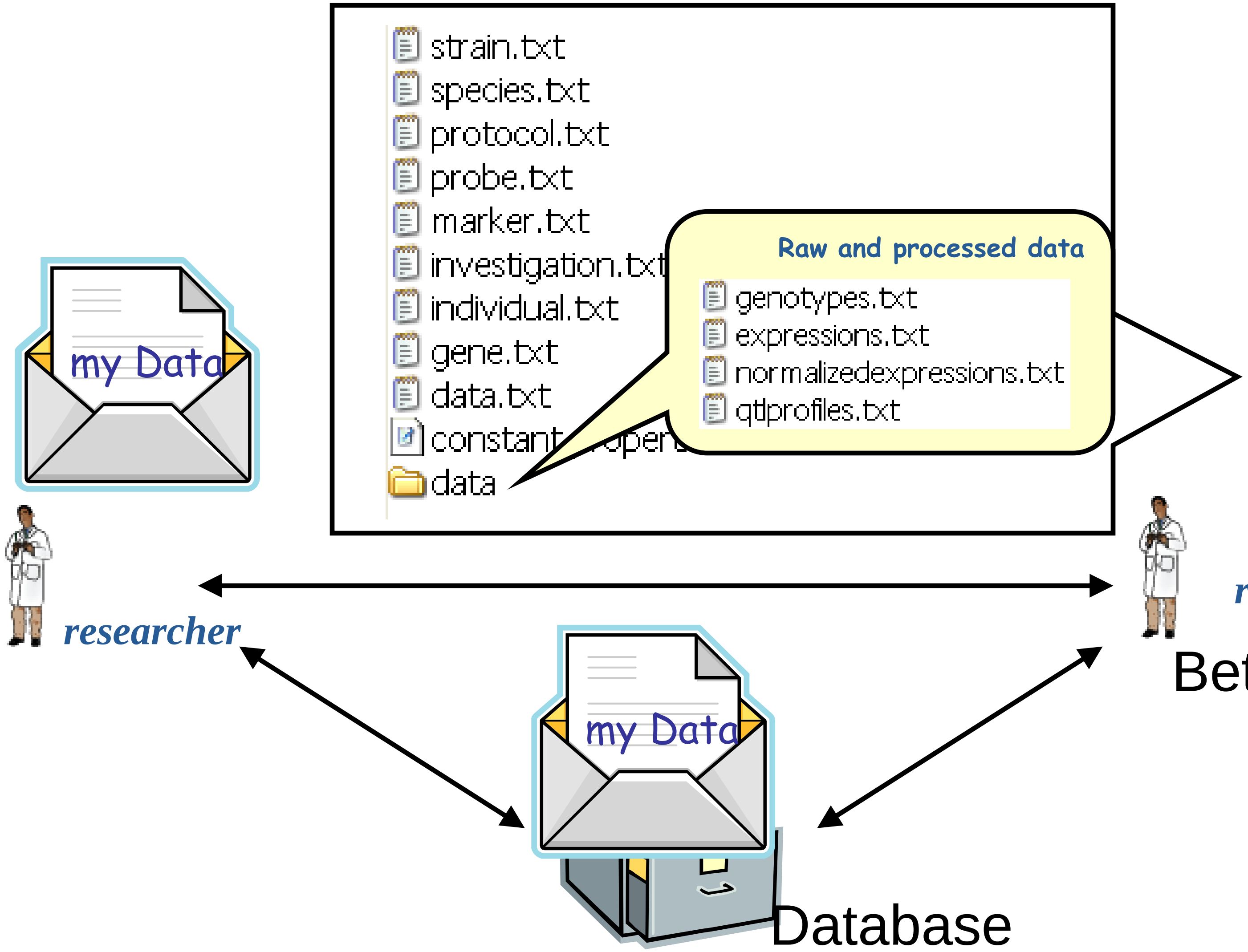
http://www.xgap.org

Swertz, van der Velde et al (2010) Genome Biology 9;11(3): R27.

Generated: common database/format

3
2

Simple text based format



strain.txt
species.txt
protocol.txt
probe.txt
marker.txt
investigation.txt
individual.txt
gene.txt
data.txt
constant oper
data

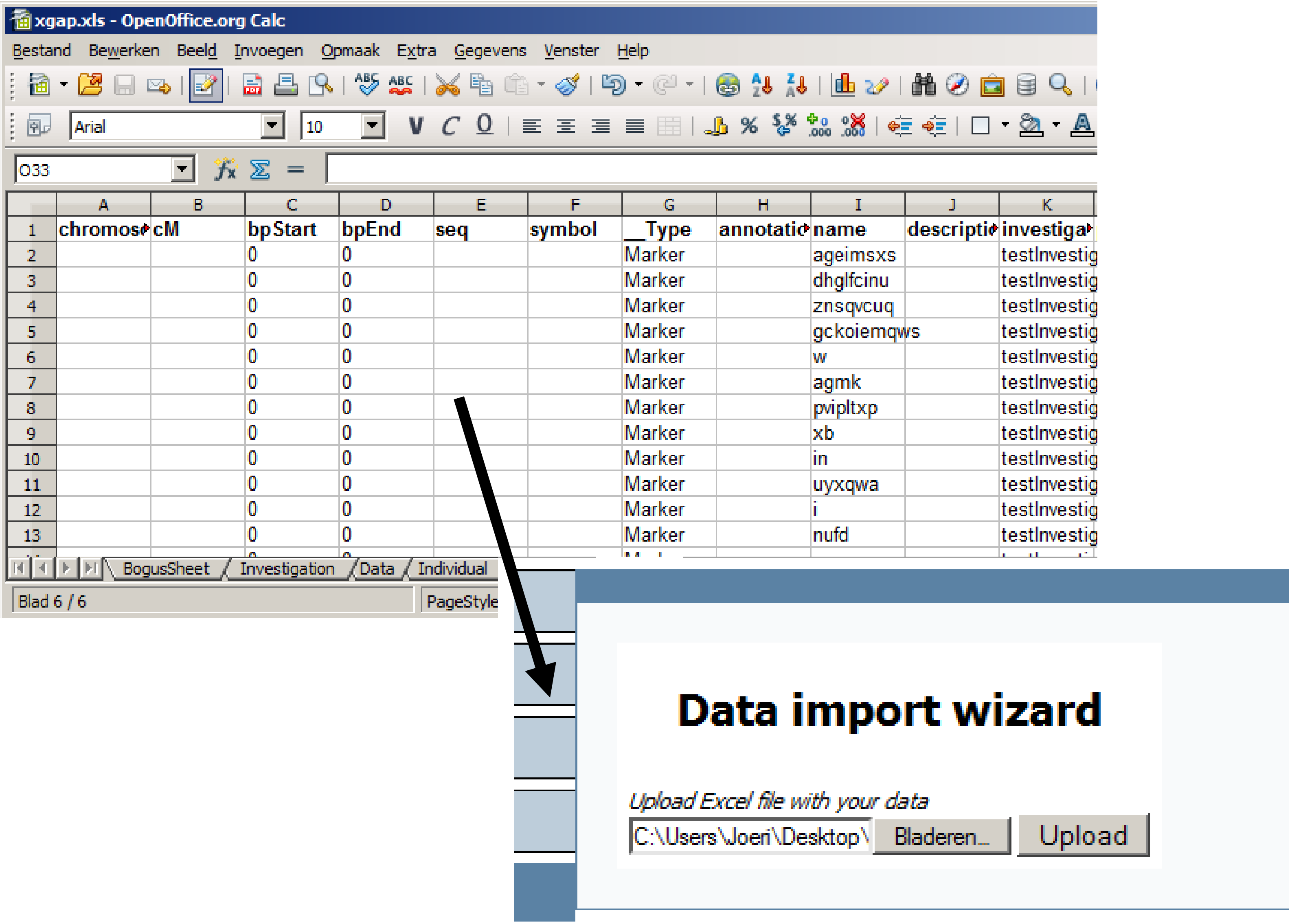
Raw and processed data

genotypes.txt
expressions.txt
normalizedexpressions.txt
qtlprofiles.txt

Between labs

Database

Generated: common database/format


xgap.xls - OpenOffice.org Calc
Bestand Bewerken Beeld Invoegen Opmaak Extra Gegevens Venster Help
Arial
10
O33
BogusSheet
Investigation
Data
Individual
Blad 6 / 6
PageStyle

	A	B	C	D	E	F	G	H	I	J	K
1	chromos	cM	bpStart	bpEnd	seq	symbol	__Type	annotatio	name	descripti	investiga
2			0	0			Marker		ageimsxs		testInvestig
3			0	0			Marker		dhglfcinu		testInvestig
4			0	0			Marker		znsqvcuq		testInvestig
5			0	0			Marker		gckoiemqws		testInvestig
6			0	0			Marker		w		testInvestig
7			0	0			Marker		agmk		testInvestig
8			0	0			Marker		pvipltxp		testInvestig
9			0	0			Marker		xb		testInvestig
10			0	0			Marker		in		testInvestig
11			0	0			Marker		uyxqwa		testInvestig
12			0	0			Marker		i		testInvestig
13			0	0			Marker		nufd		testInvestig

Data import wizard

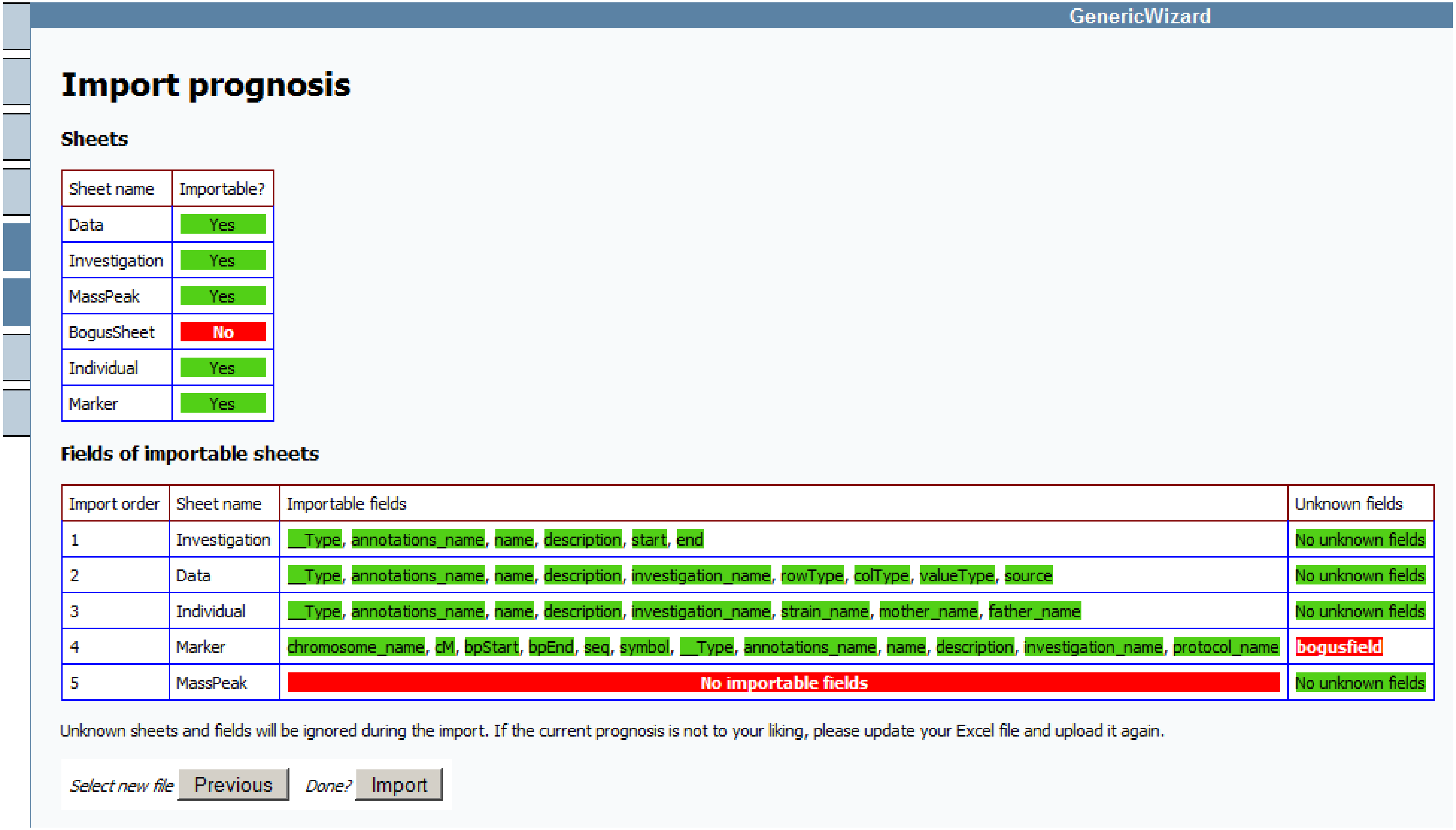
Upload Excel file with your data

C:\Users\Joeri\Desktop\

http://www.xgap.org

Swertz, van der Velde et al (2010) Genome Biology 9;11(3): R27.

Plugin: import wizard



Import prognosis

Sheets

Sheet name	Importable?
Data	Yes
Investigation	Yes
MassPeak	Yes
BogusSheet	No
Individual	Yes
Marker	Yes

Fields of importable sheets

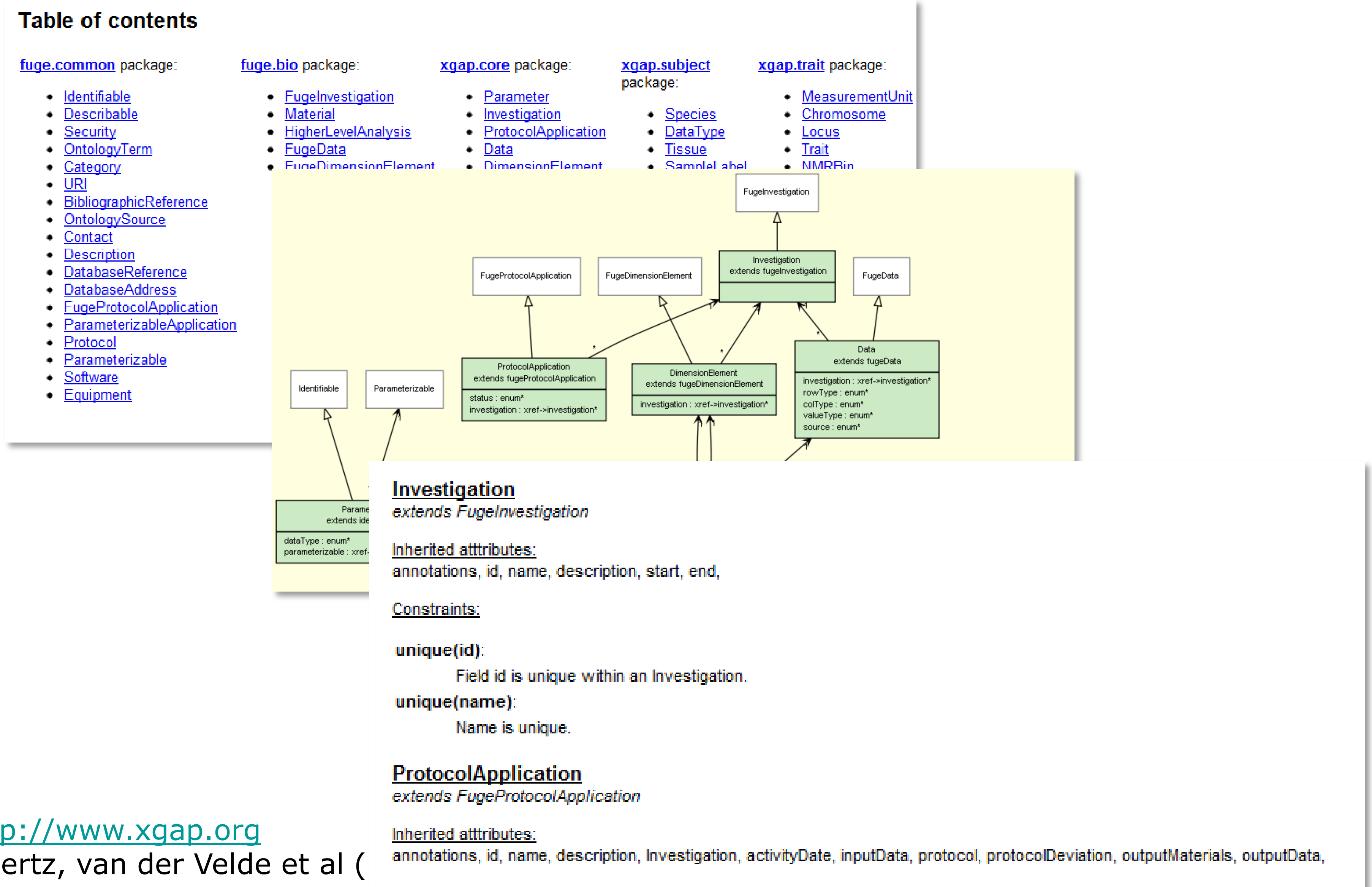
Import order	Sheet name	Importable fields	Unknown fields
1	Investigation	__Type, annotations_name, name, description, start, end	No unknown fields
2	Data	__Type, annotations_name, name, description, investigation_name, rowType, colType, valueType, source	No unknown fields
3	Individual	__Type, annotations_name, name, description, investigation_name, strain_name, mother_name, father_name	No unknown fields
4	Marker	chromosome_name, cM, bpStart, bpEnd, seq, symbol, __Type, annotations_name, name, description, investigation_name, protocol_name	bogusfield
5	MassPeak	No importable fields	No unknown fields

Unknown sheets and fields will be ignored during the import. If the current prognosis is not to your liking, please update your Excel file and upload it again.

http://www.xgap.org

Swertz, van der Velde et al (2010) Genome Biology 9;11(3): R27.

Generated: rich user documentation



XGAP 1.4 distro prototype documentation.

Table of contents

Investigation

extends FugeInvestigation

Inherited atttributes:

annotations, id, name, description, start, end,

Constraints:

unique(id):

Field id is unique within an Investigation.

unique(name):

Name is unique.

ProtocolApplication

extends FugeProtocolApplication

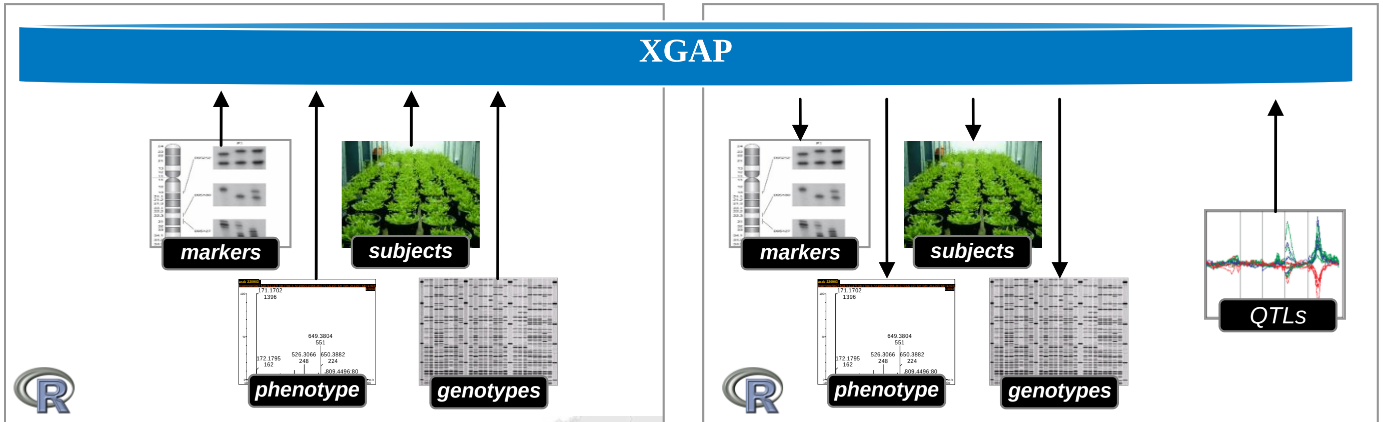
Inherited atttributes:

annotations, id, name, description, Investigation, activityDate, inputData, protocol, protocolDeviation, outputMaterials, outputData,

http://www.xgap.org

Swertz, van der Velde et al (

Generated: connection to R statistics



```
#connect to my XGAP database
source("http://aserver/xgap/api/R")

#upload my 'metanetwork' investigation
add.investigation(name="metanetwork")

#use 'metanetwork' investigation
use.investigation(name="metanetwork")

#upload subjects and traits
add.marker(name=rownames(markers),
           chr =markers$chr,
           cm  =markers$cM)
add.metabolite(name=rownames(metabolites))
add.subject(name=colnames(genotypes))

#upload genotype and phenotype data matrices
add.datamatrix(geno,
    name="geno" rowtype="marker"
    coltype="subject" valuetype="text")
add.datamatrix(mpheno,
    name="mexpr" rowtype="metabolite"
    coltype="subject" valuetype="double")
```

Scientist **A** uploads raw data

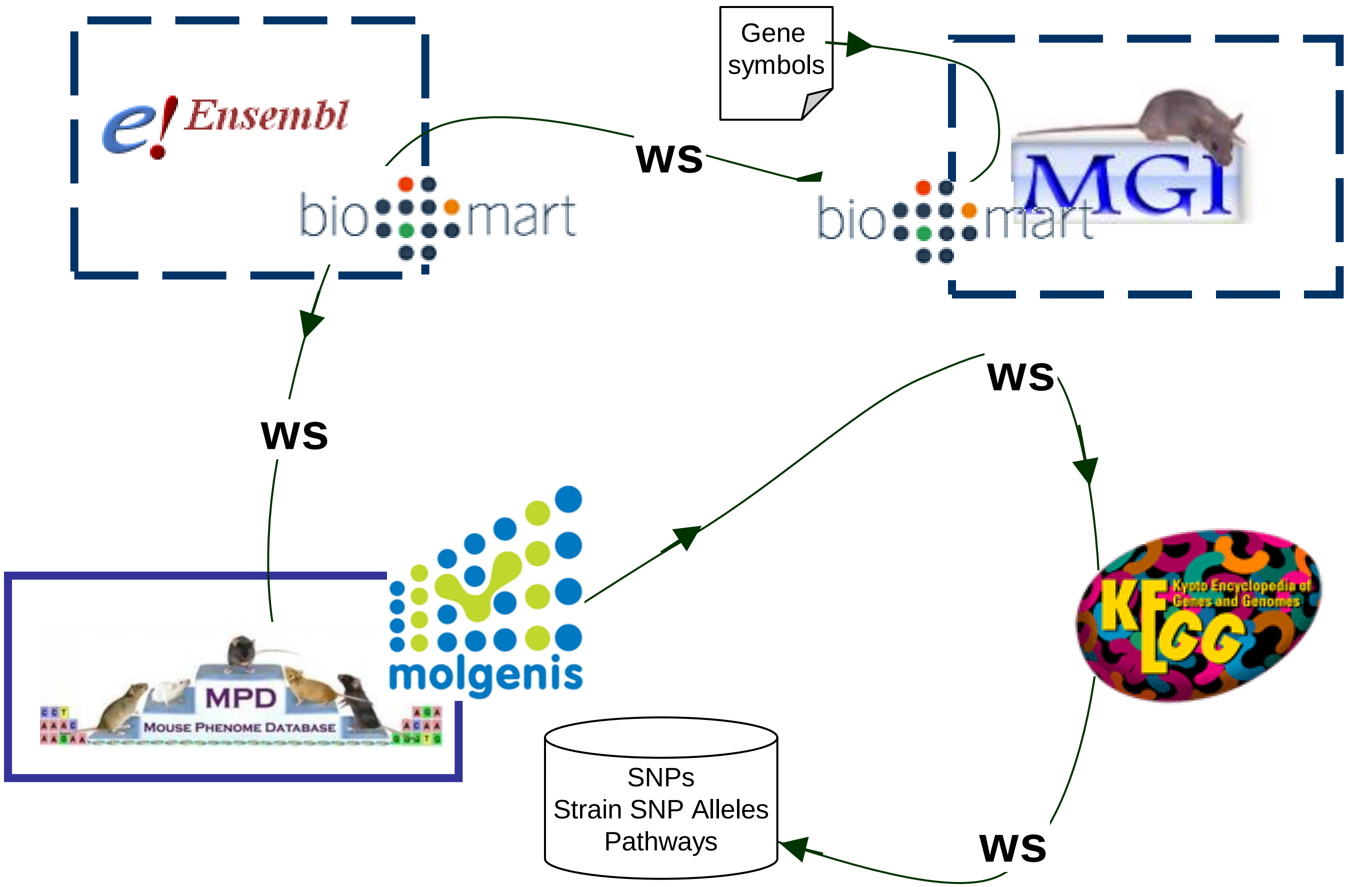
```
#connect to XGAP database
source("http://aserver/xgap/api/R")

#use 'metanetwork' investigation
use.investigation(name="metanetwork")

#list available data sets
find.datasets()

#download genotype and phenotype datasets
geno    <-find.datamatrix(name="geno")
mpheno  <-find.datamatrix(name="mexpr")
markers <-find.markers()

#calculate & plot (Fu 2007, Nature Protocols)
mqtls <-qtlMapTwoPart(geno,mpheno,spike=4)
qtlPlot(markers,mqtls, 4)

#upload qtl result matrix
add.datamatrix(mqtls,
       name="qtlprofiles"
       rowtype="metabolite" coltype="marker"
       valuetype="double")
```

Scientist **B** uploads analysis results

Swertz et al (2010) XGAP: a uniform and extensible data model and software platform for genotype and phenotype experiments. Genome Biology 11(3).

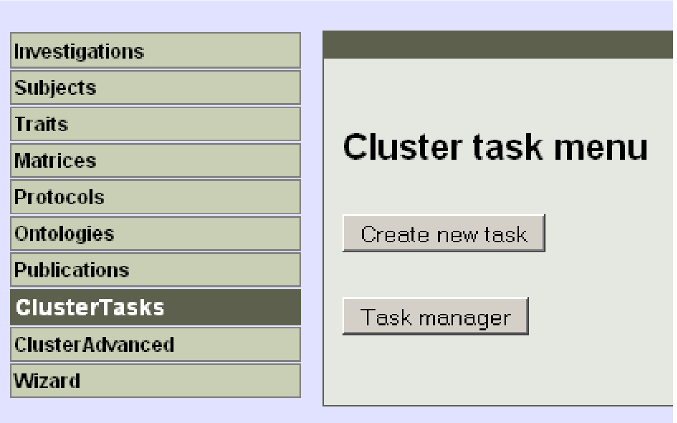
Generated: tool integration interfaces

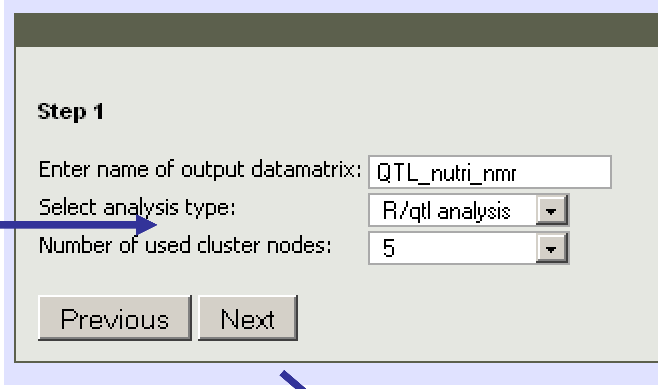
- REST, SOAP, RDF



Smedley, Swertz, Wolstencroft et al (2008) Solutions for data integration in functional genomics: a critical assessment and case study. Briefings in Bioinformatics 9(6)

Plugin: Data analysis using cloud/cluster



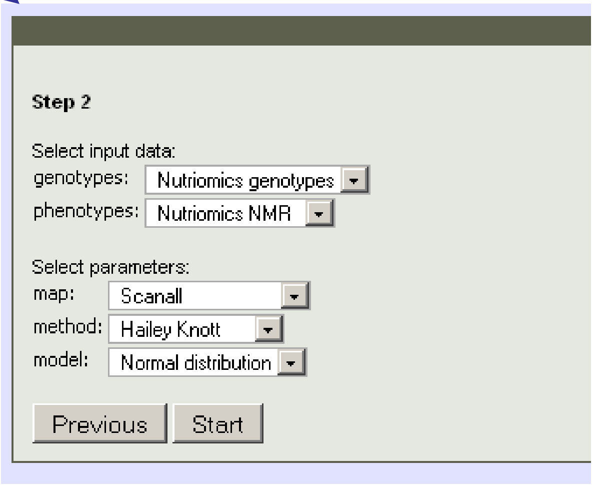


Step 1

Enter name of output datamatrix: QTL_nutri_nmr

Select analysis type: R/qtl analysis

Number of used cluster nodes: 5



Step 2

Select input data:

genotypes: Nutriomics genotypes

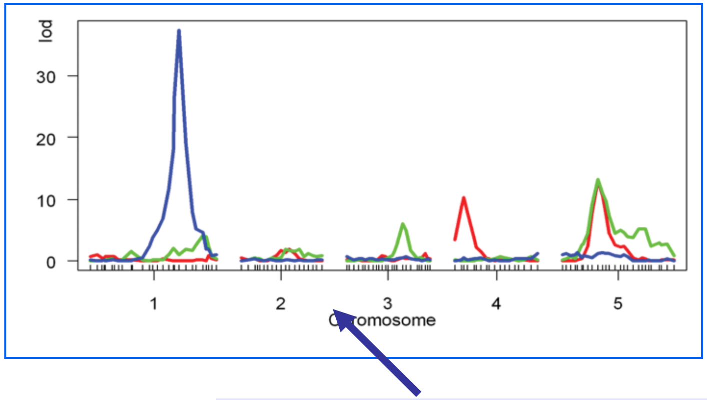
phenotypes: Nutriomics NMR

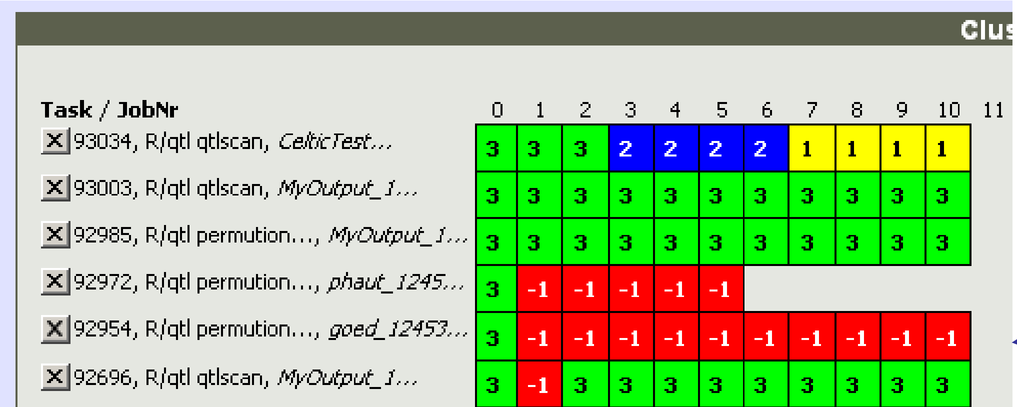
Select parameters:

map: Scanall

method: Hailey Knott

model: Normal distribution





Task / JobNr	0	1	2	3	4	5	6	7	8	9	10	11
☒ 93034, R/qtl qtlscan, *CelticTest...*	3	3	3	2	2	2	2	1	1	1	1	
☒ 93003, R/qtl qtlscan, *MyOutput_1...*	3	3	3	3	3	3	3	3	3	3	3	
☒ 92985, R/qtl permution..., *MyOutput_1...*	3	3	3	3	3	3	3	3	3	3	3	
☒ 92972, R/qtl permution..., *phaut_1245...*	3	-1	-1	-1	-1	-1						
☒ 92954, R/qtl permution..., *goed_12453...*	3	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
☒ 92696, R/qtl qtlscan, *MyOutput_1...*	3	-1	3	3	3	3	3	3	3	3	3	

PANACEA

Discussion & Conclusion

GMOD links?

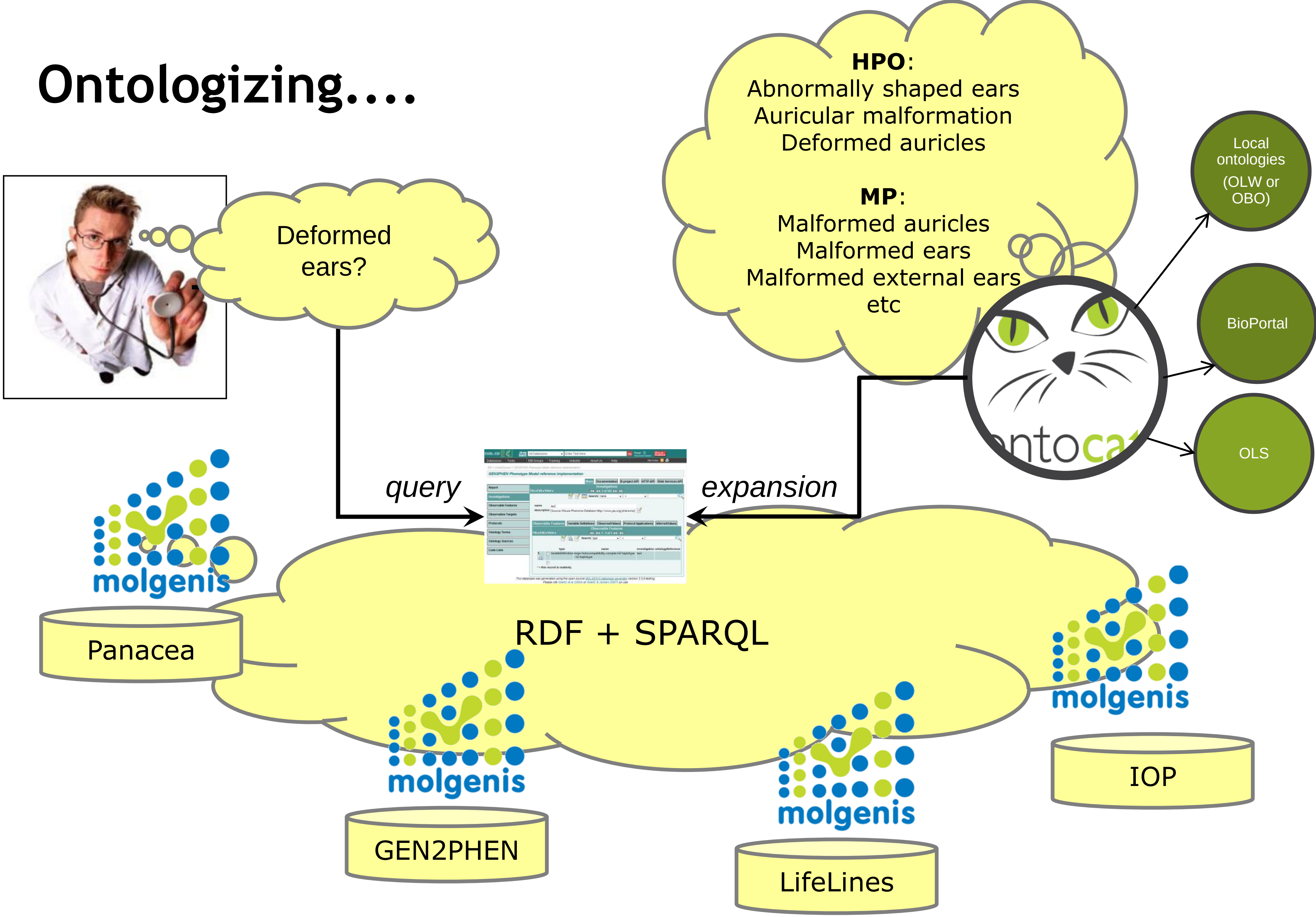
GMOD link ideas

- Chado
  - XGAP harmonization towards Chado?
  - MOLGENIS 4 Chado?

- Gbrowse & DAS
  - Have XGAP data projected on genome browser?
  - Serve XGAP data as custom tracks?

- BioMART/InterMine
  - Consume BioMART data to auto-annotate experimental data?
  - Export XGAP experiments into MART/MINE query environments?

Ontologizing....



Deformed ears?

HPO:
Abnormally shaped ears
Auricular malformation
Deformed auricles

MP:
Malformed auricles
Malformed ears
Malformed external ears
etc

Local ontologies (OLW or OBO)

BioPortal

OLS

ontocat

query

expansion

RDF + SPARQL

Panacea

GEN2PHEN

LifeLines

IOP

GEN2PHEN Phenotype Model reference implementation

name Jax2
description [Source: Mouse Phenome Database http://www.jax.org/phenome]

	type	name	investigation	ontologyReference
1.	VariableDefinition	major histocompatibility complex H2 haplotype - h2 haplotype	Jax2	

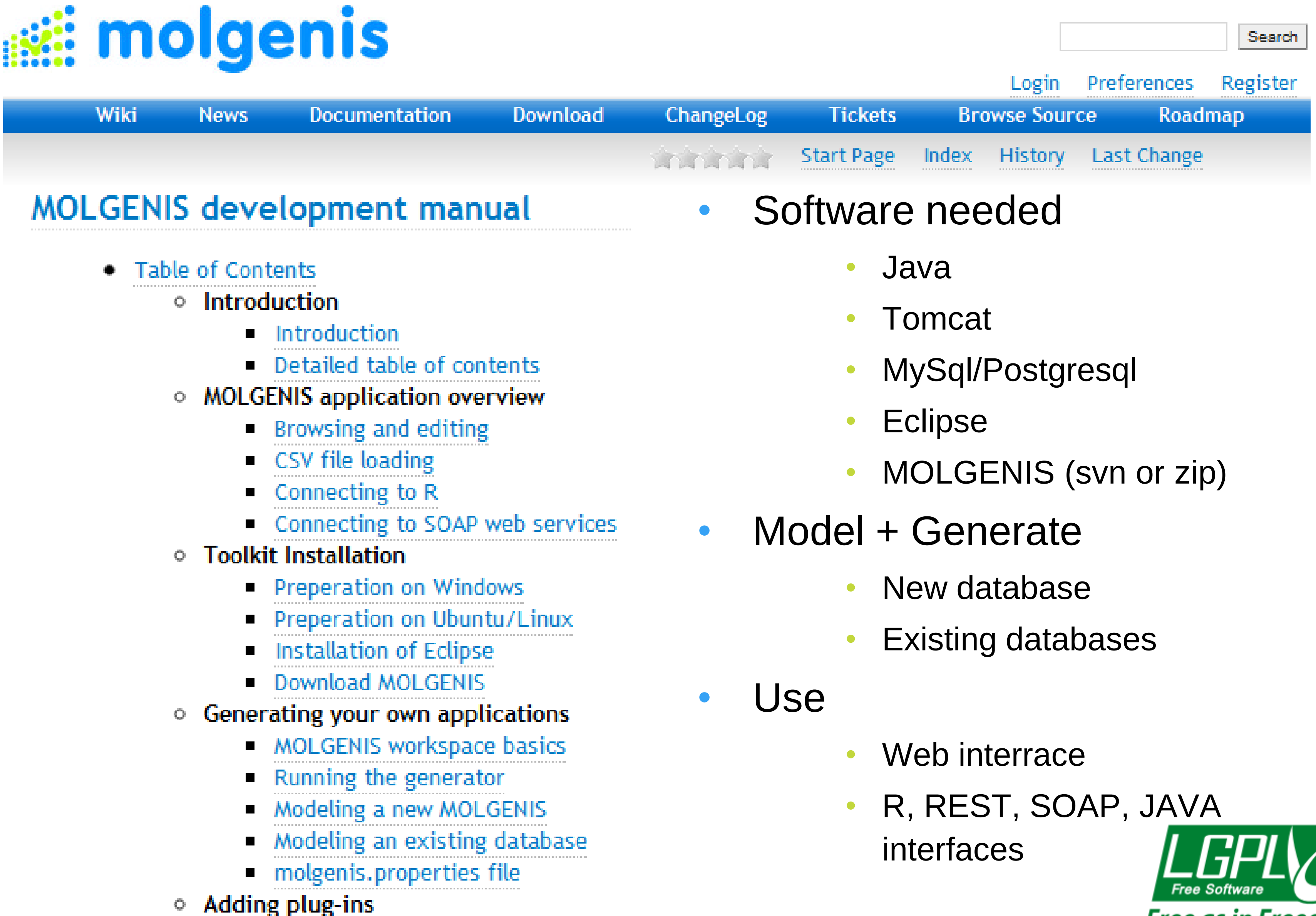
* = this record is readonly.

This database was generated using the open source MOLGENIS database generator version 3.3.0-testing.
Please cite Swertz et al (2004) or Swertz & Jansen (2007) on use

OntoCAT – Ontology common API tasks
http://www.ontocat.org and http://precedings.nature.com/documents/4666

Getting started

http;//www.molgenis.org



molgenis
Search
Login Preferences Register
Wiki News Documentation Download ChangeLog Tickets Browse Source Roadmap
Start Page Index History Last Change

MOLGENIS development manual

- Software needed
  - Java
  - Tomcat
  - MySql/Postgresql
  - Eclipse
  - MOLGENIS (svn or zip)
- Model + Generate
  - New database
  - Existing databases
- Use
  - Web interrace
  - R, REST, SOAP, JAVA interfaces

LGPLv3 Free Software

Acknowledgements

Morris Swertz
Joeri van der Velde
Joris Lops
Danny Arends
Alex Kanterakis
Erik Roos
Richard Scheltema
Martijn Dijkstra
Rudi Alberts
Bruno M. Tesson
Gonzalo Vera Rodriguez
Tomasz Adamuziak
Juha Muilu
Gudmundur Thorisson
Damian Smedley
Katy Wolstencroft

Ritsert C. Jansen
Cisca Wijmenga
Carole Goble
John M. Hancock
Andrew R. Jones
Klaus Schughart
Paul Schofield
Anthony Brookes
Helen E. Parkinson

BBMRI-NL biobanking (Hs)
EU-GEN2PHEN consortium (Hs)
EU-PANACEA consortium (Ce)
NL Brassica Nutr. consortium (At)
EU-CASIMIR consortium (Mm)
NBIC/BioAssist consortium (bioinfo)

Thank you!
Questions?

m.a.swertz@rug.nl
k.j.van.der.velde@rug.nl

Web

- MOLGENIS: http://www.molgenis.org
- XGAP: http://www.xgap.org
- OntoCAT: http://www.ontocat.org

molgenis.org
Your database at the push of a button

Pubmed

- Swertz et al (2010) *Genome Biology* 9;11(3): R27.
- Smedley et al (2008) *Briefings in Bioinformatics* 9(6): 532-544
- Swertz & Jansen (2007) *Nature Reviews Genetics* 8, 235-243
- Swertz et al (2004) *Bioinformatics* 20(13), 2075-83