TARGET Conference 2013
Probing Big Data for answers
3 - 5 April, 2013
Groningen, Netherlands

Molgenis, life science databases at the push of a button

Dr. Martijn Dijkstra

K. Joeri van der Velde, Morris Swertz,
members of the Genomics Coordination Center

umcg university of groningen genomics coordination center

Outline

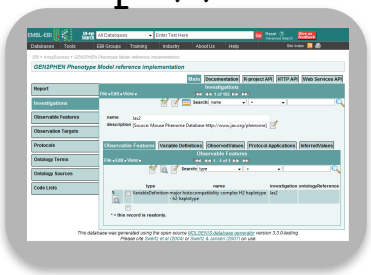
- Molgenis Overview
- Demo Movie
  - model
  - generate
  - use
- Example application
- Analysis of big data
- Summary

molgenis.org

Your database at the push of a button

1. *model in XML*
2. *generate*
3. *use*

http://…



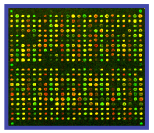
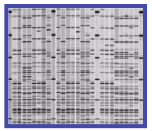
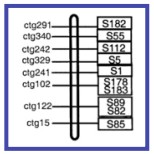
Overview

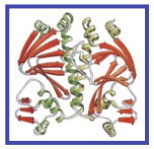
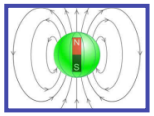
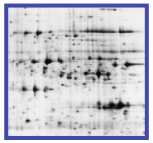
Challenge:





Species...

Experiments...

GEN2PHEN

university of groningen

FIMM

LIFELINES

SYSGENET

CASIMIR

nbioassist

WAGENINGEN UR
For quality of life

EMBL-EBI

umcg

IOP

PANACEA

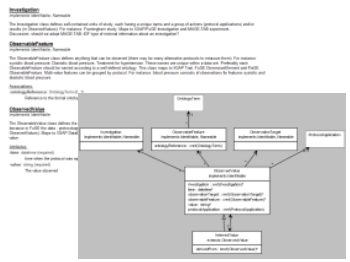
BBMRI • NL

University of Leicester

Projects ...

Needed:

❶ **Data models & Protocols**



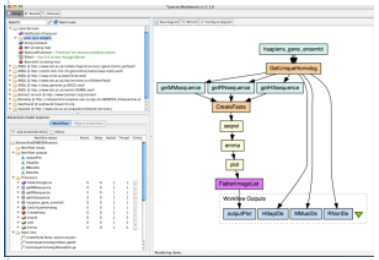
❹ **Connect to statistics**

```
find.investigation()
102 downloaded

obs<-find.observedvalue(
43,920 downloaded

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

❺ **Connect to annotation pipelines**

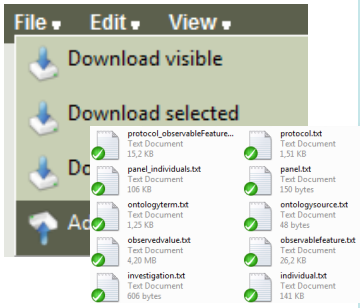


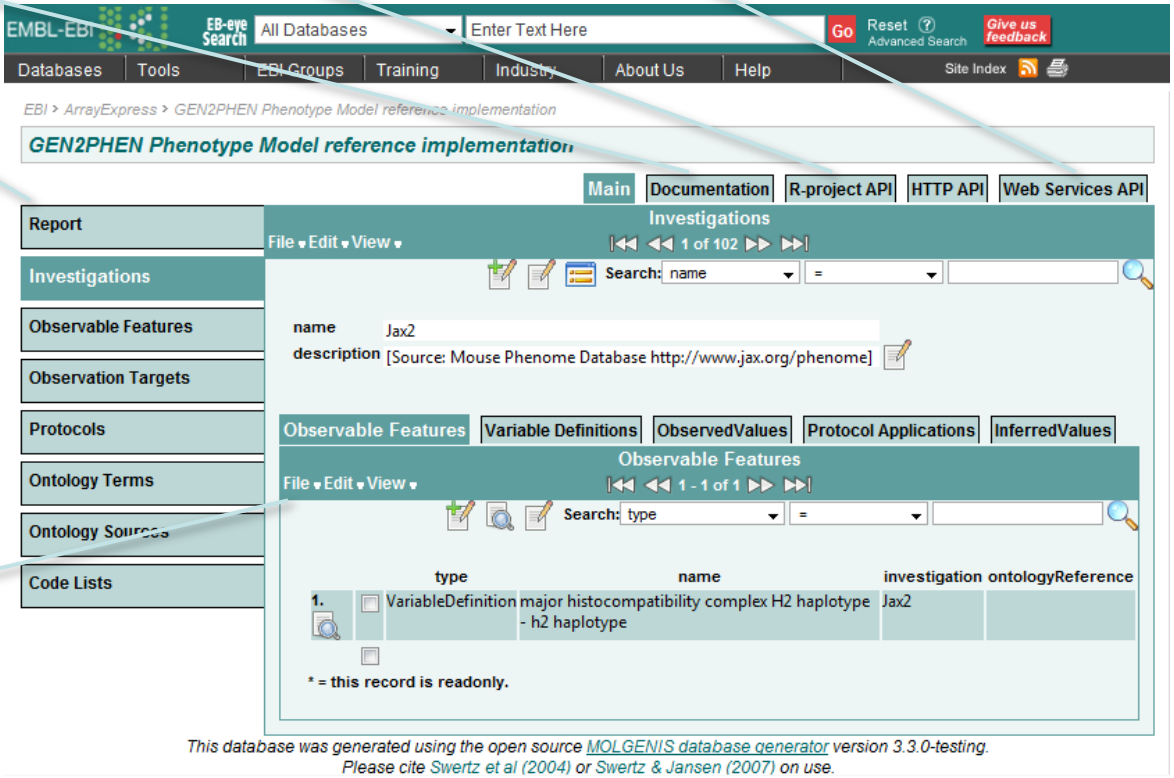
❻ **Plugin rich analysis tools**

❷ **Edit & trace your data**

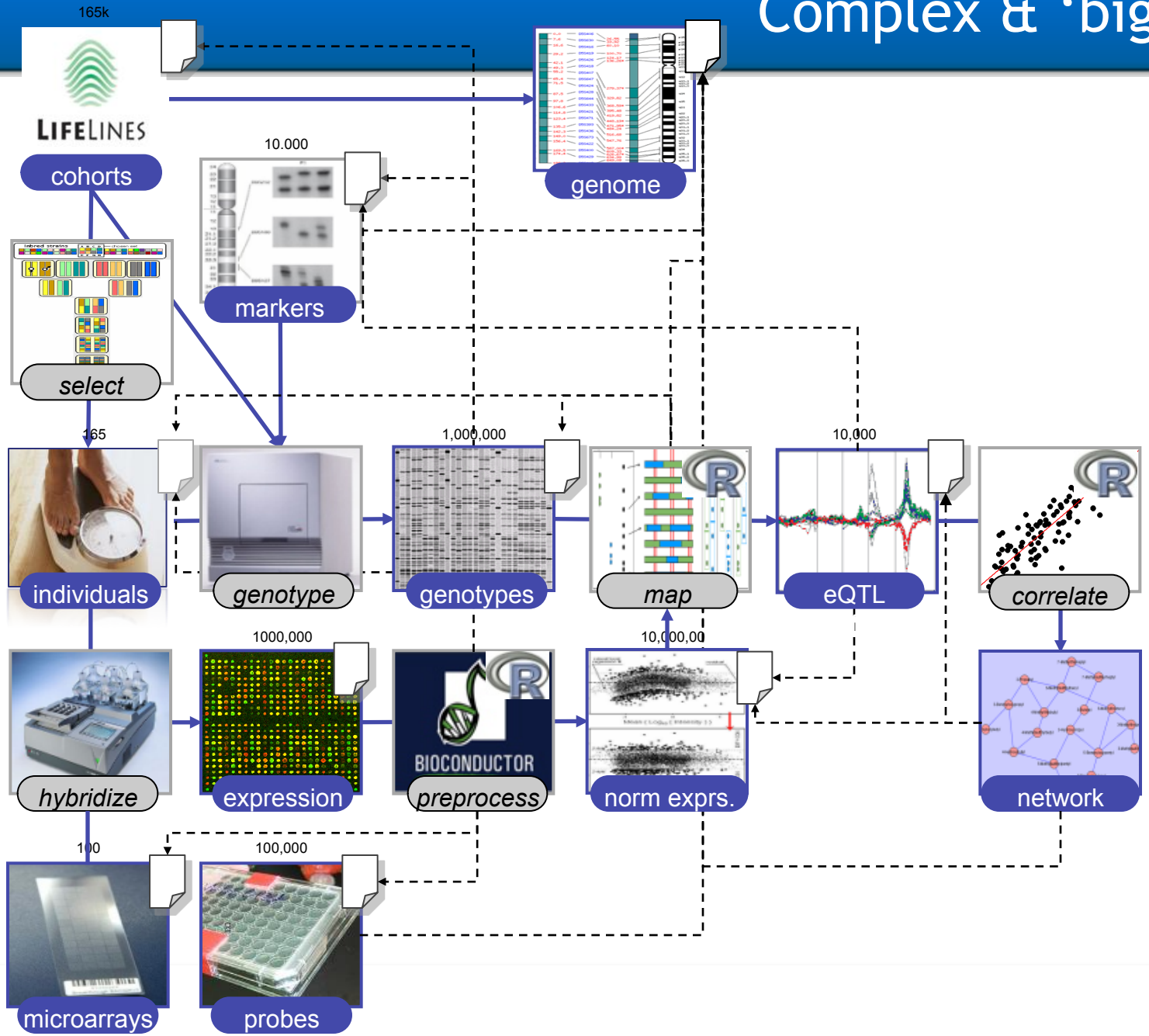
- Report
- Investigations
- Observable Features
- Observation Targets
- Protocols
- Ontology Terms

❸ **Import/export**

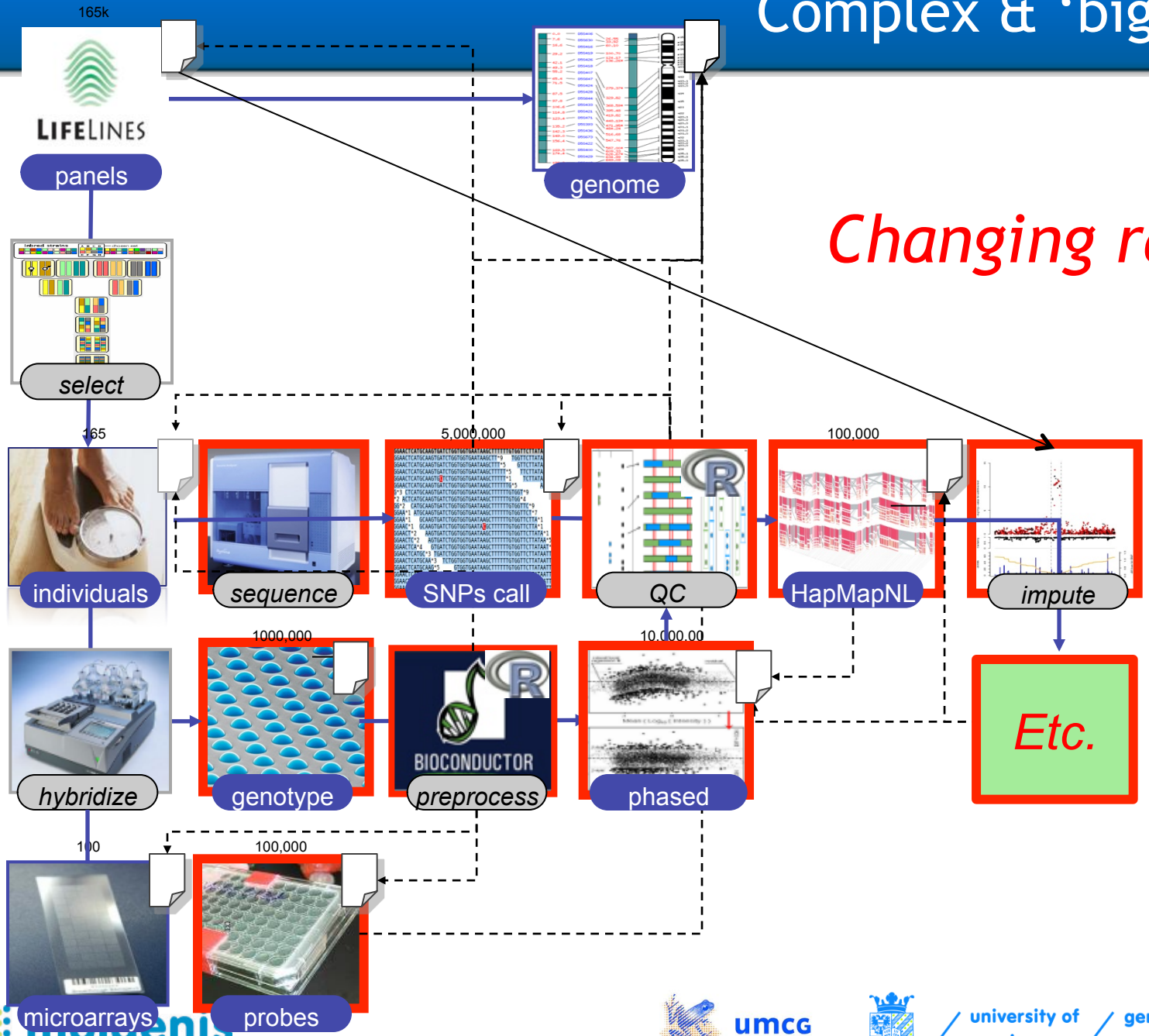




Complex & 'big data'

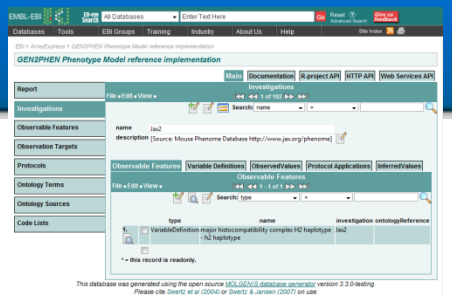


Complex & 'big data'

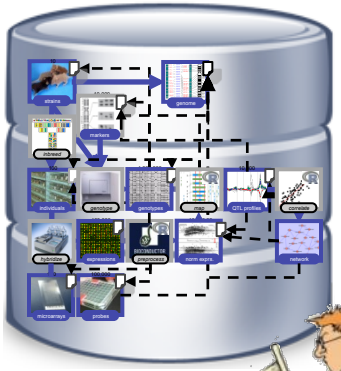


Changing rapidly

One change → many places



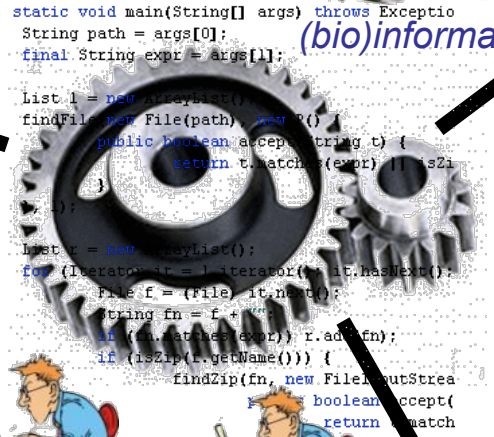
biologist



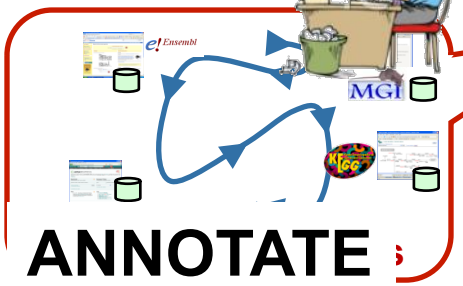
STORE

Logic

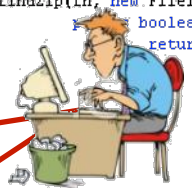
(bio)informatician



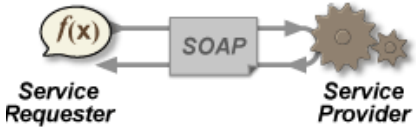
GUI



ANNOTATE



ANALYSE



bioinformatician

molgenis umcg university of groningen genomics coordination center

Model what is desired ...

Model

NextGenSeq

Describe

Model

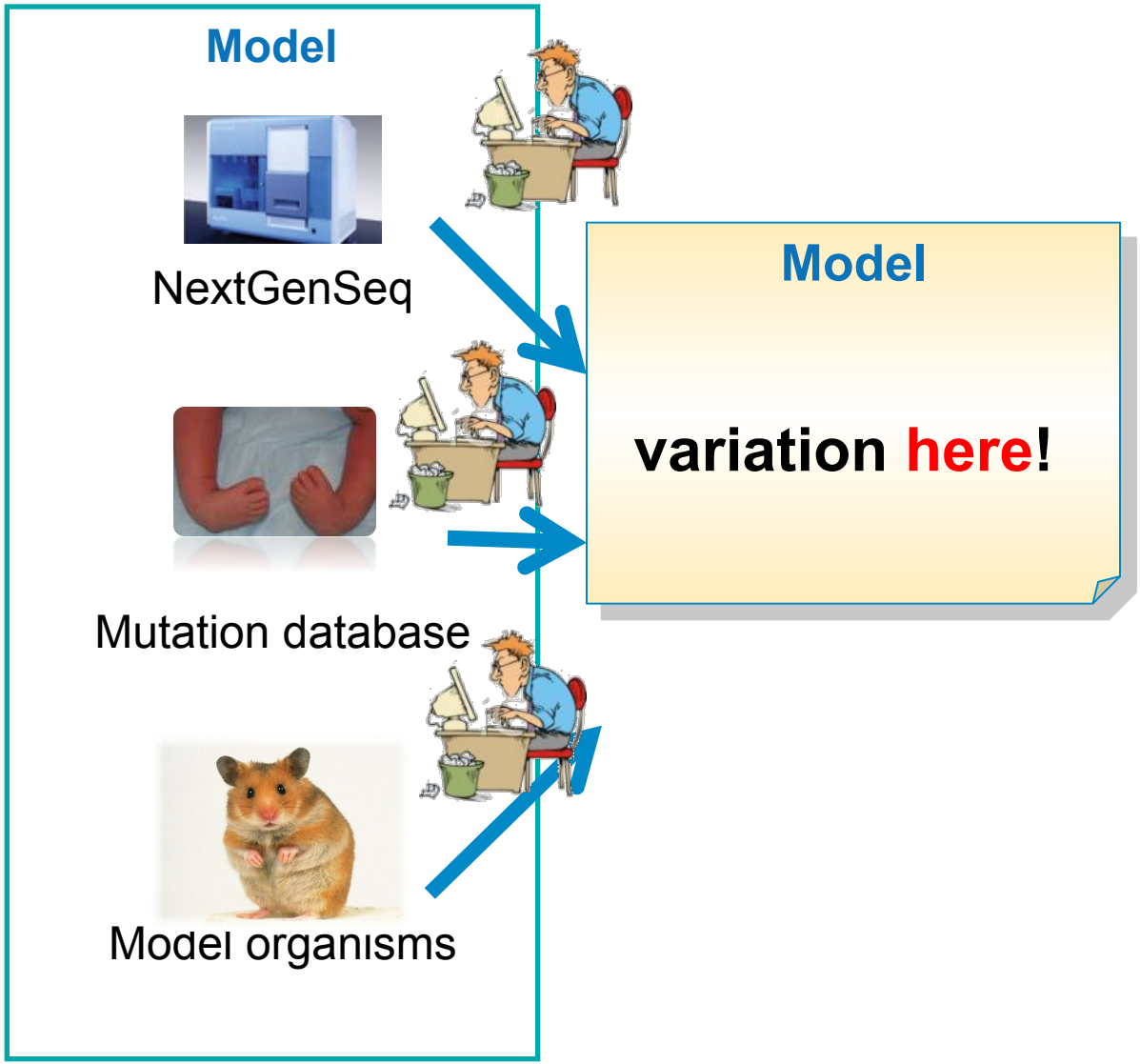
variation here!

http://www.molgenis.org

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

Swertz, Dijkstra, vd Velde et al. (2010) BMC Bioinformatics 11(suppl 12):s12

Model what is desired ...



http://www.molgenis.org
Swertz & Jansen (2007) Nature Reviews Genetics 8, 235-243
Swertz, Dijkstra, vd Velde et al. (2010) BMC Bioinformatics 11(suppl 12):s12

Autogenerate the software

Model in DSL

NextGenSeq

Mutation database

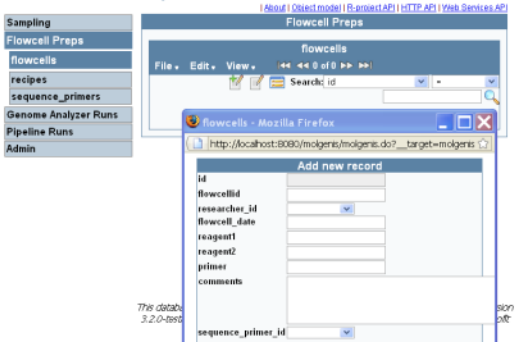
Model organisms

Generator

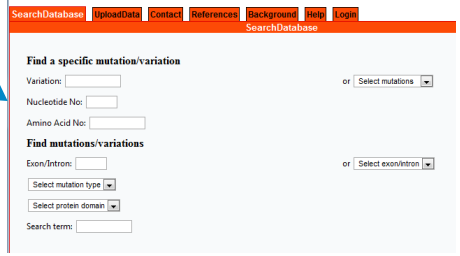
molgenis

Use generated software

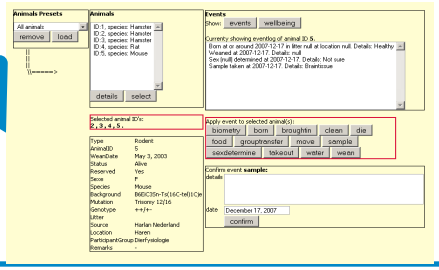
Solexa Sequencer LIMS



database of COL7A1 mutations



Animal Observatory



http://www.molgenis.org

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

Swertz, Dijkstra, vd Velde et al. (2010) BMC Bioinformatics 11(suppl 12):s12

Autogenerate the software

Model in DSL

NextGenSeq

Mutation database

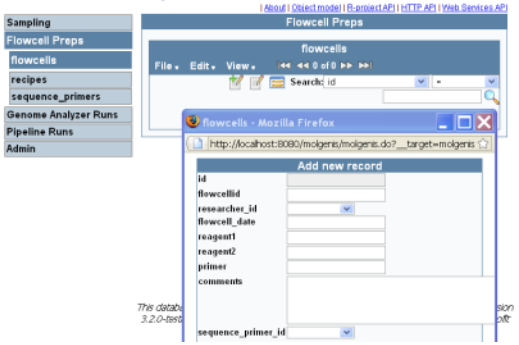
Model organisms

Generator

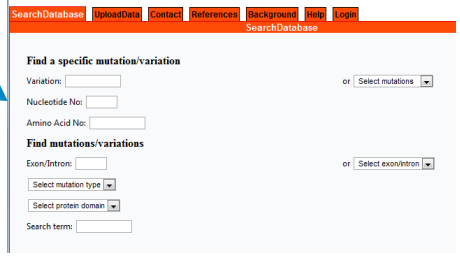
molgenis

Use generated software

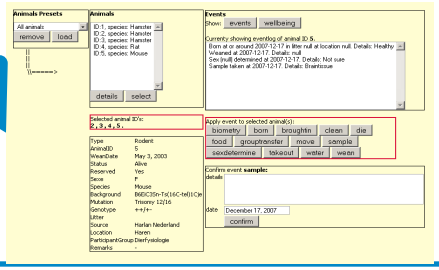
Solexa Sequencer LIMS



database of COL7A1 mutations



Animal Observatory



repeat often

http://www.molgenis.org

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

Swertz, Dijkstra, vd Velde et al. (2010) BMC Bioinformatics 11(suppl 12):s12

Reuse the improvements

Model in DSL

NextGenSeq

Mutation database

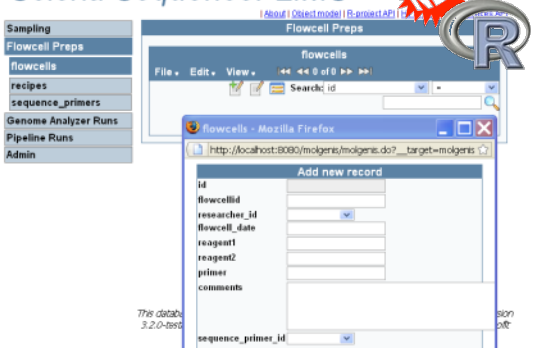
Model organisms

Generator

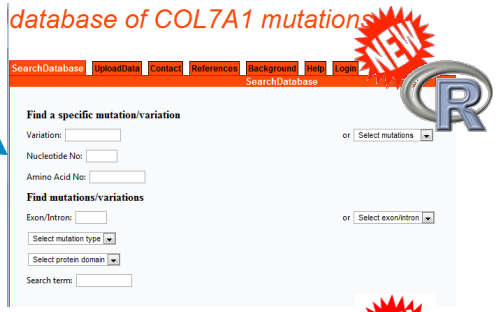
molgenis

Use generated software

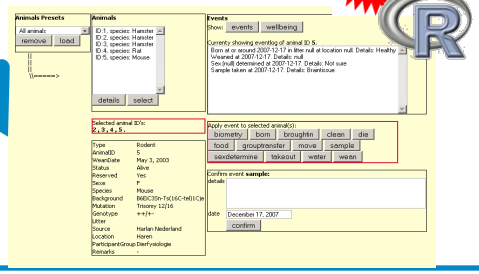
Solexa Sequencer LIMS



database of COL7A1 mutations



Animal Observatory



repeat often

http://www.molgenis.org

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

Swertz, Dijkstra, vd Velde et al. (2010) BMC Bioinformatics 11(suppl 12):s12

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 hinde
of suitable software infrastructures. Biologists need in
connects to work that is done in other laboratories, for
is helpful. However, the infrastructure must also accom
their biological system, but appropriate mechanism
are currently lacking. We argue that a minimal com
software tool called a generator, can be used to quickl
software infrastructures that 'systems biologists really

molgenis

News | Documentation | Download | ChangeLog

wiki: WikiStart

ShareThis 41

Welcome to MOLGENIS

MOLGENIS is an collaborative open source project on a mission to generate great software infrastructure for life science research. Each app in the MOLGENIS family comes with rich data management interface and plug-in integration of analysis tools in R, Java and web services.

http://www.molgenis.org

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

Swertz, Dijkstra, vd Velde et al. (2010) BMC Bioinformatics 11(suppl 12):s12

Demo

http://www.molgenis.org/wiki/MolgenisDownload

Example application

Voedingsdagboek.nl

VOEDINGSDAGBOEK - Gez

voedingsdagboek.nl

VOEDINGSDAGBOEK
GEZOND SLANK MET JE EIGEN VOEDING

Check je gezondheid | Voeding | Log in | Over ons

Feedback

Wil jij gezond slank zijn?

Dan is VOEDINGSDAGBOEK er voor jou! Je kunt hier bijhouden wat je eet. Wij geven je daarna inzicht in je voeding. Bovendien geven we je dieettips, afgestemd op je persoonlijke situatie. Zo helpen we je met gezond eten (blijvend!) naar een gezond gewicht. Dit noemen wij "*Het nieuwe afvallen*".

Begin meteen »

Je bepaalt zelf

... wat je eet. Dat is *het nieuwe afvallen*. Houd je helemaal niet van eierkoeken? Wil je niet allerlei speciale ingrediënten in huis halen voor recepten die je eigenlijk helemaal niet lekker vindt? Bij VOEDINGSDAGBOEK hoeft dat ook helemaal niet! Je eet gewoon wat je normaal ook eet en krijgt op basis van dat eten een op maat gesneden advies. Door kleine aanpassingen in je eigen voedingspatroon zul je merken dat je snel gewicht verliest en je gezonder gaat voelen! **Yes, you can!**

Gezond en blijvend afvallen

Een crash dieet van 1200 calorieën per dag lijkt snel tot resultaat te leiden. Maar schijn bedriegt, want al heel snel komt het jojo-effect om de hoek kijken. Om je streefgewicht te bereiken ?n te behouden is er wat anders nodig, namelijk een gezonde voeding. Een gezonde voeding bevat alle voedingsstoffen die je nodig hebt, maar houdt ook in dat je van niets te veel binnenkrijgt. Zo val je effectief af en houd je dit langdurig vol!

Wetenschappelijk bewijs

Uit onderzoek blijkt dat je sneller afvalt als je een VOEDINGSDAGBOEK gebruikt.

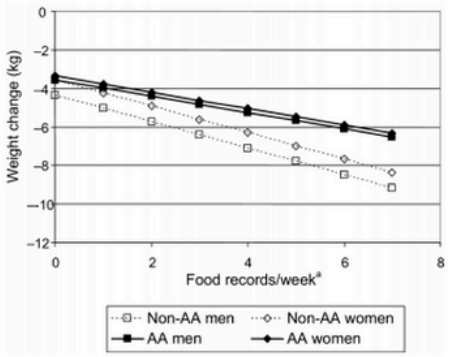


Figure 4. Estimated effect of number of food records kept per week on weight change, by gender and race [a] Evaluated at

Aangesloten

- Diëtisten
- Restaurants
- Sportscholen

Wil je gezond uit eten? Afvallen in een sportschool? Of onder begeleiding van een diëtist? De bij ons aangesloten diëtisten, restaurants en sportscholen maken dat mogelijk. Door een directe koppeling met onze database is het invoeren van gegevens erg eenvoudig.

molgenis | umcg | university of groningen | genomics coordination center

Voedingsdagboek.nl

- Model / lines of code
  - db.xml / 100 lines
  - ui.xml / 25 lines

```xml
<?xml version="1.0" encoding="UTF-8"?>
<molgenis>
    <module name="demo">
        <description>This is a demonstration module</description>
        <entity name="Experiment">
            <description>My Experiments</description>
            <field name="Id" type="autoid" description="autogenerated unique id
            <field name="Name" unique="true" description="Unique name" />
            <field name="Description" nillable="true" description="Optional des
        </entity>
        <entity name="Sample">
            <description>My Samples</description>
            <field name="Id" type="autoid" description="autogenerated unique id
            <field name="Name" unique="true" description="Unique name" />
            <field name="Experiment" type="xref" xref_field="Experiment.Id"
                xref_label="Name" />
            <field name="Species" type="enum" enum_options="[hs,mm]" />
            <field name="Created" type="date" auto="true" description="automatic date" />
        </entity>
        <entity name="Data">
            <field name="Id" type="autoid" description="autogenerated unique id" />
            <field name="Name" unique="true" description="Unique name" />
            <field name="Experiment" type="xref" xref_field="Experiment.Id"
                xref_label="Name" />
            <field name="Samples" type="mref" xref_field="Sample.Id" xref_label="Name"
                description="One or more samples" />
            <field name="Data" type="file" />
        </entity>
    </module>
</molgenis>
```

```xml
<molgenis name="example">
    <plugin name="molgenis_header"
        type="plugins.header.MolgenisHeader" />
    <menu name="Main">
        <form name="Experiments" entity="Experiment" sortby="Name">
            <menu name="ExperimentMenu">
                <form name="Samples" entity="Sample" />
                <form name="Data" entity="Data" />
            </menu>
        </form>
        <plugin name="ImportWizard"
            type="plugins.genericwizard.GenericWizard" />
        <plugin name="HelloWorld"
            type="plugins.hello.World"/>
    </menu>
</molgenis>
```

Voedingsdagboek.nl

- Model / lines of code
  - db.xml / 100 lines
  - ui.xml / 25 lines
- Generate
  - *.sql / 1722 lines
  - *.java / 46639 lines

```xml
<?xml version="1.0" encoding="UTF-8"?>
<molgenis>
    <module name="demo">
        <description>This is a demonstration module</description>
        <entity name="Experiment">
            <description>My Experiments</description>
            <field name="Id" type="autoid" description="autogenerated unique id
            <field name="Name" unique="true" description="Unique name" />
            <field name="Description" nillable="true" description="Optional des
        </entity>
        <entity name="Sample">
            <description>My Samples</description>
            <field name="Id" type="autoid" description="autogenerated unique id
            <field name="Name" unique="true" description="Unique name" />
            <field name="Experiment" type="xref" xref_field="Experiment.Id"
                xref_label="Name" />
            <field name="Species" type="enum" enum_options="[hs,mm]" />
            <field name="Created" type="date" auto="true" description="automatic date" />
        </entity>
        <entity name="Data">
            <field name="Id" type="autoid" descri
            <field name="Name" unique="true" desc
            <field name="Experiment" type="xref"
                xref_label="Name" />
            <field name="Samples" type="mref" xre
                description="One or more samples"
            <field name="Data" type="file" />
        </entity>
    </module>
</molgenis>
```

```xml
<molgenis name="example">
    <plugin name="molgenis_header"
        type="plugins.header.MolgenisHeader" />
    <menu name="Main">
        <form name="Experiments" entity="Experiment" sortby="Name">
            <menu name="ExperimentMenu">
                <form name="Samples" entity="Sample" />
                <form name="Data" entity="Data" />
            </menu>
        </form>
        <plugin name="ImportWizard"
            type="plugins.genericwizard.GenericWizard" />
        <plugin name="HelloWorld"
            type="plugins.hello.World"/>
    </menu>
</molgenis>
```

molgenis umcg university of groningen genomics coordination center

Voedingsdagboek.nl

- Model / lines of code
  - db.xml / 100 lines
  - ui.xml / 25 lines
- Generate
  - *.sql / 1722 lines
  - *.java / 46639 lines

```xml
<?xml version="1.0" encoding="UTF-8"?>
<molgenis>
    <module name="demo">
        <description>This is a demonstration module</description>
        <entity name="Experiment">
            <description>My Experiments</description>
            <field name="Id" type="autoid" description="autogenerated unique id
            <field name="Name" unique="true" description="Unique name" />
            <field name="Description" nillable="true" description="Optional des
        </entity>
        <entity name="Sample">
            <description>My Samples</description>
            <field name="Id" type="autoid" description="autogenerated unique id
            <field name="Name" unique="true" description="Unique name" />
            <field name="Experiment" type="xref" xref_field="Experiment.Id"
                xref_label="Name" />
            <field name="Species" type="enum" enum_options="[hs,mm]" />
            <field name="Created" type="date" auto="true" description="automatic date" />
        </entity>
        <entity name="Data">
            <field name="Id" type="autoid" descri
            <field name="Name" unique="true" desc
            <field name="Experiment" type="xref"
                xref_label="Name" />
            <field name="Samples" type="mref" xre
                description="One or more samples"
            <field name="Data" type="file" />
        </entity>
    </module>
</molgenis>
```

```xml
<molgenis name="example">
    <plugin name="molgenis_header"
        type="plugins.header.MolgenisHeader" />
    <menu name="Main">
        <form name="Experiments" entity="Experiment" sortby="Name">
            <menu name="ExperimentMenu">
                <form name="Samples" entity="Sample" />
                <form name="Data" entity="Data" />
            </menu>
        </form>
        <plugin name="ImportWizard"
            type="plugins.genericwizard.GenericWizard" />
        <plugin name="HelloWorld"
            type="plugins.hello.World"/>
    </menu>
</molgenis>
```

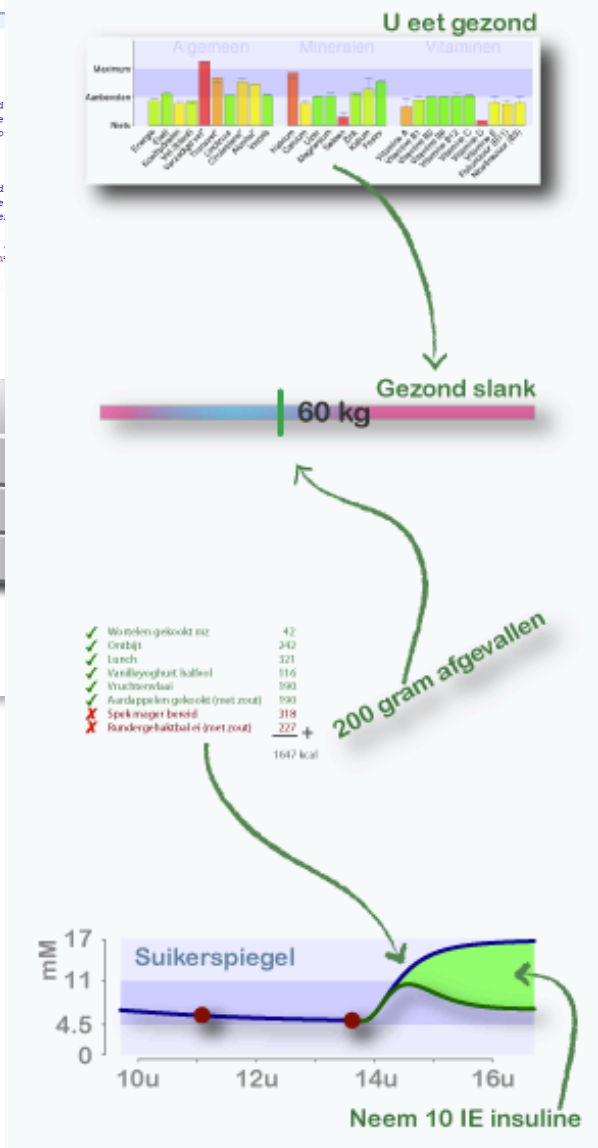
1 : 400!

molgenis umcg university of groningen genomics coordination center

Voedingsdagboek.nl

- Model / lines of code
  - db.xml / 100 lines
  - ui.xml / 25 lines
- Generate
  - *.sql / 1722 lines
  - *.java / 46639 lines
- 5 custom plugins / 3682 lines
- Use

```
<?xml version="1.0" encoding="UTF-8"?>
<molgenis>
    <module name="demo">
        <description>This is a demonstration module</description>
        <entity name="Experiment">
            <description>My Experiments</description>
            <field name="Id" type="autoid" description="autogenerated
            <field name="Name" unique="true" description="Unique name
            <field name="Description" nillable="true" description="Op
        </entity>
        <entity name="Sample">
            <description>My Samples</description>
            <field name="Id" type="autoid" description="autogenerated
            <field name="Name" unique="true" description="Unique name
            <field name="Experiment" type="xref" xref_field="Experime
                xref_label="Name" />
            <field name="Species" type="enum" enum_options="[hs,mm]"
            <field name="Created" type="date" auto="true" description
        </entity>
        <entity name="Data">
            <field name="Id" type="autoid" descri
            <field name="Name" unique="true" desc
            <field name="Experiment" type="xref"
                xref_label="Name" />
            <field name="Samples" type="mref" xre
                description="One or more samples"
            <field name="Data" type="file" />
        </entity>
    </module>
</molgenis>
```



1 : 400!

Analysis of big data

Challenge: big data and huge analyses

- Challenge 1
  - Workflows with dozens of analysis scripts
  - execute each script ~100x for different parameter values
  - 1000 – 10,000 scripts in total per workflow
- Challenge 2
  - Running on high-performance computational infrastructures
    - (Target) computer clusters, national GRID, …
    - Different scheduling systems: *PBS*, …
    - Different file management: *SRM*, …
    - Different tool management: *module system*, …
- Challenge 3
  - Monitoring execution, log files, data provenance

Solution: Molgenis Compute – huge analyses straight from db

Solution: Molgenis Compute – huge analyses straight from db

- Solution 1
  - Input: Few files
    - 1 workflow.csv
    - 1 parameters.csv
    - *n* scriptTemplates.sh
  - Output: Generate 1000s of scripts
- Solution 2: transparent execution
  - *resource management*
  - *data management*
  - *tool management*
- Solution 3
  - Monitoring execution, log files, data provenance
- **Scientist can focus on the analysis only**

molgenis compute

Closing remarks

All available as open source

Fork me on GitHub

github | Search or Type a Command | Explore Gist Blog Help | mswertz

biosoftware platform

molgenis

The Netherlands, Europe
http://www.molgenis.org
Joined on Apr 28, 2012

4 public repos **0** private repos **23** members

Repositories | Members | Edit molgenis's Profile

Find a Repository... | All Public Private Sources Forks Mirrors

molgenis_apps — Java ★ 25 ⑂ 27
MOLGENIS Advanced Application and Computation Framework for the Life Sciences
Last updated 14 hours ago

molgenis — Java ★ 27 ⑂ 26
The MOLGENIS Software generator tool for creating Dynamic Software Infrastructure used in the Life Sciences
Last updated 17 hours ago

molgenis_test — Java ★ 7 ⑂ 9
The MOLGENIS Test suite, fork this if you plan to contribute to MOLGENIS apps, so you can test commits before sending them in
Last updated 7 days ago

molgenis_distro — JavaScript ★ 5 ⑂ 4
The empty MOLGENIS Distribution, fork this and create your own MOLGENIS web application
Last updated a month ago

http://github.com/molgenis
http://www.molgenis.org/

Capture most variation in Life Sciences in one model

- Molgenis
  - Model
  - Generate
  - Use

Acknowledgements

Joeri van der Velde
Despoina Antonakaki
Tomasz Adamuziak
Rob Hastings
Sirisha Gollapudi
Gudmundur Thorisson
Chao Pang
Myles Byrne
David van Enckvort
Linda Mook
Pieter Neerincx
Ger Strikwerda
Danny Arends
Roan Kanninga
Jan Bot
George Byelas
Yang Li
Basten Snoek
Noortje Festen
Martijn Dijkstra

And more …

Lude Franke
Juha Muilu
Anthony Brookes
Helen Parkinson
Vincent Ferreti
Gert-Jan van Ommen
Jan Jurjen Uitterdijk
Ritsert C. Jansen
Jan Kammenga
Cisca Wijmenga
Paul de Bakker
Irene Nooren
Rob Hooft
Salome Scholtens
Hans Hillege
Ronald Stolk
Morris Swertz

And more…

NBIC/BioAssist consortium (bioinfo)
BBMRI-NL catalogue group(Hs)
CTMM/TraIT consortium (Hs)
EU-GEN2PHEN consortium (Hs)
EU-PANACEA consortium (Ce)
EU-BioSHARE consoritum (Hs)
EU-CASIMIR consortium (Mm)
EU-BioMedBridges cosortium (all)
NL Brassica Nutr. consortium (At)
Learning from Nature (At)
LifeLines (Hs)
TIFN (Hs)
BigGrid (info)
Target + CIT (info)

And more …

Questions?

Thank you!
Questions?
m.dijkstra.work@gmail.com

Summary

- Dynamic database
- Rich web GUI
- Interfaces for analysis
- Integrated large scale analyses

Read more

- MOLGENIS: http://www.molgenis.org
- MOLGENIS Compute: http://www.molgenis.org/wiki/ComputeStart
- xQTL: http://www.xqtl.org
- Adamusiak *et al* (2011) *BMC Bioinformatics*
- Akker *et al* (2011) *Human Mutation*
- Arends *et al* (2010) *Bioinformatics* 26: 2990-2992
- Brandsma *et al*, Norsk Epidemiologi 2012
- Snoeks *et al* (2013) *Nucleic Acids Res*
- Swertz *et al* (2010) *Genome Biology* 9;11(3): R27.
- Smedley *et al* (2008) *Briefings in bioinformatics* 9(6):532-44.
- Swertz & Jansen (2007) *Nature Reviews Genetics* 8, 235-243

molgenis.org
Your database at the push of a button

molgenis umcg university of groningen genomics coordination center

Genomics Coordination Center, UMCG, Groningen

Diagnostics & clinic

Biobanking

Core facility of~15 programmers, postdocs and PhD students on a mission to research, develop and support high throughput life sciences with database and analysis e-infrastructures.

Large scale multi-omics

Human and model organism research: genotype 2 phenotype