MOLGENIS by example: generating an extensible platform for genotype and phenotype experiments.

Morris A. Swertz[1,2] (m.a.swertz@rug.nl), K Joeri van der Velde[2], Rudi Alberts[3], Bruno M. Tesson[2], Richard A. Scheltema[2], Gonzalo Vera[2], Damian Smedley[6], Katy Wolstencroft[7], Paul Schofield[4], Klaus Schughart[3], John M. Hancock[5], Engbert O. de Brock[2], Andrew R. Jones[8], Helen E. Parkinson[6], Ritsert C. Jansen[2] and members of the EU-CASIMIR consortium for mouse, EU-GEN2PHEN consortium for human, EU-Panacea consortium for C. elegans and NL-NWO QTL Express consortium for A. thaliana.

[1]Department of Genetics, University Medical Center Groningen, P.O. Box 30.001, 9700 RB Groningen, The Netherlands; [2]University of Groningen, The Netherlands; [3] Helmholtz Centre for Infection Research, Braunschweig, Germany; [4]MRC Harwell, United Kingdom; [5]University of Cambridge, United Kingdom; [6]European Bioinformatics Institute, United Kingdom; [7]University of Manchester, United Kingdom [8]University of Liverpool, United Kingdom.

Software and source: http://molgenis.sourceforge.net and http://www.xgap.org/
Licence: LGPLv3.

MOLGENIS is a general toolbox to auto-generate complete database software from compact models, including user and programmatic interfaces. Here we describe its use to generate XGAP, an eXtensible software platform for high-throughput Genotype And Phenotype experiments:

A growing array of genetic, transcript, protein and metabolite profiling technologies, natural and model organism populations, and GWA, GWL and mutagenesis experimental designs are at our disposal with each different strengths and weaknesses. Integrated analysis of all these genotypes and (molecular) phenotypes provides new opportunities to unravel the genome-to-phenome trajectory. However, the data formats and tool interfaces used by each experiment often also differ, impeding comparison, exchange and integration of data and tools within and between experiments, laboratories and consortia.

To mold existing and new data sets and analysis tools into a singular medium we developed an open software platform named XGAP benefitting both extremes on the user spectrum: experimentalists and computational researchers. To quickly generate a uniform looking software and ease extension with new profiling technologies and methods, we used the open source MOLGENIS software platform. In MOLGENIS, most parts of the software infrastructure can be blueprinted in a compact XML model file that is automatically converted into the many Java, SQL and R code files needed via generator templates written in Freemarker.

We blueprinted a minimal core data model building on standards such as FuGE, MAGE-TAB and OBO and added profiling method specific extensions. Then we auto-generated the basic platform having (i) an easy to create delimited file format to load and exchange data, (ii) a suitable database for storage and querying, (iii) programmatic interfaces to java, web services and the statistical R tools for computational researchers to plug-in tools and (iv) web user interfaces to manage and query data and run plugged-in analysis tools for experimentalists. Hand-written features were plugged into the generated software, such as import/export wizards and a large data matrix viewer, bridges to R/QTL, Ontology Lookup Service, and KEGG services; more plug-ins to GMOD/Gbrowse, standard GWA software and Bioconductor packages are planned.

Other researchers can (and have) edit(ed) the blueprint and add plug-ins to generate an extended XGAP version that suits their particular needs but still adheres to the standard XGAP formats. We are optimistic that XGAPs uniform data representation and MOLGENIS-based extensible software infrastructure will help the communities of genotype/phenotype researchers to share data and tools notwithstanding large variation between research aims.