MOLGENIS: rapid prototyping of biosoftware at the push of a button

Morris A. Swertz*[,1-6], K Joeri van der Velde[1], Alexandros Kanterakis[1], Juha Muilu[2,7] Tomasz Adamusiak[2,6], Martijn Dijkstra[1,7], Gudmundur A. Thorisson[2,9], George Byelas[1], Danny Arends[1], Anthony J. Brookes[2,9], Ritsert C. Jansen[1] and Helen Parkinson[2,3,6]

[1]Genomics Coordination Center, Groningen Bioinformatics Center, University of Groningen & Dept of Genetics, University Medical Center Groningen, The Netherlands. [2]EU-GEN2PHEN consortium. [3]EU-CASIMIR consortium. [4]BBMRI-NL. [5]NBIC/BioAssist consortium, [6]European Bioinformatics Institute, Hinxton, United Kingdom. [7]Institute for Molecular Medicine Finland (FIMM), Helsinki, Finland. [8]Center for Medical Biomics, University of Groningen, Groningen, The Netherlands. [9]Department of Genetics, University of Leicester, United Kingdom. Contact: Morris Swertz (m.a.swertz@rug.nl)

Project website: http://www.molgenis.org
Code: http://www.molgenis.org/svn/molgenis/3.3/
License: LGPLv3

Abstract

MOLGENIS provides bioinformaticians with a simple model to automatically generate flexible web platforms for all possible genomic, molecular and phenotypic experiments. Each generated MOLGENIS includes biologist friendly user interfaces, semantic interfaces to RDF, SPARQL and ontologies, processing interfaces to R and SOAP/REST web services, a text-file exchange format and full documentation. Galaxy compatible model extensions are currently being developed to integrate cloud computing into this mix.

Introduction

An increasing array of biotechnologies is producing unprecedented amounts of *omics data. There is a huge demand on bioinformaticians to provide their biologists with user friendly, scalable software infrastructures to capture, exchange, and exploit all these new data and provide semantic and programmatic interfaces to connect analysis tools [1]. While standardization is helpful, new research must be quickly accommodated for which efficient software variation mechanisms are needed [2]. We here present MOLGENIS, a model driven software toolkit to efficiently produce the software needed.

Methods

MOLGENIS uses templates to automatically convert a compact XML model into running software. Writing 500 lines of MOLGENIS model replaces 15.000 lines of code in Java, D2RQ-N3 [3], SQL and R. Existing databases can be quickly enriched with a MOLGENIS front-end using the 'ExtractModel' procedure. The standard generated platform of user interfaces, R interfaces, semantic interfaces, text file parsers and writers, and REST/SOAP web services can be extended via plug-ins to integrate research specific processing protocols. Obviously, the generator can be re-run often to accommodate new research (or when a new generator is added which features you want to add; like R and semantic web last year).

Results

We evaluated the MOLGENIS toolbox for many types of biomedical experiments ranging from sequencing to proteomics building on various community consultations [3], including:

XGAP: an eXtensible Genotype And Phenotype platform [4] for systems genetics (GWAS, GWL) on transcript, metabolic and protein data. See http://www.xgap.org

MAGETAB-OM: a microarray experiment data platform based on the MAGE-TAB data format standard. See http://magetab-om.sourceforge.net/

Pheno-OM: to integrate any phenotype data from locus specific annotations to rich cohort reports with the help of the OntoCAT ontology toolkit[4] . See http://wwwdev.ebi.ac.uk/microarray-srv/pheno/

FINDIS: a mutation database for monogenic diseases belonging to the Finnish disease heritage. See http://www.findis.org/molgenis_findis/

Conclusion

MOLGENIS enables rapid prototyping of biological software and eases sharing of data models and software components. New generators are added frequently; for example a Galaxy model parser to generate GridGain based cloud computing jobs that work MOLGENIS data entities. All this jazz greatly helps to reuse the best design patterns to timely compose intelligent software systems that "biologists want to have".

References

1. Thorisson GA, Muilu J, Brookes AJ (2009) Genotype-phenotype databases: challenges and solutions for the post-genomic era. **Nature Reviews Genetics.** 10(1):9-18.
2. Swertz MA, Jansen RC (2007) Beyond standardization: dynamic software infrastructures for systems genetics. **Nature Reviews Genetics**. 8(3).
3. Juha Muilu: schemalet website. http://www.schemalet.org
4. Swertz *et al* (2010) XGAP: a uniform and extensible data model and software platform for genotype and phenotype experiments. **Genome Biology**. 9;11(3):R2
5. Common Api for ontology Tasks.