# Towards a federated microarray gene expression repository using MOLGENIS and MAGE-TAB

Alexandros Kanterakis[1,2], Tomasz Adamusiak[3,4], Juha Muilu[4,5], Helen Parkinson[3,4], Despoina Antonakaki[1]
Morris A. Swertz[1-4,6]

[1.] Genomics Coordination Center, Department of genetics, University Medical Center Groningen & Groningen Bioinformatics Center, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen.[2.] BBMRI-NL consortium, http://www.bbmri.nl. [3.] European Bioinformatics Institute, EMBL-EBI, Hinxton, UK[4.] EU-GEN2PHEN consortium, http://www.gen2phen.org[5.] Institute for Molecular Medicine Finland, University of Helsinki, Finland. [6.] NBIC BioAssist biobanking taskforce.
Contact: Alexandros Kanterakis: a.kanterakis@medgen.umcg.nl

Project website:   http://magetab-om.sourceforge.net/
Code:              http://sourceforge.net/projects/magetab-om/develop
License:           LGPLv3

## Abstract

The national consortium for Netherlands biobank research infrastructure BBMRI-NL is embarking on microarray meta-analyses aiming to exploit the wealth of microarray and GWAS data currently fragmented between individual biobanks (>6500 samples). To this end there was a need for an easy to populate, customize and federate infrastructure to submit, host and share data, annotations and tools between local and central installations. Here we report the first results.

## Introduction

Primary objective is to (a) establish a web-based national repository for microarray gene expression data and (b) to populate it with well-annotated microarray experiment data from participating biobanks. Secondary objective is to (c) share the software as 'microarray database in-a-box' such that all BBMRI biobanks can reuse it locally and (d) can easily share/federate data and tools between local and central installations.

## BBMRI-NL and GEN2PHEN

The envisioned system should include suitable user interfaces for researchers, programmatic interfaces for analysis protocols and data federation, and should be easily extended to accommodate diverging local needs. None of the available (open source) systems seemed to provide this and meanwhile GEN2PHEN [2] started 'database-in-a-box' projects including a microarray system based on the MAGE-TAB file format [3] and the MOLGENIS [4,5] biosoftware platform. BBMRI-NL chose to sponsor this project with the following results:

## MAGE-TAB object model

We created an object model that matches the MAGE-TAB v1.1 [3] format which is adopted by many major institutes to share microarray investigations, samples, protocols, and data. As a tab-delimited, spreadsheet based format, it is easy to create targeting also BBMRI facilities that lack specialized bioinformatics support.

## MOLGENIS implementation

We encoded the MAGE-TAB model in 850 lines of MOLGENIS [3] XML and auto-generated 60K lines of software code. The result is a MAGE-TAB data management suite including web user interfaces, programmatic interfaces to R, JAVA, SOAP, REST and semantic interfaces to RDF and SPARQL.

## Data submission and sharing

We created data parsers to enable MAGET-TAB files submission and sharing, building on MOLGENIS import/export mechanisms which resolved much of the complicated foreign key dependency and performance issues when importing related data items into a database.

## Future work

The next phase in this project is to populate the system with data, deploy local installs, plug-in analysis tools, and research how MOLGENIS' interfaces can be used in practice to enable Web 2.0 levels of data and tool sharing within the privacy sensitive biobanking community.

## References

1. BBRMI-NL is a branch of Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-EU).
2. GEN2PHEN aims to unify human databases towards holistic views into Genotype-To-Phenotype data.
3. Tim F. Rayner et al. (2006) A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. BMC Bioinformatics, 7, 489
4. Swertz MA, Jansen RC (2007) Beyond standardization: dynamic software infrastructures for systems genetics. Nature Reviews Genetics 8(3).
5. MOLGENIS software: http://www.molgenis.org