

EMBRACE – BioMart Developments and Future

Syed Haider¹, Benoit Ballester¹, Richard Holland¹, Damian Smedley¹, Peter Rice¹, Arek Kasprzyk²

¹EMBL European Bioinformatics Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK.

²Ontario Institute for Cancer Research, Toronto, Ontario, Canada, M5G 0A3.

The EMBRACE network of excellence aims to provide a standardized application programming interface (API) to a majority of the core biomolecular databases including Ensembl, UniProt, MSD, ArrayExpress, Wormbase and several other information resources hosted at but not limited to EMBRACE partner institutes. The project has significantly enhanced the usability of these resources by improving data presentation, web services, APIs and e-learning hence facilitating their use by ‘wet-lab’ biologists.

BioMart features a built-in query optimization engine and provides support for data federation between remote repositories without physical relocation and post processing. Each data source can be accessed through any of the following:

- Web service
- BioMart-DAS server
- Perl/Java API
- URL Access
- MartView (web based GUI)
- MartExplorer (Java desktop application)
- MQL (shell based Mart Query Language).

Substantial improvements to the BioMart web service and web interface have been carried out to simplify the interaction with the system. As a result the BioMart web service interface has now been adopted by a number of third party software packages such as Taverna, BioConductor-biomart, Galaxy and Bioclipse, adding new ways to manipulate the data. This includes, an *out of the box* BioMart-DAS server which enables lateral data-federation offering data/annotation sharing between disparate clients. For advanced users, complex scientific workflows are also possible over any of the mart repositories through Taverna's BioMart plugin.

The web service interface provided by the BioMart project has significantly reduced difficulties with data-integration. Optimization features included in the package have been particularly useful when working with large data repositories storing high throughput experiment data such as genomic sequence or microarray experiments. Consequently, EMBRACE-Grid has adopted BioMart as its primary data management system, aiming to offer in the near future a new, versatile platform for biological data analysis.

The process of the conversion of a data source into a BioMart compliant schema followed by its configuration and deployment is completely automated now using *martj* tools. Further developments to ease the deployment process, scalability, secure access and empowering the system with visualizations and analysis plugins are underway.

BioMart is completely Open Source, licensed under the LGPL, and freely available to anyone without restrictions.

www.biomart.org

www.embracegrid.info

Project Source code: www.biomart.org/install-overview.html