The Mouse Resource Browser (MRB) – A near-complete registry of mouse resources

Michael Zouberakis, Christina Chandras, John M. Hancock, Paul N. Schofield, and Vassilis Aidinis

***Abstract*—The Mouse Resource Browser MRB (http://bioit.fleming.gr/mrb) is an easy-to-use database for searching and retrieving mouse resource information. Currently, MRB hosts a list of 200 mouse resources which are divided in 33 different categories. Apart from core information such as URL(s), contact information and free text descriptions of domain content, MRB holds valuable technical information for each resource; server technology used, relational database management system(s) utilized, programming language(s) of implementation, schema descriptive documents or actual database dumps and most importantly information about the integration and interoperability services provided. MRB also hosts an index of ontologies – both OBO and non-OBO – and minimum information standards for biological investigations. Moreover, MRB has implemented the maturity scheme developed by the CASIMIR consortium and attempts to assess resources accordingly. Ultimately MRB's goal is to provide a complete registry of mouse resources and their programmatic accessibility methods.**

I. Introduction

In recent years, through the use of recent technological advances like gene cloning and transgenic technologies, the laboratory mouse has come to serve as a premier animal model in studying the complex mechanisms involved in human disease. As researchers progressively rely on the mouse as a model organism to examine the complex mechanisms regulating disease and its pathological processes through the application of functional genomic platforms, such as transgenesis, targeted mutagenesis, expression profiling and bioinformatics, an increase in data production and the generation of numerous, scattered biological resources for the storage and sharing of data and biological material has been noted. These informatics infrastructures have subsequently become an important tool in assisting scientists to further their understanding of the biology of human disease. The wide range of data types, both genomic and phenomic, together with their world wide distribution among many differently specialised databases, makes it potentially difficult and frequently impossible to ensure preservation and consistency of information, data quality and future retrieval.

In this context, CASIMIR (Coordination and Sustainability of International Mouse Informatics Resources; www.casimir.org.uk), a coordination action of the 6th Framework Programme of the European Commission [1], focuses on the coordination and integration of databases set up in support of the 5th and 6th Framework Programme projects containing experimental data, including sequences, and material resources such as biological collections, relevant to the use of the mouse as a model organism for human disease.

According to recommendations generated by PRIME (Priorities for mouse functional genomics research across Europe, FP6 Coordination action: http://www.prime-eu.org/) and information obtained from Peters et al., 2007 [2], we have undertaken the collection of available databases of relevance to mouse functional genomics and assembled a useful list including a range of mouse bio- resources available in Europe. While compiling this list of biological resources, a need for providing a dynamic and interactive way of presenting this list was identified. As this list of resources is constantly enriched and/or updated, in order to stay informed with the latest developments and handle this ever-growing inventory, an efficient management system was required. As a result, CASIMIR, decided to create a database of mouse databases, the Mouse Resource Browser – MRB (http://bioit.fleming.gr/mrb), containing information on most available European and International mouse resources. MRB is a resource management project that provides a dynamic and interactive way of presenting an updated mouse resource list, with basic categorization of all resources, a search interface that helps users to detect the most appropriate database(s) according to their request, an abstract assessment of the scientific content and technical aspects of resources based on quality criteria defined by the CASIMIR consortium, indexes of both ontologies and minimum information standards. Furthermore having identified the most widely implemented interoperability technologies utilized by the mouse community [3], MRB projects services based on these technologies and aims to become a portal of integration.

II. Database Design and Implementation

MRB is the front-end of a PostgreSQL fully normalized relational database. The front-end itself is a typical Java EE

Manuscript received July 3, 2008. This work was supported in part by the Sixth Framework Programme CASIMIR under Grant FP6-037811 (European Union) and in part by the Sixth Framework Programme MUGEN under Grant LSHG-CT-2005-005203 (European Union).

M. Zouberakis is with the B.S.R.C. Alexander Fleming, Vari, Greece (e-mail: zouberakis@fleming.gr).

C. Chandras is with the B.S.R.C. Alexander Fleming, Vari, Greece (e-mail: chandras@fleming.gr).

J. M. Hancock is with the Bioinformatics Group, MRC Harwell, Harwell, Oxfordshire OX11 0RD, U.K. (e-mail: j.hancock@har.mrc.ac.uk).

P. N. Schofield is with the Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3DY U.K. (e-mail: PS@mole.bio.cam.ac.uk).

V. Aidinis is with the B.S.R.C. Alexander Fleming, Vari, Greece (corresponding author phone: +30-210-9654382; fax: +30-210-9654210 e-mail: v.aidinis@fleming.gr).

application that follows the MVC architectural pattern generating three transparent layers:

1) the EJB layer,

2) the intermediate Session layer and

3) the interface/web layer.

Keeping the database schema as simple as possible and avoiding extended use of stored procedures and DBMS specific functions and types, the EJB layer – an object oriented (OO) API mapped to the database – is restricted to essential practicality. On the contrary the intermediate layer handles most relational and combinatorial functionality mounting to greater degrees of complexity. Finally the interface layer handles data representation.

Currently MRB is deployed on Sun's open source application server Glassfish.

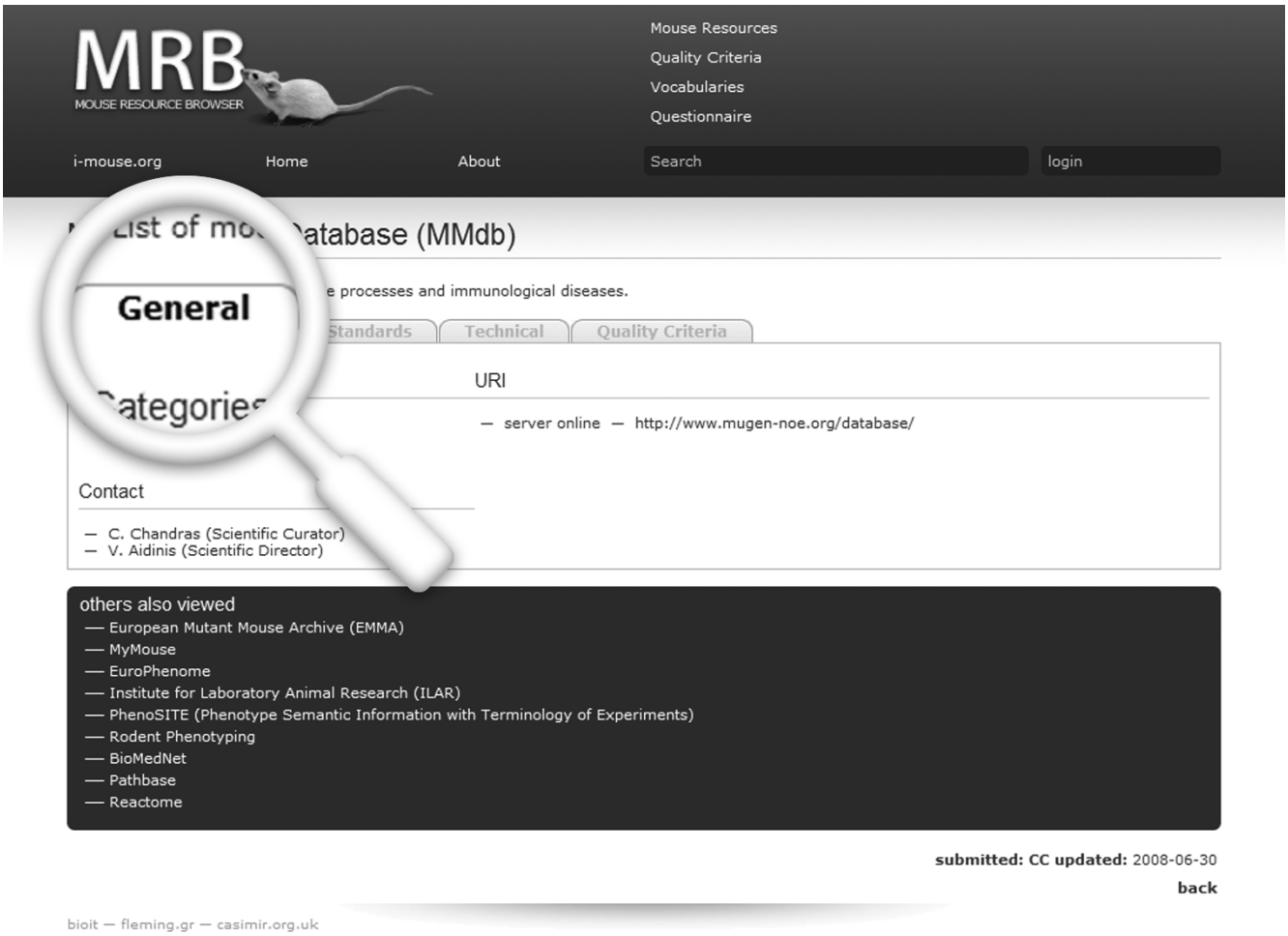

Fig. 1. Screen shot of MRB; the view of a mouse resource demonstrating the use of tabs per data set section. Here the 'General' tab is on focus.

III. Content Management

MRB's data collection was compiled and is being updated through extensive literature review and web browsing, direct or indirect resource personal contact – mostly via questionnaires – and occasionally via user recommendations. Data submission is only possible through MRB's curation team, who carefully check collected or submitted data for accuracy and completeness. In order to maintain a constantly updated version of MRB for the user to refer to, mouse resource information is under regular curation and is continuously revised. The date that the information for the respective biological resource was last revised is also available for the user.

Contrary to all other provided information, programmatic access related information is analyzed and presented on the fly, based solely on the provided interoperability services without any intervention from MRB's curators.

Even though MRB can support multiple user groups with different levels of access rights in an ordinary CMS fashion, it was opted to simply enable read access to all data and to all visitors with absolutely no restriction. For obvious reasons only administrative staff is granted write permissions and are thus required to log in.

IV. Data Modeling and Content Delivery

Conceptually MRB is governed by five primary entities; all of which are fully searchable and provided with an index page within the discussed application.

A. Mouse Resource

The primary entity of MRB is – as expected – accompanied by a detailed data set of all hosted entities. Mouse resource data sets are divided into four smaller sections according to the context of the details provided. Each section is visualised as a separate tab on the central page of every resource (Fig. 1).

The first section presents in detail general information on the particular biological resource. These include the "About" section where a short description of the resource is given, a content based categorisation of the particular database, the URL(s) of the particular database with an interactive link provided and a contact for the user to get in touch with the personnel of the particular resource.

The second section lists the ontologies and minimum information (MI) standards that are utilized by a resource, whereas the third section is devoted to technical information. From a technical point of view mouse resources are split into 3 categories:

1) relational databases,
2) object oriented and
3) flat files

The technical tab defines the category each resource belongs to, lists the programming languages and the database management system(s) used to develop the resource and indicates the server technology it is deployed on. Additionally, file downloads related to the resource's schema – i.e. complete database dumps, schema dumps, image diagrams modeling the schema etc. – are provided where available.

With regard to interoperability the most important provisions in the technical section are the web service links. Presently this set includes links to plain web pages describing how to programmatically access a resource, links to BioMart, query interfaces and direct links to WSDL files. All links come with an additional indicator that reflects the status of the server each link is pointing to and an '[a]' link (Fig. 2); a pointer to a web service analysis servlet entitled **wsAnalyzer**.

wsAnalyzer aims to break down documents that describe programmatic access to a resource on the fly, detect the methods that enable remote access and analyze their input and output parameters, displaying findings in a human readable format. In the long run collecting the findings of wsAnalyzer will be a first step to define:

1) an ontology of remote procedures and parameters
2) a MI standard for programmatic access services

Both points are considered by the members of the CASIMIR consortium to be milestones for the standardization of programmatic accessibility and in extent the integration of mouse resources.

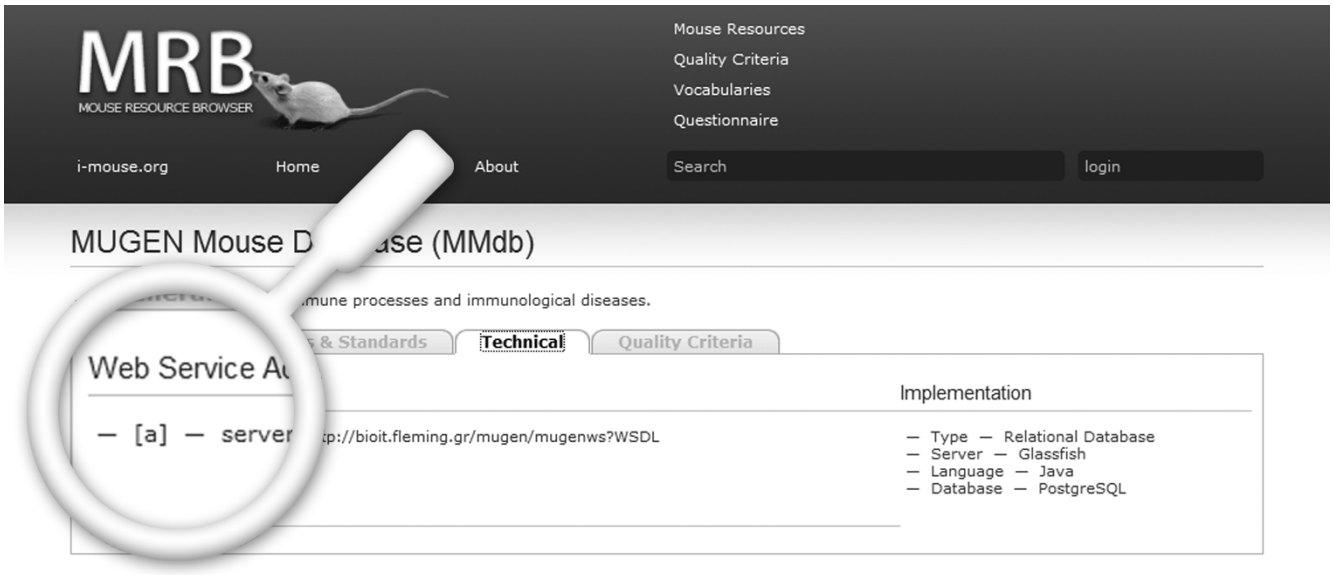

Fig. 2. The 'Technical' tab of a mouse resource view. Focusing on the 'Web Service Access' section's '[a]' link to 'wsAnalyzer'.

The current version of wsAnalyzer fully supports WSDL 1.1. Support for version 2.0 WSDL is in beta and future releases of MRB are planned to support WADL [4].

The fourth section is dedicated to a set of quality criteria as defined by CASIMIR participants [5]. Each criterion exemplifies a specific topic or area of importance and has three different levels of maturity. Resources that address these specific topics to a certain maturity level are awarded with an appropriate indication.

B. Category

Category – as mentioned above – is an entity mainly used to group mouse resources according to their content and usage. Resources usually fall into more than one category, so each category's dedicated view displays the number of related resources and a short description. Category entity entries are indexed and fully searchable.

C. Maturity Criteria

When compiling a list, there's an almost fontal tendency to evaluate one's findings. Naturally a set of assessment criteria is required to proceed with such a task. Within this context, members of the CASIMIR consortium developed a scheme of criteria to classify the diverse collection of resources. Confronting variability, these criteria – referred to as maturity models – are of a broad outline with three possible levels of variation – referred to as maturity levels. Contrary to traditional and perhaps more strict assessment criteria, the maturity model approach offers a rather loose evaluation method, abstractly summarizing the range and quality of a resource's provided services rather than providing a subjective ranking system of hierarchical character.

In the context of MRB, maturity model entities are indexed and fully manageable by administrative users. Satellite entities maturity levels – also fully manageable by administrative users – comprise maturity models and are therefore listed in every model's detailed view.

D. Ontology

Considering the importance of ontologies in biomedical data integration [6], [7] and their penetration in the mouse community, MRB's provision of an ontology index was important to support and facilitate interoperability through accepted semantic standards. Closely related to the central entity of MRB, the ontology entity is provided with an index view, listing all entries. Moreover each ontology is accompanied by a small dataset consisting of a short description, an indication of whether the ontology is an OBO or non-OBO ontology and an indication of whether it is made in-house or not. A list of external links to various resources – i.e. the ontology's home, its latest downloadable version in OBO or OWL format etc. – are also provided. All indexed ontologies are fully searchable.

E. MIBBI

Since standardization is such a pivotal prerequisite for the integration of – mouse – resources, projects falling under the MIBBI (Minimum Information for Biological and Biomedical Investigations) umbrella tend to be crucial points of reference. The rationale of developing MI checklists to describe methods, data and analyses [8] has resulted in the consolidation of standards such as MIAME and is constantly winning ground.

An index of fully searchable and manageable MIBBI projects is provided by MRB. The MIBBI entity – within MRB's context – is followed, in a similar fashion to the ontology entity, by a short description and a list of related resources.

V. Querying and Other Database Features

MRB provides a browsing/filtering interface to the underlying data, allowing formulation of queries in order to narrow down the list of potential results and to coordinate screening of certain biological databases and resources. Specifically, the mouse resource index page bears two drop-down menus; one to sort the collection either alphabetically or according to the date they were last revised and a second one to filter data according to their resource category. Alternatively stored data can be queried formulating *ad hoc (string)* queries. At all times, the user can enter the desired query word(s) into the available text box on the top of the page and seek for the particular word(s) at all levels of the database. In all cases, a chart is automatically generated with the corresponding results. Returned results may include mouse resources, resource categories, ontologies and minimum information standards, which in turn can be accessed by simply clicking on their appearing title.

Tracking mouse resource page loads, MRB generates a list of recently visited resources similar to the 'Customers Who Viewed This Item Also Viewed' feature of commercial web sites. The underlying principle of the feature is to introduce users to potentially useful resources.

VI. Conclusion and Future Prospects

The Mouse Resource Browser exceeding its initial goal – to become a content management system of mouse resources and databases – is aiming to become an intercessional link between resources of the mouse community. Focusing on integration and interoperability, collected information includes – amongst others – technical and programmatic accessibility details, ontologies and MI checklists.

Use of standards both in the framework of ontologies and metadata figuration is considered essential for the integrational potential of resources and – at least within the mouse community – is slowly becoming common practice.

On the contrary, interoperability services provided by mouse resources clearly need improvements. The review of interoperability technologies and the use-case example developed by CASIMIR members [3] pinpoints future directions and demonstrates how technologies like web services [9] and software like Taverna [10] and Molgenis [11] can be utilized to accomplish interoperation.

Following recommendations by and discussions with fellow CASIMIR members, future versions of MRB will include:

--An enhanced version of the wsAnalyzer feature, so as to support WSDL version 2.0 and WADL.

--Web Services in order to allow programmatic access to MRB's mouse resource collection.

--A Web Service generator for indexed mouse resources based on Molgenis. The central idea is to provide an application programming interface for resources that do not provide interoperability services with the simplest possible way by just opening a port and granting direct access to the actual database.

The source code of the Mouse Resource Browser application is available under the GNU General Public License (GPL) as a binary download and via cvs from the CASIMIR sourceforge project page (http://sourceforge.net/projects/casimir-org-uk/).

REFERENCES

[1] J. M. Hancock, P. N. Schofield, C. Chandras, M. Zouberakis, V. Aidinis, D. Smedley, N. Rosenthal, K. Schughart, The CASIMIR Consortium, "CASIMIR: Coordination and Sustainability of International Mouse Informatics Resources (Periodical style–Accepted for publication)", *8th IEEE International Conference on Bioinformatics and Bioengineering*, to be published.

[2] L.L. Peters, R.F. Robledo, C.J. Bult, G.A. Churchill, B.J. Paigen, K.L. Svenson The mouse as a model for human biology: a resource guide for complex trait analysis. *Nat Rev Genet*., vol. 8, pp. 58-69, Jan. 2007.

[3] S. Smedley, M.A. Swertz, K. Wolstencroft, G. Proctor, M. Zouberakis, J. Bard, J.M. Hancock, P. Schofield and other members of the CASIMIR consortium, "Solutions for data integration in functional genomics: a critical assessment and case study," *Briefings in Bioinformatics* , submitted for publication.

[4] M. J. Hadley, Sun Microsystems Inc. (2006, November 9). "Web Application Description Language (WADL)" [Online]. Available: https://wadl.dev.java.net/wadl20061109.pdf

[5] J. Bard, J.M. Hancock, P. Schofield, V. Aidinis, S. Smedley, G. V. Gkoutos, M.A. Swertz, M. Gruenberger, J. A. Blake, S. Green, M. Zouberakis, CASIMIR meeting in Cambridge, May 2008.

[6] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene Ontology: tool for the unification of biology," *Nat Genet.*, vol. 25, pp. 25-29, May 2000.

[7] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, OBI_Consortium, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S. A. Sansone, R. H. Scheuermann, N. Shah, P. L. Whetzel, and S. Lewis, "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration," *Nat Biotechnol.*, vol. 25, pp. 1251-1255, Nov. 2007.

[8] C. F. Taylor, D. Field, S.A. Sansone, J. Aerts, R. Apweiler, M. Ashburner, C. A. Ball, P. A. Binz, M. Bogue, T. Booth, A. Brazma, R. R. Brinkman, A. M. Clark, E. W. Deutsch, O. Fiehn, J. Fostel, P. Ghazal, F. Gibson, T. Gray, G. Grimes, N. W. Hardy, H. Hermjakob, R. K. Jr. Julian, M. Kane, C. Kettner, C. Kinsinger, E. Kolker, M. Kuiper, N. Le Novere, J. Leebens-Mack, S. E. Lewis, P. Lord, A. M. Mallon, N. Marthandan, H. Masuya, R. McNally, A. Mehrle, N. Morrison, J. Quackenbush, J. M. Reecy, D. G. Robertson, P. Rocca-Serra, H. Rodriguez, H. Rosenfelder, J. Santoyo-Lopez, R. H. Scheuermann, D. Schober, B. Smith, J. Snape, K. Tipton, P. Sterk, S. Wiemann, "Promoting Coherent Minimum Reporting Requirements for Biological and Biomedical Investigations: The MIBBI Project" [Online]. Available: http://mibbi.sourceforge.net/docs/papers/MIBBI.pdf

[9] The World Wide Web Consortium. "Web Service Activity" [Online]. Available: http://www.w3.org/2002/ws

[10] D. Hull, K. Wolstencroft, R. Stevens, C. Goble, M. R. Pocock, P. Li and T. Oinn, "Taverna: A tool for building and running workflows of services," *Nucleic Acids Research*, vol. 34, pp. 729-732, 2006.

[11] M.A. Swertz, E.O. de Brock, S.A.F.T. van Hijum, A. de Jong, G. Buist, R.J.S. Baerends, J. Kok, O.P. Kuipers and R.C. Jansen, "Molecular Genetics Information System (MOLGENIS) alternatives in developing local experimental genomics databases," *Bioinformatics*, vol. 20, pp. 2075-2083, Sept. 2004.