

WEB SERVICES

TECHNOLOGY & DATABASE INTEROPERABILITY FOR MOUSE RESOURCES

WHAT'S OUT THERE

- **32** relational database projects were spotted
- **4** different RDBMSs (MySQL, PostgreSQL, ORACLE & DB2)
- (Probably) **32** different development strategies & technologies
- **1** model: **THE MOUSE**
- **1** problem: **INTEROPERABILITY**

for a full list of resources visit the MUGEN Mice Database @ www.mugen-noe.org/database

(WHAT IS) INTEROPERABILITY

"The capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units"

-Information Technology Vocabulary, Fundamental Terms

In our case replace '*functional units*' with **databases** & '*users*' with **developers**.

Also note that '*unique characteristics*' refers to technical unique characteristics.

DB CONNECTIVITY (SO FAR)

- **Hyperlinks & Web Form Querying**
- **BioMart**
- **Web Services**
- **...MOBY**

Hyperlinks & Web Form Querying

The easiest way to provide access to other (on-line) databases:

- Hyperlinks
(i.e. click here to visit...)
- Web form querying
(involves web forms located in other websites/db front-ends)

A typical example is the **MMdb**, currently providing direct links to:

MGI | Ensembl | ArrayExpress | Entrez | Eurexpress | EMMA | JAX

In the context of the '**one-stop-shop**' rationale and attempting to provide as much information as possible within a '**max three click away**' limit, hyperlinks & web form querying provide **sufficient** interconnectivity.

DB interoperability – **ACTUAL DATA FLOW** - requires more than simple forwarding to other pages. We need **direct, dynamic** and on **request** data transfer technology.

BioMart

BioMart is an **advanced data management system** that has been integrated across multiple institutional sites with variable species coverage. It is incredibly well developed and documented and is fully integrated into the **Ensembl** genome browser.

The **good** points:

- The database can be included in BioMart's central project page and can be queried directly from there.
- Utilizing MartView the database can bear a nice web interface.
- Utilizing MartService web services are automatically installed.

The **weak** points:

- Currently BioMart supports only MySQL, PostgreSQL & Oracle.
- MartView & MartService need Apache web server installation.
- Can MartView absorb unlimited web services from other databases?

Web Services

Web Services are a *“software system designed to support interoperability between information systems over a network.”*

Technically speaking web services are *SOAP* standard *XML* over *HTTP*.

The **good** news:

- Web Services do not depend on the RDMS and the schema of a database.
- Web Services are not dependant on the front-end's programming language.
- Web Services can be freely and easily absorbed by other databases.

One **minor drawback**:

- Web Services have to be developed for **every single** database.

Web Services & Mouse Resources

Why Web Services seem to be the reasonable choice:

- Our resource index includes dbs that are built on DB2. BioMart does not support DB2. Web Services do not depend on the RDBMS.
- Web Services provide access to XML 'shaped' data that can be processed, handled, imported...by **ANY** programming language.
- Web Services do not require a specific web server to be installed on your (physical) server.
- Web Services can allow data flow between **ALL** database projects.

A Web Service Client Example

Currently we have two projects running: **MMdb** & **TgDb**.

MMdb is a virtual mouse repository.

It's being heavily curated by our curation team and features:

- availability,
- genotypic and
- phenotypic data for mouse models.

The screenshot shows the MUGEN MICE DATABASE interface. The browser address bar displays `http://localhost:8080/mugen/mde.jsp`. The page title is "mugen mice". Below the title, there are navigation links: "home | help | contact | about | search:" and "mutants | genes | alleles | mp terms". A search bar is present with the text "mugen mice". Below the search bar, there are filters for "Research Applications", "Gene", "Institution", "Researcher", "Mutation Type", and "MP Terms". The main content area displays a table of mouse models with the following columns: "MUGEN ID", "Common Line Name", "Mutation(s)", "Research Application", and "Updated". The table lists 14 mouse models, including M119001 (TNFalpha-/-), M119002 (MOG-Cre), M139001 (CD11c-Cre-ER^T), M139004 (ST33.396), M139006 (KL25), M139007 (CD19-Cre), M139008 (IL-10 KO (Roer)), M141001 (PLP-Cre-ER^T), M141002 (IL-10 KO (Can)), M141003 (IL-10R KO), M141004 (IFNgamma-Receptor), M141005 (PI3Kgamma, p110gamma), M141006 (penetraxin 3, TSG-14), M141007 (TIR8), M143002 (218), M143003 (ST-NP), M143004 (Smarta), M145001 (Line 7), M145002 (VIL0), and M145003 (Cmy-Cre, cre deleter, deleter strain). The table also includes navigation controls and a login form on the left side.

MUGEN ID	Common Line Name	Mutation(s)	Research Application	Updated
M119001	TNFalpha-/-	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-09-13
M119002	MOG-Cre	targeted mutation	TRANSGENIC TOOL	2007-09-13
M139001	CD11c-Cre-ER ^T	transgenic	TRANSGENIC TOOL	2007-08-28
M139004	ST33.396	transgenic	TRANSGENIC TOOL	2007-09-04
M139006	KL25	insertion, targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-03
M139007	CD19-Cre	targeted mutation	TRANSGENIC TOOL	2007-08-03
M139008	IL-10 KO (Roer)	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-03
M141001	PLP-Cre-ER ^T	transgenic	TRANSGENIC TOOL	2007-09-13
M141002	IL-10 KO (Can)	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-03
M141003	IL-10R KO	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-09-13
M141004	IFNgamma-Receptor	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-03
M141005	PI3Kgamma, p110gamma	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-06
M141006	penetraxin 3, TSG-14	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-06
M141007	TIR8	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-06
M143002	218	transgenic	MODEL OF IMMUNE PROCESSES	2007-09-12
M143003	ST-NP	transgenic	TRANSGENIC TOOL	2007-09-13
M143004	Smarta	transgenic	MODEL OF IMMUNE PROCESSES	2007-08-06
M145001	Line 7	transgenic	MODEL OF IMMUNE PROCESSES	2007-09-13
M145002	VIL0	insertion, targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-06
M145003	Cmy-Cre, cre deleter, deleter strain	transgenic	TRANSGENIC TOOL	2007-09-13

visit @ www.mugen-noe.org/database

A Web Service Client Example

TgDb is also a virtual mouse repository, but focusing on transgenic mutants.

It's still being actively developed & curated.

Some key features:

- availability,
- genotypic and
- anatomy data.

The screenshot shows the TgDb web application interface. The browser address bar displays 'http://localhost:8080/tgdb/'. The page header includes the TgDb logo and 'BSRC Fleming'. A notice states 'TgDb beta version 0.7d.2c is still being heavily curated.' and provides statistics: 'TOTAL MOUSE STRAINS HOSTED: 185' and 'PARTIALLY CURATED & PUBLICLY AVAILABLE STRAINS: 18'. There are search and login fields. The main content area is titled 'Transgenic Mice Index' and shows a table of 18 transgenic mouse strains. Below the table are navigation controls and search filters for 'transgene' and 'researcher'.

TgDb ID	Line Name	Date
TgMUS:143003	ST-NP	2006-08-29
TgMUS:145004	K14-Cre	2006-08-30
TgMUS:1207009	CD21-Cre	2007-01-22
TgMUS:139004	ST33.396	2007-03-13
TgMUS:139003	RAGE ^{fl}	2007-03-17
TgMUS:143001	Z/EG	2007-03-17
TgMUS:171001	CD11b-Cre	2007-03-18
TgMUS:139002	aAT-Cre-ER ^T	2007-03-20
TgMUS:215012	Foxp3 EGFP	2007-04-12
TgMUS:211009	Langerin-DTREGFP	2007-06-12
TgMUS:119002	MDGI-cre	2007-06-12
TgMUS:139007	CD19-Cre	2007-06-12
TgMUS:145003	Cmv-Cre	2007-06-12
TgMUS:147002	LysM-Cre ^{K/I}	2007-06-12
TgMUS:193001	Langerin-GFP ^{K/I}	2007-06-12
TgMUS:139001	CD11c-Cre-ER ^T	2007-06-13
TgMUS:141001	PLP-Cre-ER ^T	2007-06-13
TgMUS:145005	Mx-Cre	2007-06-13

visit @ www.fleming.gr/tgdb

A Web Service Client Example

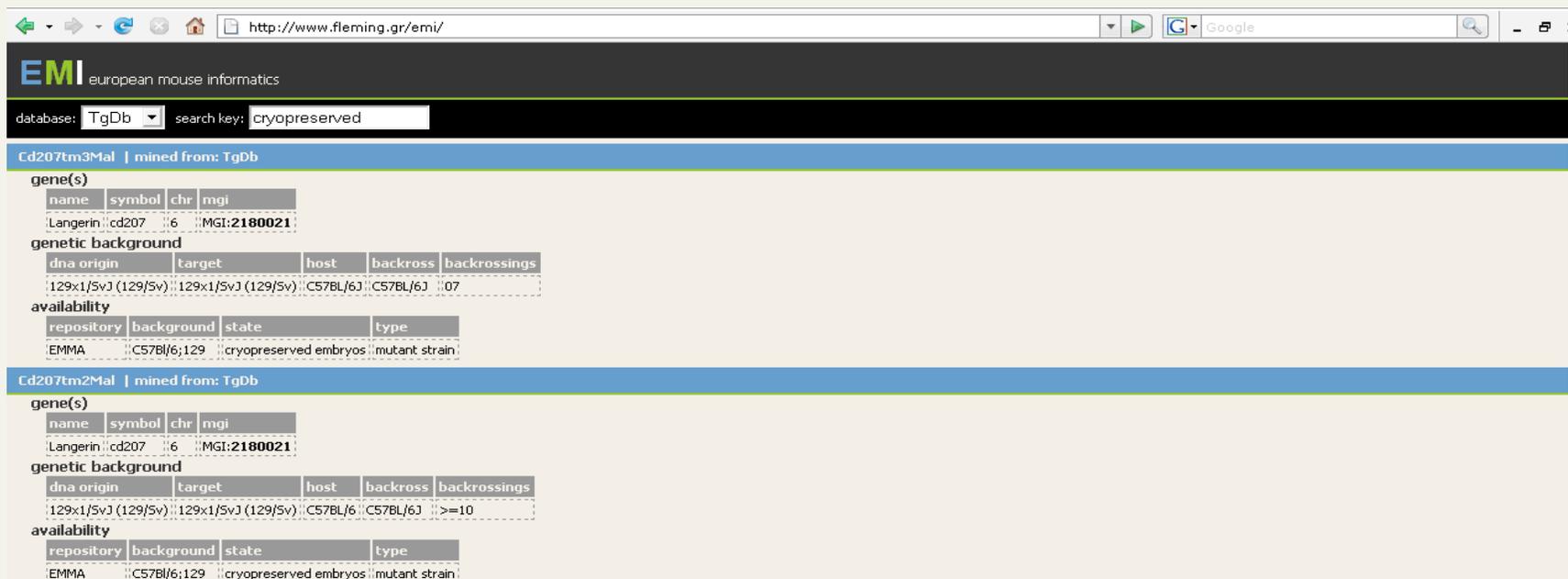
We are developing Web Services for both projects. In fact we have completed the core for both Services, so naturally we have deployed a little Web Service client project to absorb the two Web Services for testing and demonstrational purposes.

This client was called **EMI** (European Mouse Informatics) and just like our two main projects, it is **mouse centric virtual repository**.

EMI is not the front-end of a relational database. It is a simple web application that handles data representation for both MMdb & TgDb.

What is important to note is that MMdb & TgDb have significant differences in their schemas. EMI is overcoming this issue as it handles raw data fed from the two Web Services. Whatever field is provided via a Web Service is displayed.

A Web Service Client Example



The current screenshot is querying TgDb for 'cryopreserved'. Strain, genes, genetic background, availability and phenotypic information can be displayed and queried.

EMI can be found @ www.fleming.gr/emi

The Bigger Picture

CASIMIR's WP7 has compiled an impressive resource list which involves remarkable database projects that approach the mouse model from a different perspective (virtual mouse repositories, phenotype and anatomy projects, mouse genome projects etc.).

The biggest challenge yet is to integrate those resources in a '**one-stop-shop**'.

Considering the fact that many mouse resource projects have already implemented Web Services and the benefits of utilizing this technology, we can only encourage further Web Service development in order to ...

The Ultimate Goal

...to achieve the ultimate goal; the launch of a project that will absorb all relevant available Web Services, and will provide:

- A useful and easy to use interface, providing an all-embracing view/dataset of the mouse model.
- A combinatorial optimization platform for mouse model data, paving the way towards systems biology.

