

## WHY ANOTHER MICROARRAY LIMS

Joel Arrais<sup>\*</sup>, Luis Silva<sup>\*</sup>, Maria Rodrigues<sup>\*</sup>, Laura Carreto<sup>\*\*</sup>,  
José Luis Oliveira<sup>\*</sup> and Manuel A.S. Santos<sup>\*\*\*</sup>

<sup>\*</sup>Institute of Electronics and Telematics Engineering of Aveiro (IEETA), 3810-193, Aveiro

<sup>\*\*</sup>Faculty of Science and Technology, University of Algarve, Campus de Gambelas,  
8005-139 Faro; <sup>\*\*\*</sup>Department of Biology, University of Aveiro,  
3810-193 Aveiro;

jpa@ieeta.pt

**Abstract: Microarray technology holds the promise to revolutionize the future of biology and medicine unlike any other previous technology. However, the full power of microarrays can only be achieved, if researchers are able to efficiently store and share their results. In the past years, traditional LIMS have evolved in order to cope with microarrays technology. In this paper we evaluate a set of current microarray LIMS. Thereafter, and despite the several solutions available, we realise that none of them meet the requirements of our research centre. Here, we also present MIND, a Microarray Information Database, that intends to be not only a new microarray LIMS, but also an all-in-one package, that allows data management, analysis and sharing.**

### Introduction

In the last few years, microarray technology changed not just the way biologists conduct their laboratorial experiments as also the way clinical diagnosis and disease treatment can be made [1-3].

However, a single experiment may involve dozens of microarrays, each containing tens of thousands of spots for which several measurements need to be recorded, resulting in millions of elements of information [4]. Because each element must be properly treated and stored, the point where a LIMS (Laboratory Information Management Systems) is required is quickly reached [5].

In the past years, traditional LIMS have evolved in order to cope with microarrays technology. New versions and new products, commercial and public domain solutions, have appeared, increasing the difficulty in choosing one.

In our research centre, a facility of microarrays is being installed and, with the increase of the data, the necessity of having a LIMS increased too [6]. Moreover, the need of sharing the results of our experiments created a demand for solutions that also could act like shared repositories.

Here, we review some microarray LIMS and describe the challenges we face in choosing one. Then, we present MIND, a MIAME compliant Microarray Information Database [7, 8], that consists in a new

proposal to accomplish what we would expect from a microarray LIMS.

However, MIND intends to be more than just a simple microarray LIMS. MIND will represent a web based system that handles the storage, management, reporting and analysis of microarray data.

Effectively, with the development of MIND we have reached a system capable to improve the efficiency of the management of microarray laboratorial data.

### Microarray LIMS

Due to the raised amount of data that needs to be structured and stored, the use of traditional registry solutions, like paper, doesn't comply with the necessities of the laboratory. So, it is required the use of an information system, that helps the management of all the laboratorial process. However, due to the large amount of the existent commercial and public domain solutions, the choice of one LIMS isn't an easy task.

The first decision to be made is between the use of an existent application and the development of a new one. In the first case, two alternatives still remain: a commercial or a public domain solution.

A commercial solution has, as main advantage, the access to an application in a steady and mature state that results from several years of development. However, these products are typically expensive, and there is always the risk of reaching a point where the current system no longer satisfies the actual requirements. In these situations, beyond the costs associated with the installation of a new system, it also should be taken into account the costs related with the migration of the existing data.

In the existent microarrays LIMS, it can be outstood the high relevance that the public domain solutions hold. Beyond not possessing acquisition costs, these solutions are, in many cases, fruit of the work of important groups. But, on the other hand, these solutions can be oriented to a specific end, which implies that they could not be reasonable for the requirements of every laboratory. Some of these applications, especially the web based ones, oblige to an advanced knowledge for the installation, which may also imply a cost. There are also problems with the non-existence of technical support, as well as the

possibility of the application be discontinued, what implies the transference for a new system, with all the problems and costs that this situation entails.

The development of a solution since the ground stage, and despite the involved risks, can be a very rewarding option. In the first place, it is necessary to have in consideration that this option requires one group of computer technicians, as well as the existence of future users to validate all the development stages. In this type of implementation, it also has to be taken into account the imperative of a vital period, corresponding to the development of the application. Although some disadvantages, the main advantages of this options are to cope with laboratory's necessities, as well as to always assure that future improvements can be carried through.

### Evaluation of microarray LIMS

Our aim for a microarray LIMS is a consequence of the installation of a microarray facility in our research centre. In fact, in a short period of time, large amounts of data will be generated and an efficient and complete system will be needed. However, due to the large quantity of existent systems, and to the threat that could come from the choice of one less suitable, several efforts were spent in finding the right system. To help in this task, it has been considered an evaluation methodology, in which the features of all the system are equally compared (Table 1).

Based on the available systems and on their relevance the following set was selected:

- BASE, *BioArray Software Environment* (<http://base.thep.lu.se>);
- MiamExpress/ ArrayExpress (<http://www.ebi.ac.uk>);
- maxd (<http://bioinf.man.ac.uk/microarray/maxd>);
- LIMaS, *Laboratory Information Management for Array Systems* (<http://www.mgu.har.mrc.ac.uk/facilities/microarray/limas>);
- MADAM: *MicroArray DAta Manager* (<http://www.tigr.org/software/tm4>);
- LAD, *Longhorn Array Database*, (<http://www.longhornarraydatabase.org>).

In spite of all the above systems have the same purpose – to manage microarray data –, this goal is achieved by them in slightly distinct ways. Different system requirements, different user interfaces and different auxiliary tools are just some of the divergences between all the presented systems. In accuracy, in this evaluation, and for each LIMS, the following information has been collected: use's license; congruence with MGED norms; image's storing capabilities; possibility to introduce references for external databases; analysis and visualization tools; operative systems; database management system; web server; interface's type, installation's complexity and programming languages used.

One of the common features between all the

Table 1: Global evaluation of microarray LIMS

	BASE	MiamExpress/ ArrayExpress	maxd	LIMaS	MADAM	LAD
License	GNU	open source	open source	Academic use	open source	open source
MGED	MIAME	yes	yes	yes	yes	yes
	MAGE-OM	no	yes	no	no	no
	MAGE-ML	no	yes	yes	no	no
Image storage	yes	yes	yes (Link)	yes (Link)	yes (Link)	yes
External references	yes	no	yes	no	yes	yes
Analysis and processing tools	built in	none	MaxdView	none	MIDAS e MeV	built in
Operative System	Linux	Linux	any	any	any	Linux
DBMS	MySQL, PostgreSQL	Oracle	MySQL	MySQL, Sybase and Oracle	MySQL	PostgreSQL
API	Web	Web	Java package	Java package	Java package	Web
Web server	Apache	Apache	-	-	-	Apache
Installation complexity	high	high	medium	medium	medium	medium
Computer languages	Php, Java, Javascript, C++, Perl, R	Perl, Javascript	Java	Java	Java	Perl, Javascript

presented LIMS is that they are non-commercial solutions. Despite that, some of them have some restrictions. For example, LIMaS is only free for academic use and the ones of GNU license type can be used and modified, but all the modifications must be made available.

Maxd is another example, because, in spite of its license type is open source, an initial web connection is required, restricting its operability.

Due to their importance, we had into account the MGED standards (<http://www.mged.org>) [7-9]. From the comparison table, it can be verified that all of the presented applications follow MIAME recommendations, but just one use MAGE-OM and two allow MAGE-ML export [10]. Indeed, almost all the presented solutions announce that soon they will support MAGE-ML, but, by now, only ArrayExpress and maxd really does it. In fact, for an effective microarray data sharing, the high importance of MAGE-ML exportation option should be noticed.

The image of the microarray hybridisation is the only valuable proof of the realization of an experiment. Therefore, its storage is important, and it would be even better, if done by the LIMS. However, in some LIMS, just an external link of the image location is stored, which may promote information inconsistency.

After the completion of a microarray experiment, a pipeline of stages of data treatment and analysis need to be done. As it is possible to observe, these functionalities can be enclosed in the LIMS or can constitute an independent application [11, 12]. In the second case, better functionalities are achieved, but a better relation between the data and their analysis results is still required.

In the presented LIMS, do exist two types of interfaces: desktop application and web application. In the first kind, all the data are stored locally and data sharing can only be accomplished by the explicit exportation and publication in any public database. Only this way, different users can access the same information with fast and easy sharing of the results.

In the tested systems, and independently of the interaction model, it can be verified a preference for the use of solutions without acquisition costs. Indeed, the choice of the operative system (Linux), of the database (MySQL or PostgreSQL) and of the programming languages (Java, Perl, PHP, R) implies that. As the tested applications are free to use, it wouldn't be expected that the necessary technologies to its use weren't also free.

The installation's complexity, and despite it isn't an accurate measure, tells if each system is easy, or not, to install, configure and maintain. This way, using a three level scale (*low*, *medium* and *high*) each system was classified as *low* if it could be installed by any user, *medium* if it requires some technical knowledges and *high* if advanced technical skills are required. From the presented LIMS, all the desktop applications are easy to install, having as unique difficulty the database's configuration, and so they are classified as *medium*.

However, the web application's systems, and in special the ArrayExpress and the BASE, require advanced installation's knowledge and so they are classified as *high*. It should be noticed that despite of all the presented systems are free to use, the technical support costs must be taken into account, especially in a system that requires high installation's skills.

Now that all the characteristics of the presented systems have been explored it is possible to make a better judgment. Indeed, all the presented systems are valid solutions but none of them is able to comply with our expectations. From all the presented systems the maxd seems to be a best option, once that it allows data management and data analysis, it is easy to install and, the most important, it allows data exportation in MAGE-ML format. Nevertheless, it is a desktop application, which implies some limitations in a collaborative environment. From the web server systems, the BASE seems a suitable proposal, once that it allows the data management and analysis in the same interface. However, BASE still doesn't allow the data exportation in MAGE-ML, what is considered a very important feature [10].

So, if we are interested in a web application with visualisation and analysis tools and with the MAGE-ML data exporting option, none of the proposed systems totally satisfies our requirements. This led us to develop a new microarray LIMS proposal.

## MIND

MIND stands for Microarray Information System and consists in a new proposal of a database server to manage large amounts of data generated by microarray experiments. MIND intends to be a user friendly web interface to handle the storage, management, reporting and analysis of microarray data. Moreover, MIND aims to be a platform prepared to integrate new functionalities and communicate with other microarray information systems.

MIND is meant to be installed in one server that might be shared by one or several microarrays laboratories. The server is accessed via a web browser, using personal login accounts. Each user is admitted to a private area, where he can store data related with ongoing experiments, and to a public area, where all the public experiments are visible.

MIND is prepared to be installed in any operative system and uses only freely available software. MIND runs in Tomcat web server and MySQL is the used database server. In the interface were applied the following languages: Java, Javascript and R. MIND is available for test in our server at <http://bioinformatics.iceta.pt/mind> (Figure 1a-b).

MIND's development, firstly, considered the flow of a microarray experiment. For each microarray, it is required the storage of the information about each spot, as well as the organization of all the spots in the microarray. It is also important to keep the information about the samples: their sources and the applied

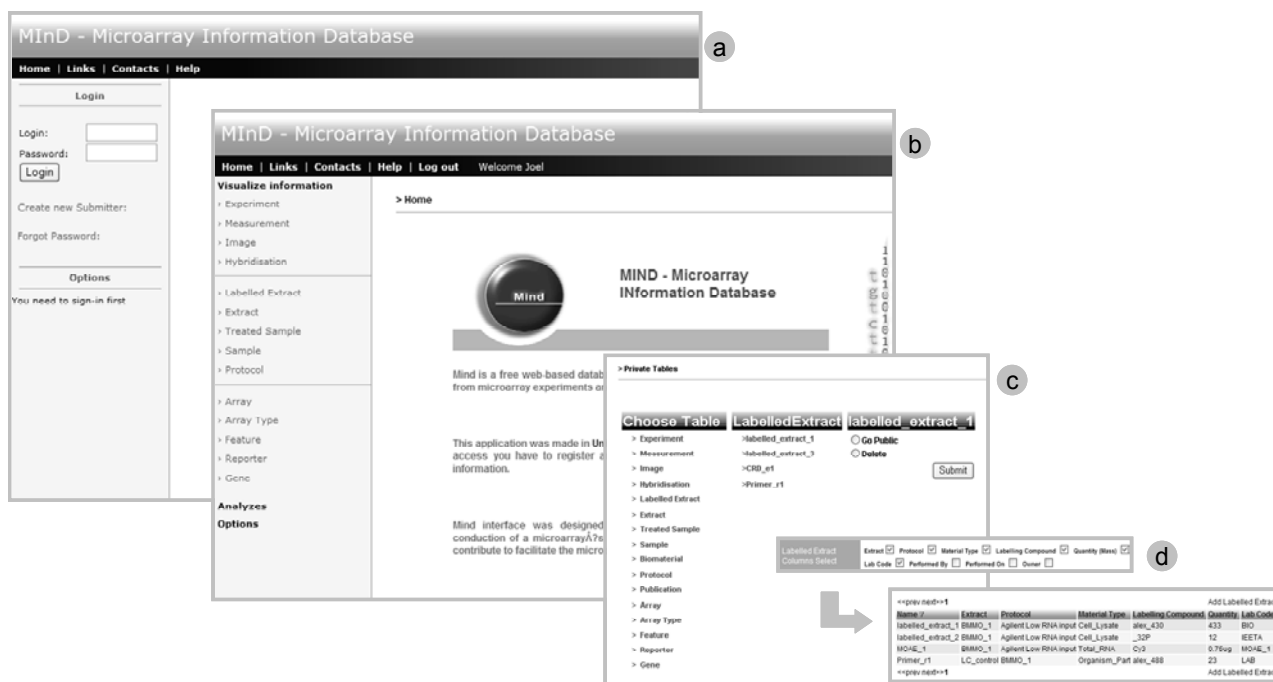


Figure 1: Mind interface

- a) Login interface;
- b) Main interface;
- c) Publish private data;
- d) Configure user interface.

methods in the extraction and treatment. After the hybridisation, the microarray is scanned, the image is normalized, quantified and stored. Then, for each image, information about each spot is acquired. Over these data, several analyses can be made, being the first, and the most important, the one that intends to validate the quality of the microarray.

In order to promote an easy and effective share of the data and to allow automatic software updates to all users, a web implementation has been chosen for MIND. However, if the inserted data stays immediately publicly available, several problems can come from the sharing of preliminary experiments results. So, in MIND, when new experiment data is inserted, it is kept in a private area. At any moment, the private data can be made public (Figure 1c).

The information that defines a microarray experiment can be very extensible, and consequently, a microarray LIMS could be a complex tool. However, depending on the profile of each user, only subsets of all the available information will be required. Based on that, MIND allows the users to select his own data visualisation profile. In the options menu, for each data item, a flag can set that specific data as visible or invisible. It is also possible to define, for each user profile, the default user data (Figure. 1d).

In some of the previously presented LIMS, just data storage was considered. However, due to the large amounts of data analysis tools in the LIMS, it becomes crucial to assure the correct correspondence between

data and their results. In MIND, a generic and extensible data analysis implementation has been made. Indeed, using the R language (<http://www.r-project.org>), several standard analysis scripts are already available, in MIND interface. Thus it is possible to verify the MA plots and a virtual false colour representation of the array. And, in the future, any other R script for a specific problem can be added by MIND' users.

The MIND platform is prepared to integrate further functionalities and to communicate with other microarray information systems. Indeed, due to the high value of the results of microarray experiments in clinical area, our future goal is to promote MIND interoperability with Hospital Information Systems.

## Conclusion

Microarrays LIMS are an indispensable tool in a microarray laboratory, due to their capability to manage large amounts of data. However, choosing a LIMS, from the several commercial and public domain solutions, can be a daunting task.

Here, we reviewed the most relevant existent LIMS. Then, observing their strengths and weaknesses, we identified the major problems they evidence. Finally, based on that and with the help of microarray experts, MIND is presented as a new proposal.

MIND is a powerful but yet easy to use application, that wants to make the bridge between the complex web

applications and the restricted desktop ones. It is fully compliant with MIAME and allows data exportation in MAGE-ML format.

By now, our efforts are to finish the development of the basic platform of MIND. After this stage, to allow a more creative and flexible use of MIND by researchers or data miners who wish to be able to automate complex queries of the available data, other functionalities will be implemented. Techniques for determining data quality in an automated fashion also remain an active area of research. Another possibility is the study for the improvement of our interface interoperability with clinical information systems.

When finished, the MIND intends to be not just a LIMS, but also a repository of microarray data results. We hope that our work contributes to facilitate the microarray experts task and to promote microarray data into a clinical level.

### Acknowledgement

This work was developed under the endorsement of the NoE INFOBIOMED, funded by the European Community, under the Information Society Technologies (IST) program.

### References

- [1] FADIEL A., NAFTOLIN, F. (2003): 'Microarray applications and challenges: a vast array of possibilities', *Int Arch Biosci*, pp. 1111-1121
- [2] CHOUDHURI S. (2004): 'Microarrays in Biology and Medicine', *J. Biochem Molecular Toxicology*, 18:4, pp. 171-179
- [3] COJOCARU, G.S., REHAVI, G., KAMINSKI, N. (2001): 'The Use of Microarrays in Medicine', *Microarrays in Medical Research*, 3, pp. 292-296
- [4] LEMIEUX, B., AHARONI, A., SCHENA, M. (1998): 'Overview of DNA chip technology', *Molecular Breeding*, 4, pp. 277-289
- [5] NING, J. (2004): 'Development of an Integrated LIMS for Microarray Facility Center', presented at Eighth Annual International Conference on Research in Computational Molecular Biology (RECOMB 2004), San Diego
- [6] CARRETO, L., SANTOS, M.D.C., ARRAIS, J., RODRIGUES, M., SILVA, L., OLIVEIRA, J.L., SANTOS, M.A.S. (2005): 'Implementation of a National Facility for DNA-Microarrays in Portugal', presented at 8th International Meeting of the Microarray Gene Expression Data Society, Bergen
- [7] BRAZMA, A., HINGAMP, P., QUACKENBUSH, J., SHERLOCK, G., SPELLMAN, P., STOECKERT, C., AACH, J., ANSORGE, W., BALL, C.A., CAUSTON, H.C., GAASTERLAND, T., GLENNISON, P., HOLSTEGE, F.C.P., KIM, I.F., MARKOWITZ, V., MATESE, J.C., PARKINSON, H., ROBINSON, A., SARKANS, U., SCHULZE-KREMER, S., STEWART, J., TAYLOR, R., VILO, J., VINGRON, M. (2001): 'Minimum information about a microarray experiment (MIAME) – toward standard for microarray data', *Nat. Genet.*, pp. 365-371
- [8] BALL C.A., SHERLOCK G., PARKINSON H., ROCCA-SERA P., BROOKSBANK C., CAUSTON H.C., CAVALIERI D., GAASTERLAND T., HINGAMP P., HOLSTEGE F., RINGWALD M., SPELLMAN P., STOECKERT C.J.JR., STEWART J. E., TAYLOR R., BRAZMA A., QUACKENBUSH J. (2002): 'Standards for Microarray Data', 298
- [9] STOECKERT, C. J. JR., CAUSTON, H. C., BALL, C.A. (2002): 'Microarray databases: standards and ontologies', *Nature Publishing Group*, pp. 469-473
- [10] SPELLMAN, P.T., MILLER, M., STEWART, J., TROUP, C., SARKANS, U., CHERVITZ, S., BERNHART, D., SHERLOCK, G., BALL, C., LEPAGE, M., SWIATEK, M., MARKS, W. L., GONCALVES, J., MARKEL, S., IORDAN, D., SHOJATALAB, M., PIZARRO, A., WHITE, J., HUBLEY, R., DEUTSCH, E., SENGER, M., ARONOW, B. J., ROBINSON, A., BASSETT, D., STOECKERT, C. J. JR., BRAZMA, A. (2002): 'Design and implementation of microarray gene expression markup language (MAGE-ML)', *Genome Biology*, 3
- [11] DUDOIT, S., GENTLEMAN, R. C., QUACKENBUSH, J. (2003): 'Open Source Software for the Analysis of Microarray Data', *BioTechniques*, 34, pp. 45-51
- [12] AMARATUNGA, D., CABRERA, J. (2004): 'Exploration and analysis of DNA Microarray and protein array data', *Wiley*