

LAS: a software platform to support oncological data management

*Original*

LAS: a software platform to support oncological data management / Baralis, E.M., Bertotti, A., Fiori, A., Grand, A.. - In: JOURNAL OF MEDICAL SYSTEMS. - ISSN 0148-5598. - ELETTRONICO. - 36:1(2012), pp. 81-90. [10.1007/s10916-012-9891-6]

*Availability:*

This version is available at: 11583/2502973 since:

*Publisher:*

Springer Science+Business Media New York

*Published*

DOI:10.1007/s10916-012-9891-6

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

# LAS: a software platform to support oncological data management

Elena Baralis · Andrea Bertotti · Alessandro Fiori · Alberto Grand

Received: 18 June 2012 / Accepted: 9 October 2012

**Abstract** The rapid technological evolution in the biomedical and molecular oncology fields is providing research laboratories with huge amounts of complex and heterogeneous data. Automated systems are needed to manage and analyze this knowledge, allowing the discovery of new information related to tumors and the improvement of medical treatments. This paper presents the Laboratory Assistant Suite (LAS), a software platform with a modular architecture designed to assist researchers throughout diverse laboratory activities. The LAS supports the management and the integration of heterogeneous biomedical data, and provides graphical tools to build complex analyses on integrated data.

## Author's version

Published in:

Journal of Medical Systems, November 2012 (Online version)

DOI: 10.1007/s10916-012-9891-6

URL: <http://www.springerlink.com/openurl.asp?genre=article&id=doi:10.1007/s10916-012-9891-6>

E. Baralis · A. Grand

Dipartimento di Automatica e Informatica, Politecnico di Torino, Corso Duca degli Abruzzi 24 10129, Torino, Italy  
E-mail: {elena.baralis, alberto.grand}@polito.it

A. Bertotti

Laboratory of Molecular Pharmacology, Institute for Cancer Research and Treatment (IRCC), Str. Prov. 142 Km. 3.95, 10060, Candiolo, Italy

A. Bertotti

Department of Oncological Sciences, University of Torino Medical School, Str. Prov. 142 Km. 3.95, 10060, Candiolo, Italy

E-mail: andrea.bertotti@ircc.it

A. Fiori

Fondazione Piemontese per la Ricerca sul Cancro-Onlus (FPRC), Institute for Cancer Research and Treatment (IRCC), Str. Prov. 142 Km. 3.95, 10060, Candiolo, Italy  
E-mail: alessandro.fiori@ircc.it

Furthermore, the LAS interfaces are designed to ease data collection and management even in hostile environments (e.g., in sterile conditions), so as to improve data quality.

**Keywords** Biological data management · Data integration · Web application

## 1 Introduction

In the post-genomic era with the advent of new technologies (e.g., microarrays), a huge amount of complex molecular data are generated with high throughput. The management of clinical, molecular and biological data is definitely a challenging task due to the complexity and heterogeneity of data. Moreover, the integration of heterogeneous molecular data with clinical and biological information is a major prerequisite to discovering the main causes of tumors.

Numerous efforts have been devoted to building management systems that assist researchers during laboratory activities. Many commercial Laboratory Information Management Systems (LIMS) are available, but they are usually very expensive and require significant human and economic resources to adapt to specific laboratory requirements [7]. The researcher community has also proposed different tools focused on the management and the analysis of different biological data. For instance, TreeSNPs [2] manages data generated through single nucleotide polymorphism (SNP) identification. Differently, Screensaver [12] supports the storage and the comparison of small molecule and RNA screening data. Other proposed LIMS focus on managing data associated with mouse colonies. For instance, MausDB [10] integrates standard mouse colony management, phenotyping workflow scheduling and mouse phenotyping

result data management. However, these systems are often mainly focused on managing few kinds of data related to a particular set of experiments or kind of source. Moreover, many of these tools do not currently allow the integration and analysis of experimental data.

From 2004, a number of tools have been developed under the Cancer Biomedical Informatics Grid (caBIG) project [8], addressing both the management and the analysis of biomedical data. For instance, caTissue provides biospecimen inventory, tracking, and basic annotation, while VISDA [13] allows clustering analyses on high dimensional and complex biomedical datasets. However, the end user data accessibility in caBIG tools may be compromised by the object model complexity and overly generic terminologies [9]. Even if these tools belong to the same project, the integration of experimental data is in many cases troublesome or unfeasible.

In this paper, we propose the Laboratory Assistant Suite (LAS) platform that assists the researchers in different laboratory activities. Its modular architecture allows managing different kinds of raw data (e.g., biological, molecular) and tracking experimental data. Each LAS module is tailored to handle specific activities or kinds of data, but it is plugged into a broader and uniform framework and thus integrates with the activities managed by other modules almost effortlessly. The platform can be easily extended by developing and integrating new modules, should the need arise to manage new laboratory activities. User interfaces are designed to be practical in hostile environments where researchers should minimize their interactions with the system (e.g., in sterile conditions). Furthermore, the platform provides a graphical query tool to integrate different resources and performs analyses in a grid environment in order to extract knowledge related to tumors. The paper is organized as follows. Section 2 describes the context in which the LAS has been studied and developed. Section 3 discusses the main features of the system architecture, while Section 4 presents the six modules that currently make up the system. Section 5 describes how the proposed system can support laboratory research in diverse real scenarios. Finally, Section 6 draws conclusions and presents future developments.

## 2 Context

A multitude of procedures are performed daily in research laboratories to analyze different biological and medical aspects of tumors, to the aim of discovering new knowledge and improving therapies. To model our environment, we analyzed a subset of procedures developed and performed in the research laboratories of our institution. In particular, we focused our attention

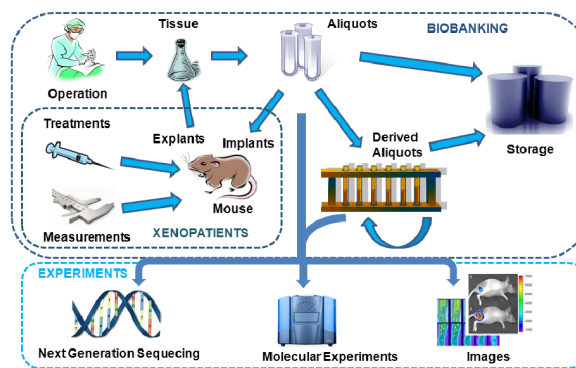


Fig. 1 Dataflow including the xenopatient approach.

on the procedures based on the *xenopatient* experimental pipeline [1]. This approach is based on the serial transplantation of human tumor specimens in immunocompromised animals. The aim is to help in translating the correlative information emerging from data integration into clinically relevant and functionally validated biomarkers. In the following we briefly introduce the context in which the LAS platform has been proposed by describing the main activities of a research study including the xenopatient approach (see Figure 1).

Tumor specimens are initially collected from surgical interventions. From each individual patient-derived material a set of aliquots are generated, i.e., vital, RNA Later and snap-frozen. According to the characteristics of the aliquots and the purpose of the research study, different operations can be performed: (i) storage in dedicated containers, (ii) extraction of derived aliquots (e.g., DNA, RNA), and (iii) implantation in immunocompromised animals (i.e., xenopatient). Researchers can define treatment protocols to monitor the evolution of the tumor mass within the implanted animals by means of measurements, and evaluate the response to drugs. Furthermore, they can generate new tissue samples from these animals for further analyses. Indeed, all (derived) aliquots can be exploited for experimental analyses with different technologies, such as microarray and DNA sequence experiments.

To the best of our knowledge, the procedures adopted in our experimental pipeline are largely standardized and they reflect common practice in the oncological research field. Thus, we believe that most functionalities offered by the system we are developing could be useful to other researchers from different institutions. The LAS software platform is freely available from the authors for use by interested researchers.

### 3 Architecture

The LAS architecture is the result of a joint effort by both IT and biomedical researchers, in an attempt to design a system to effectively manage the entire experimental and analysis process. Since the laboratory-related procedures can be subdivided into different levels according to data complexity and purpose, the LAS has been designed as a three-tier architecture. Each tier addresses a specific issue and interacts with the other tiers by means of well-defined interfaces. We modeled the platform in the following tiers: (i) operative, (ii) integration, and (iii) analysis.

In the general architecture of the LAS, each tier includes a set of fully-fledged applications, or modules. While the lower tier is mainly concerned with the collection of experimental data, the modules and data managed by the upper tiers are characterized by an increasing level of abstraction. Lower tiers can service requests generated by the upper tiers and provide the data needed to carry out complex tasks (e.g., data integration and/or analysis).

The *operative* tier is responsible for collecting, storing and tracking raw experimental data from different sources (e.g., xenopatients, microarrays), each handled by a specific software module. Modules in this tier are meant to work in close interaction with the researchers in a laboratory setting. Therefore, graphical user interfaces (GUIs) are tailored to ease data entry in hostile environments and assist researchers throughout their experiments in order to improve data quality. Interaction is designed to be especially lean with the aid of special input devices, such as touch-screen notepads and barcode scanners, available in our research institute.

The *integration* tier is aimed at integrating different types of raw experimental data by means of complex queries. Ad-hoc identifiers have been adopted throughout the databases that make it possible to correlate different biological entities. The integrated data can be browsed or visualized as graphs (e.g., genealogy trees). In addition, they can be fed to the analysis tier and enriched by means of annotations. For instance, a population of xenopatients can be annotated as responsive to a given drug according to statistical analyses.

The *analysis* tier allows the definition of a workflow for the analysis of integrated data. Users can design their analysis session by selecting and cascading one or more analysis building blocks. The analysis process ultimately generates annotations, which are sent to the integration tier and stored in the database, and it can optionally export data for visualization with external tools.

Each LAS module is a web-based application implemented in Python and exploiting Django <sup>1</sup>, an open-source web application framework which follows the Model-View-Controller [5] architecture. In this way, end users can access the applications from virtually any Internet-enabled device and without almost any system requirements. This aspect is fundamental to provide data access also in hostile environments (e.g. sterile conditions) with wireless devices. According to the data structure and type, a relational DBMS (MySQL <sup>2</sup>) and/or a non-relational database engine (MongoDB <sup>3</sup>) are exploited to provide persistent storage for the experimental data and the annotations. Different user privileges can be defined in order to grant or restrain access to different application functionalities. Furthermore, all user activity is tracked within each module by means of audit trail procedures. A set of APIs (Application Programming Interface) are provided to make some of the functionalities offered by each application available to the others and thus permit data exchange among different applications. Finally, thanks to the modular architecture, different modules can be installed on different machines, so as to distribute the workload of the entire system. A grid environment is also being integrated to increase the computational resources available for the analysis tasks.

### 4 Modules

The LAS platform is designed to be easily updated when new laboratory or analysis activities need to be managed. To date, the BioBanking, Xenopatients, and Storage Management Modules have been fully implemented and deployed for use by the research staff. Most functionalities of the Multi-Dimensional Data Manager are already available, while the Microarray Management Module is currently under testing. The Analysis Developer Module is still being developed. In the following we present the main functionalities of each module.

#### 4.1 BioBanking Management Module

With the rapid advances in biomedical and genetic technologies, collections of human biological materials have attracted increasing attention from the researcher community. Such collections, named biobanks, represent a fundamental resource for the research and the diagnosis of different pathologies, and the study of possible therapeutic applications. Based on the types of biological

<sup>1</sup> <https://www.djangoproject.com/>

<sup>2</sup> <http://www.mysql.com/>

<sup>3</sup> <http://www.mongodb.org/>

materials they store, biobanks are commonly divided into tissue and genetic biobanks. The BioBanking Management Module (BBMM) addresses both issues.

The scope of the BBMM spans a wide range of activities, including management of biological samples and laboratory instrument data, as well as support to a number of laboratory-related procedures. In particular, the module can currently handle the following activities and the corresponding data.

(i) Collection of biological material from surgical intervention and acquisition of aliquots from external laboratories. Aliquots stored in the system are characterized by features such as tumor type (e.g., colorectal), tissue type (e.g., liver metastasis), number of fragments and source hospital or laboratory.

(ii) Measurement of aliquot physical characteristics, such as volume, concentration, purity and quality.

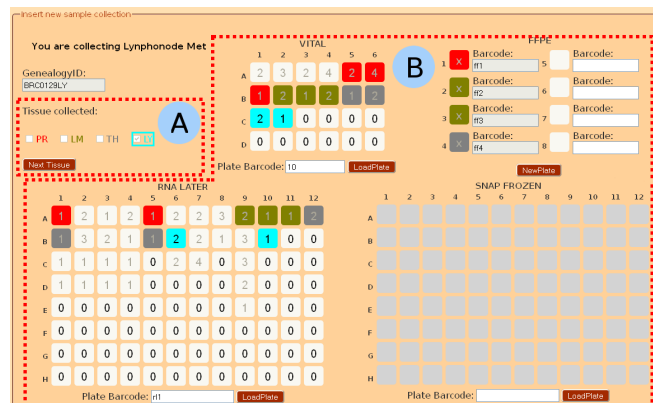
(iii) Derivation of new biological materials (e.g., DNA, cDNA) from the aliquots stored in the biobank.

(iv) Tracking of consumables involved in laboratory operations, to manage the warehouse and monitor associated quality indexes.

Particular care has been devoted to the design of the graphical user interfaces, so that they closely match the researcher's physical working environment. They have been designed to also work on tablet devices, to allow use in hostile environments, such as in sterile conditions or during surgical interventions. For instance, Figure 2 shows the aliquot collection interface. The application shows several panels, i.e., vital, FFPE, RNA later and snap-frozen (block B in the figure), that graphically represent the physical working containers in which sample aliquots will be stored. Containers of each type can be quickly loaded by scanning their barcode, or by manually typing it. The layout of each container is automatically retrieved from the Storage Management Module. Each cell, representing an aliquot, is colored according to the type of tissue, and reports the number of aliquot fragments it currently stores. Clicking (or tapping) on an empty cell stores one fragment of aliquot of the current tissue type in the corresponding container location; each further click (or tap) increments the fragment count by one. The tissue type being collected is shown on the top left hand side (block A).

## 4.2 Storage Management Module

Research laboratories make use of several types of containers (e.g., freezers, racks, plates, tubes) to store biological material. The main issue in designing a storage management system is to model how containers can interact with each other. For instance, a freezer of a given



**Fig. 2** The BBMM aliquot collection interface.

manufacturer and model can host different racks and/or plates. The inner layout of the freezer may also change according to the type of hosted containers.

This issue is addressed in the Storage Management Module through a flexible database schema. Different container types can be defined, along with a geometry, describing how child containers are laid out, a parent container (if any), and a set of relevant features that will be recorded for each container of that type (e.g., manufacturer, model and storage temperature for freezers, aim and aliquot type for plates).

The module shows the current usage of each container, and allows the end user to move containers and change their aim. All the modules belonging to the LAS can easily retrieve the information associated with a given container through the provided API, using its unique identifier, such as a barcode. In this way, laboratories can exploit barcode readers to streamline data entry procedures.

## 4.3 Xenopatients Management Module

One of the crucial activities in our institution, management of the immunocompromised animals (i.e., xenopatients) is supported by this module. It enables tracking and monitoring of the xenopatient life cycle, from their acquisition by the research institute to their death. The functionalities provided by this module include the following.

(i) Loading of data (i.e., barcode and gender) for a batch of mice (characterized by the same status, strain, age, source, and owner) into the system. Mice can be equipped with RFid tags, allowing quick insertion of each animal into the database and easy identification at a later time.

(ii) Management of tumor tissue implants into the available xenopatients. The XMM effectively handles this operation by accessing the BioBanking and Storage

modules (through the appropriate APIs) and retrieving the tumor aliquots currently stored in the source container (e.g., plate, tube), whose available quantity is updated accordingly.

(iii) Monitoring of tumor growth within the xenopati-ents. Qualitative and quantitative measurements of the tumor mass are tracked for each mouse, and treat-ments, aimed at assessing the effectiveness of given drugs, can be assigned to measured xenopati-ents. Each treat-ment is composed of several phases, each associated with different kinds of information. For instance, the drug, the administration mode, the dose, and the ad-ministration frequency can be defined.

(iv) Management of tumor explants, marking the end of the xenopati-ent life cycle. A number of tumor tissue aliquots are generated and made available for fu-ture implants. Explant operations also cooperate with the BioBanking and Storage Management Modules, to display destination plates and transmit data about gen-erated aliquots.

During the xenopati-ent life cycle, the user can as-sign each mouse different operations to be performed, which can be subsequently validated by a supervisor. The module supports the supervisor in the decision-making process by means of graphical utilities, and tracks all decisions (see Section 5.1).

#### 4.4 Microarray Management Module

Microarray technology has become a fundamental tool in genomic research to analyze thousands of gene ex-pression values with a single experiment [4,6]. The anal-ysis of this huge amount of data aims to identify the set of genes that are highly correlated with the phenotype. Thus, tracking both the biological information related to the sample source and the entire experimental pro-cess is necessary to assess the quality of the analyses and improve further integration tasks. The Microarray Management Module tracks all information related to the various steps of a microarray experiment, providing support to the following tasks.

(i) Planning of different experiments by assigning each biological sample a position within a microarray chip. Automatic assignment is also possible to optimize the chip exploitation.

(ii) Hybridization of chips according to specific pro-tocols, tracked by the system.

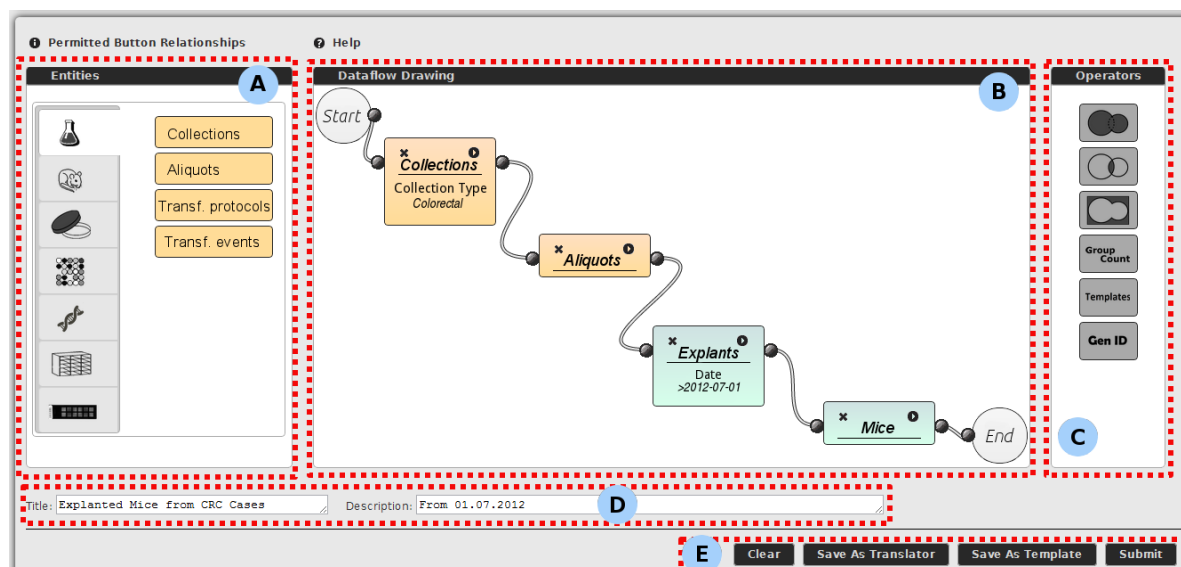
(iii) Scan events. Based on additional data provided by the chip manufacturer, the system generates an in-put file for the microarray scanner machines. Quality measures about the scan event are also tracked for fur-ther analyses.

(iv) Collection and storage of the microarray raw data obtained from the scans. Such data are stored in a content repository and linked to the corresponding sample within the database. To automatically feed the content repository and update the database each time a scan is performed, we provide a client/server tool writ-ten in Java. The client program retrieves the microarray raw data produced by the scanners and sends it to the server hosting the content repository. When new data are received by the server, the relevant information is extracted to feed the content repository, and links to the raw data are added to the database.

#### 4.5 Multi-dimensional Data Manager

The integration of different biological and molecular data is a fundamental task to discover new knowledge related to tumors. The Multi-Dimensional Data Manger (MDDM) aims at extracting all information of interest from the operative tier in a uniform way by exploiting graphical tools. In the following we present the main ideas on which the functionalities are built.

The MDDM allows performing heterogeneous and complex queries on the data managed by the opera-tive tier modules. The user defines the result set of each query, designed as a workflow, by cascading dif-ferent query blocks. For each query block, objects of the corresponding type will be retrieved (e.g., aliquots, xenopati-ents), based on the filtering conditions speci-fied by the user. All SQL queries actually built by the system to process each query block are hidden to the final user. The MDDM query editor assists the user in designing her query by performing coherency checks at run-time and forbidding operations that are not well-defined (e.g., cascading blocks corresponding to unre-lated object types or intersecting disjoint sets of ob-jects). To ease the use of the system by unexperienced users, a set of wizards based on predefined workflows are also provided. In Figure 3 the query editor screen is reported. The query blocks are shown on the left hand side of the editor (block A in Figure 3) and cat-egorized according to the module from which the data are drawn (e.g., flask icon for biobank data, mouse icon for xenopati-ent data). The user can drag and drop the query blocks onto the workflow editor (block B) and build the workflow of the query by connecting them. Filtering conditions may be specified for each block by clicking the arrow button. Set operators, listed in block C, can also be used in the workflow. Once the workflow has been defined, the user can assign a title and a de-scription to the query (block D). Finally, she can simply run the query or save it in the system to be reused in the future for different purposes (block E).



**Fig. 3** The MDDM query editor interface.

Each object returned by a query is inserted in a non-relational database. With this technology, heterogeneous objects can be stored without defining a fixed structure for the information. Thus, MDDM can easily store different biological and molecular objects and add other information (i.e., annotations) derived from further analyses. For instance, researchers can annotate a set of objects with their response to a drug treatment, according to statistical analysis performed on corresponding molecular and biological data. The integrated data can also be browsed or visualized by means of graph representations (e.g., genealogy trees).

Communication among MDDM and the other modules exploits the APIs defined by each module. Each API of interest for the integration task, and its related parameters, are registered into MDDM, so that a corresponding query block is created and made available in the editor. Since some filter parameters may change over time, or new operative modules may become available, a procedure is included to update the MDDM configuration and avoid inconsistencies.

#### 4.6 Analysis Developer Module

Genomic data encode many biological insights whose investigation can be the basis for dramatic advances in the field of molecular oncology. The analysis of different (integrated) biological and molecular data is fundamental in any research laboratory. Several tools have been proposed with different purposes. However, to the best of our knowledge, a tool that can integrate many different analysis methods is missing. The Analysis Developer Module aims to provide the end user with a web-

based tool to define an analysis workflow. The graphic interface is based on the same idea exploited in the MDDM for the query construction. In this way, the user can focus her attention on the analysis tasks and needs not concern herself with any issue related to data formats and conversions. Each block represents an algorithm integrated in the platform, while data exchange among the blocks and the collection of results are handled by the Analysis Developer Module in a transparent way.

To enhance the computational resources available for analysis, we are currently building a grid environment based on the Apache Hadoop framework<sup>4</sup>. It allows the distributed processing of large data sets across clusters of computers, using a simple programming model.

## 5 Use cases

The following sections illustrate how the LAS platform can concretely support laboratory research. We present two real application scenarios we encountered in our laboratories, and describe the LAS tools addressing these issues. Even though the idea and the design of such tools were based on the functional requirements of our laboratories, we believe they reflect common practice in the oncological research field and are general enough to be useful to other researchers from other institutions.

<sup>4</sup> <http://hadoop.apache.org>

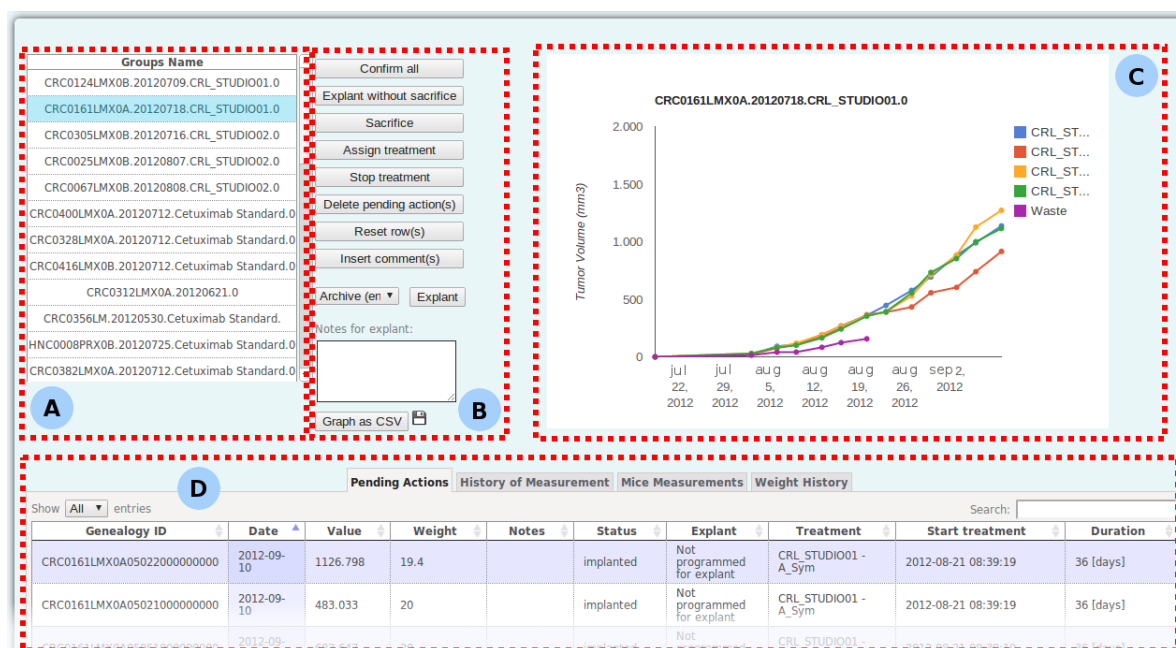


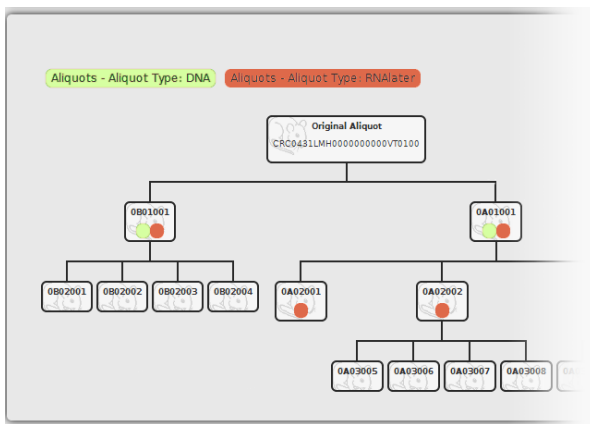
Fig. 4 The XMM experiment review interface.

### 5.1 Use case 1: Experiment review

Immunocompromised mice that have been implanted with a tumor mass are regularly measured by laboratory technicians to monitor the growth of the tumor and assess the effectiveness of treatments. Through the LAS interface, technicians record the observed values and may propose a given course of action for each mouse (e.g., start/stop a treatment, explant). Such measurements and actions are reviewed by a supervisor through the “Experiment review” screen (see Figure 4). The screen initially displays a list of all experimental groups for which new measurements exist (block A). When an experimental group is selected, its measurements and all associated pending actions are shown in the bottom table (block D). To ease the supervisor’s task, any pending action awaiting approval is highlighted, while further information is available in the other tabs (e.g., historical measurement record for each mouse). Moreover, a plot showing the average variation of the tumor mass over time for each treatment arm (block C) provides a useful tool to support the supervisor in the decision-making process. The supervisor may accept the actions proposed by the technician for any measurement, or she may modify or reject them, by means of a set of buttons (block B). She may also add comments, such as explanations or future directions. Once the review is complete, a report is displayed, including all actions (both approved and rejected or modified ones), and sent to both the supervisor and the technician.

### 5.2 Use case 2: Paternity test

In the xenopatient experimental pipeline, we define the xenogenealogy as the genealogy tree representing the relationships among the xenopatients implanted with original or passaged tumors. To support the robustness and the predictive power of this experimental procedure, serially passaged tumors should retain the morphologic and genomic features of their original counterparts. Thus, the researchers in our institution perform mutational analyses on both a representative series of matched primary and xenografted cases. The LAS can automatically select the aliquots, explanted from the xenopatients belonging to a given xenogenealogy, to perform these paternity tests. The selection procedure is scheduled to be automatically performed for each xenogenealogy stored in the system twice per month, but it can also be triggered by the researcher on a specific case. For instance, the user can retrieve the genealogy of interest by performing a query with a set of constraints using the MDDM module. Figure 3 shows a query retrieving the mice explanted in a given time period (i.e., after July, 1<sup>st</sup> 2012) and belonging to a colorectal tumor collection. The workflow of the query starts retrieving all colorectal tumor collections (Collections block). This set of collections is exploited to filter the aliquots in the biobank (Aliquots block) and select the explants from mice, performed after July, 1<sup>st</sup> 2012, that have generated these aliquots (Explant block). Finally, the mice involved in these explants are returned to the user (Mice block). The result set is pre-



**Fig. 5** Xenogenealogy with marked xenopatients.

sented as a table reporting the features of each mouse. By selecting the genealogy identifier associated with a given mouse, additional information can be retrieved by means of query templates and visualized on the xenogenealogy. For instance, in Figure 5, mice whose DNA and RNALater aliquots are stored in the biobank are marked on the xenogenealogy. Finally, the user can ask the system to identify the aliquots for the paternity test on the xenogenealogy. The selection of a set of representative xenopatients is exploited to cover all the branches of the genealogy tree and identify xenopatients from which DNA has not yet been extracted (i.e., on which the paternity test has never been performed before). In both scenarios (manual and automatic selection of candidates for the paternity test), an email with the set of identified aliquots and the associated xenogenealogy with marked mice is sent to the researcher and her supervisor.

### 5.3 Usage statistics

The first modules of the LAS platform have been actively employed for research in our institution since March 2012. To date, about 15 users exploit the BioBanking, Storage and Xenopatients Management Modules in their daily activity. As of September 2012, the BioBanking database thus includes about 30,000 aliquots of different types (e.g., RNALater, SNAP frozen, DNA, RNA) belonging to 560 collections, and tracks the collection of tens of new aliquots every day. The Storage Management module manages around 200 containers, while the Xenopatients Management Module is tracking all the activities performed on approximately 2,900 mice.

By adopting the LAS, the time required for retrieving and analysing data has significantly decreased. For instance, to perform the experiment review (described in Section 5.1), the supervisor would previously need to

manually identify and examine all measurements performed by different technicians to make his final decisions, which should then be transmitted to the technicians. This would generally require from half to a full working day. With the current system, which automates data retrieval, aggregation and communication, the same task only takes a few minutes.

## 6 Conclusions and future work

In this paper we presented the Laboratory Assistant Suite (LAS) platform, designed to assist the researchers of biological and biomedical laboratories in their activities. The modular architecture of the system can manage heterogeneous and complex data and perform different analyses. Moreover, the graphical interfaces and the web-based architecture enable the researchers to use the platform even in hostile environments (e.g., in sterile conditions). Some operative modules are already in use in the research laboratories of our institution, with promising results.

Besides completing the MDDM with annotation management capabilities, we are developing other operative modules of the LAS platform. In particular, we are designing a SangerSeq Management Module to handle sequencing experiments, track consumables and biological materials (e.g., primers), and define quality controls, and a Cell-line Management Module to track the life cycle of cell-lines and monitor the experimental conditions applied to each cell. Furthermore, we are interfacing some state-of-the-art analysis approaches with the Hadoop framework to increase the computational power available for the Analysis Developer Module. These methods will be applied on integrated data, including microarray and xenopatient data.

## 7 Conflict of Interest

The authors declare that they have no conflict of interest.

## References

1. Bertotti A., Migliardi G., Galimi F. et al., A molecularly annotated platform of patient-derived xenografts ('xenopatients') identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer, *Cancer Discovery* (2011)
2. Clément S., Fillon J., Bousquet J. et al., TreeSNPs: a laboratory information management system (LIMS) dedicated to SNP discovery in trees, *Tree Genetics & Genomes*, 6, 3, 435-438 (2010)

3. Eckerson W. W., Three Tier Client/Server Architecture: Achieving Scalability, Performance, and Efficiency in Client Server Applications. *Open Information Systems*, 10 (1995)
4. El Akadi A., Amine A., El Ouardighi A. et al., A two-stage gene selection scheme utilizing MRMR filter and GA wrapper, *Knowledge and Information Systems*. (2010)
5. Gamma E., Helm R., Johnson R. et al., *Design patterns*, 1. Addison-Wesley Reading (2002)
6. Golub TR, Slonim DK, Tamayo P. et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science*, 286, 5439, 531 (1999)
7. Haquin S., Oeuillet E., Pajon A. et al., Data management in structural genomics: an overview, *METHODS IN MOLECULAR BIOLOGY-CLIFTON THEN TOTOWA-*, 426, 49 (2008)
8. Kuhn, K. et al., The Cancer Biomedical Informatics Grid (caBIG): Infrastructure and Applications for a Worldwide Research Community, *Medinfo 2007*, 1, 330 (2007)
9. London J.W. and Chatterjee D., Using the Semantically Interoperable Biospecimen Repository Application, caTissue: End User Deployment Lessons Learned, *IEEE International Conference on BioInformatics and BioEngineering*, 316-317 (2010)
10. Maier H., Lengger C., Simic B. et al., MausDB: An open source application for phenotype data and mouse colony management in large-scale mouse phenotyping projects, *BMC bioinformatics*, 9, 1, 169 (2008)
11. Mierswa I., Wurst M., Klinkenberg R et al., YALE: Rapid Prototyping for Complex Data Mining Tasks, in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2006)
12. Tolopko A., Sullivan J., Erickson S. et al., Screensaver: an open source lab information management system (LIMS) for high throughput screening facilities, *BMC bioinformatics*, 11, 1, 260 (2010)
13. Wang J., Li H., Zhu Y. et al., VISDA: an open-source caBIG analytical tool for data clustering and beyond, *Bioinformatics*, 23, 15, 2024 (2007)