



HAL
open science

SITRANS: a web information system for microarray experiments

Frederique Laforest, A. Tchounikine, T. Chaari, H. Charles, Federica Calevro

► **To cite this version:**

Frederique Laforest, A. Tchounikine, T. Chaari, H. Charles, Federica Calevro. SITRANS: a web information system for microarray experiments. *Studies in Health Technology and Informatics*, 2005, 116, pp.33-38. hal-00391451

HAL Id: hal-00391451

<https://hal.science/hal-00391451>

Submitted on 30 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SITRANS: a Web Information System for Microarray Experiments

Frédérique Laforest^a, Anne Tchounikine^a, Tarak Chaari^a, Hubert Charles^b, Federica
Calevro^b

^aUMR CNRS 5205 LIRIS Laboratoire d'InfoRmatique en Images et Systèmes d'information
^bUMR INRA 203 INSA de Lyon BF2I Biologie Fonctionnelle, Insectes et Interactions
INSA de Lyon, F-69621 Villeurbanne Cedex, FRANCE

Abstract

Microarray experiments aim at analyzing expression levels of genes using DNA probes. The amount of data managed for each experiment is very large. It is thus essential to provide electronic support for the capture and the management of information describing microarray experiments. We present here the SITRANS Web information system, the aim of which is to help research workers storing, browsing, sharing and publishing data.

Keywords

Information Systems, Microarray Analysis, World Wide Web

1 Introduction

Microarray analysis has become a widely used tool for the generation of gene expression data on a genomic scale. The amount of data managed for each experiment is very large. The steps used in a microarray experiment are numerous and various : optimization of probes positioning on the slide, preparation of slides surface, spotting, purification and labeling of messenger RNA, hybridization, washing, signal measurement, normalization and data analysis. The novelty of the technique makes that interesting data are also not well known nor normalized. The description of an experiment and moreover the comparison between experiments are thus difficult and often impossible. The reproduction of published experiments is also impracticable as many details are often not given or not precise enough. The MIAME model [1] has provided a big advance on the standardization for presenting and exchanging microarray data. But the fields it provides are not yet formatted (they are often textual descriptions) and the precision of information can thus differ from one MIAME description to the other.

The complexity of the process and the volume of data to be managed require the use of a computerized information system for the management of experiments, their publication and the exchange of results. In this context, the SITRANS project aims at answering three main objectives:

- Allow research workers to manage their experiments i.e. to have a history of each experiment step, from the microarray design to its analysis,
- Allow result publication in a standardized format (MAGE-ML [2]),

- Build a common database for all researchers of the platform in DTAMB (University Claude Bernard Lyon I) so that they can share both their know-how and their data.

To do so, we have developed and tested SITRANS, a Web information system that could be defined as a shared and electronic lab booklet. It does not include analysis tools, it is only dedicated to the storage and the retrieval of information about experiments. Data stored range from raw data on spotter configuration to scanned images, including results of normalization and analysis. It is based on a 4-tier architecture.

The next section presents the SITRANS information system. After a brief presentation of the technical architecture, we present the way we have defined the database schema and the functional view of the system. Section 3 then presents the navigation tool we have written, based on the topic maps theory.

2 The SITRANS information system

The SITRANS information system aims at capturing, storing, retrieving and publishing data on microarray experiments. The first and non trivial work has been to design the database schema. The second step has concerned the functional view of the application. We have collaborated a lot with end-users for these two first steps, and worked with simulations and demonstrators. We thus have designed the overall application architecture and developed it. The first subsection shows rapidly the technical environment we have chosen. The following subsections detail each step.

2.1 Technical environment

The SITRANS information system is a Web application. It is based on a multi-tier architecture. The database management system is PostgreSQL. It interacts with EJBs on a jboss server. EJBs are divided into two layers: entity EJBs represent the database objects, while session EJBs represent the application functionalities. The user interface uses JSP pages that work in collaboration with session entities, using Data Transfer Objects (DTOs) [3]. As a matter of fact, as data exchanged between JSPs and EJBs are numerous, we have chosen to build DTOs that minimize the number of transfers (even if each transfer is heavier). JSP pages are managed by a tomcat server, coupled with an apache HTTP server that serves HTML pages to the end-user. HTML pages include some javascript code for captures validation before transfer to JSP pages.

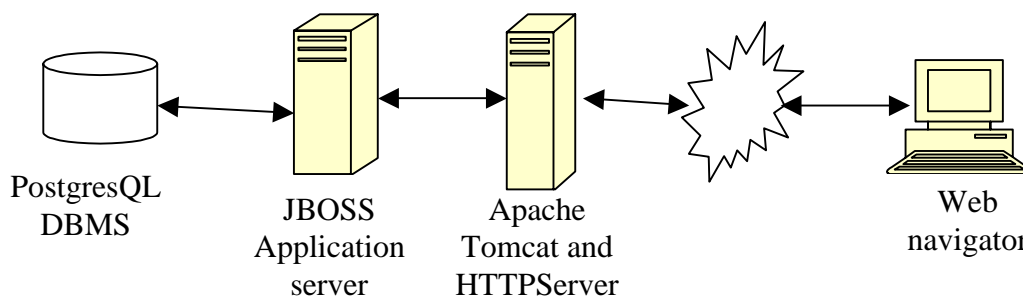


Figure 1: Technical architecture of SITRANS

2.2 Database schema

The design of the database schema was initially developed in 2002 and has been made on two bases:

- first on the previous work made by Pr. Barbry Team in Nice, France (CNRS UMR 6097). His team gave us the database schema they built for their microarray experiments, called MedLab [4]. Their database relies on PCR microarrays.
- second on experiments done on the DTAMB platform located in Lyon. It has been designed in the objective to rapidly be accessible to the public research staff.

We thus have modified the Medlab database schema. The database we have built contains data on the experiment environment : it provides the list of scanners, spotters, analysis tools, experimental protocols and so on... that can be used on the DTAMB platform; it also manages the localization of new and used slides, allowing for a basic stock management. We have also modified the Medlab Schema so that it can manage both oligonucleotide-based and PCR-based experiments.

We have also defined the concept of project: a project concerns one or more experiments on the same slide model. A slide model defines the slide spots and their contents. Most projects use duplicated or quite similar slides that follow the same model, for validation purpose.

In this article, we cannot provide the database schema, as it contains 44 tables. It is available at the following address: http://liris.cnrs.fr/~sitrans/sitrans_v2.3_create.sql. The resulting schema conforms to the MIAME standard, and extractions to the MAGE-ML format will be automatized soon.

2.3 Functional view of the application

The functional view of the application has been built in close collaboration with end-users. We have proposed many different versions and discussed the overall functioning as well as details. Meetings made conclusions on the different pages to be shown, as well as their content, positioning of elements, navigation between pages and so on. We have concluded with a user interface organized as follows :

- The data capture menu follows exactly the steps of a typical project: project description, array design (slides, biochip/slide model and probes), target preparation (samples, extracted products, amplified products, labeled products), hybridization, and biochip analysis (raw images, raw data, and normalized data). We have thus designed a page for each step. Figure 2 shows on the left the list of available pages, and on the right a page being edited. A popup page allows for getting more details.
- Precedence rules have been defined for some steps, others may be filled in any desired order. The practice proved that the “logical” precedence of steps in a typical experiment are quite always followed. As soon as the data recording step is performed, a researcher has the possibility to “freeze” it, i.e. nobody can modify the project data anymore. A researcher may also “publish” project data, i.e. make them available to other end-users of the system.
- Each time it is possible, the hand-based capture is minimized. We have implemented it in two ways: first the integration and interpretation of robots description files (e.g. spotting map) with use of mediators, and second the choice of values in lists of available tools (protocols, machines and so on coming from the database itself).
- Lists of available tools are managed by the platform administrator, who can update them in function of the platform evolution. We have called them platform data, in

opposition to experiments data, that are created and managed by research workers themselves.

- We have defined 3 types of end-users: the platform administrator manages platform data ; the research worker creates and works on microarray projects, the visitor can explore published data.

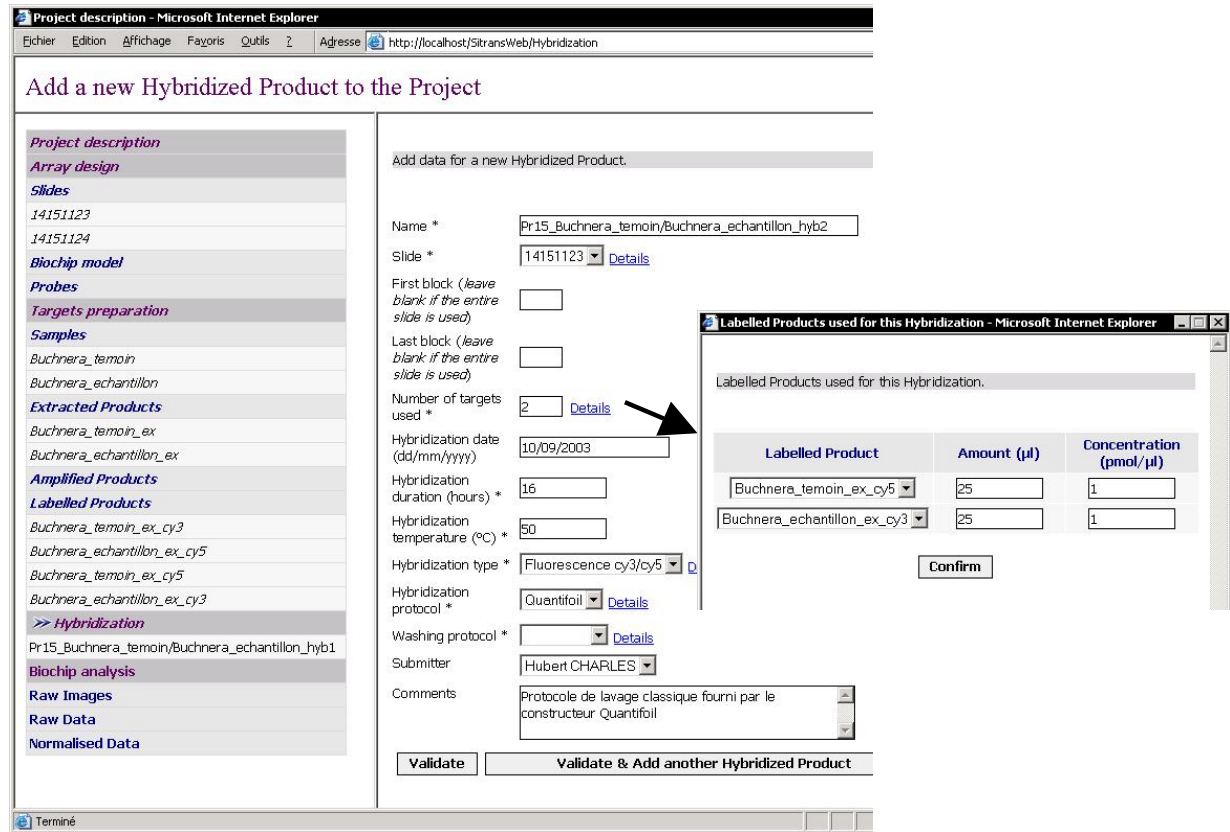


Figure 2: Capture page of an hybridized product

- The data consultation can be made in two ways. First a standard way allows for routing among capture pages of the different steps of the project. Pages are exactly the same ones as for capture: the only difference is that no data may be modified nor deleted. Second we have defined a navigation way, using a map of data. This is explained in more details in the following sub-section. No classical SQL-based querying has been implemented at the moment. As a matter of fact, end-users were first not able to define query forms embedding fixed SQL queries. Second, they are not enough familiar with the SQL language for writing themselves queries on the database

3 Data visualization using topic maps

Data visualization using capture forms is not flexible. That's why we have chosen to provide another visualization paradigm. It is based on navigation in a data map and has been developed as an applet. The data map is based on the topic maps theory [5]. In the navigation tab (see figure 3), we present topics to the end-user. In this first version, topics correspond to tables in the database schema, but logic predicates can be used to define topics as selections on the database [6]. Each topic is assorted with the number of corresponding occurrences in the database. The selected topic is also linked to other related

topics. The values associated to the related topics provide the number of occurrences corresponding to the selected topics occurrences. Changing the selected topic is easy: a click on the desired topic centers it in the middle of the tab and provides its neighbor topics.

Changing to the “selection” tab, the end-user can make a selection of occurrences in the selected topic. The selection can be made on the list of occurrences or with constraints on columns values. The navigation tab then updates its occurrences counts in consequence.

A third tab (“history”) provides the history of navigation. Buttons “precedent” (back) and “suivant” (next) allow for moving in the navigation history.

The data visualization is thus based on both navigation in topics and selection of instances.

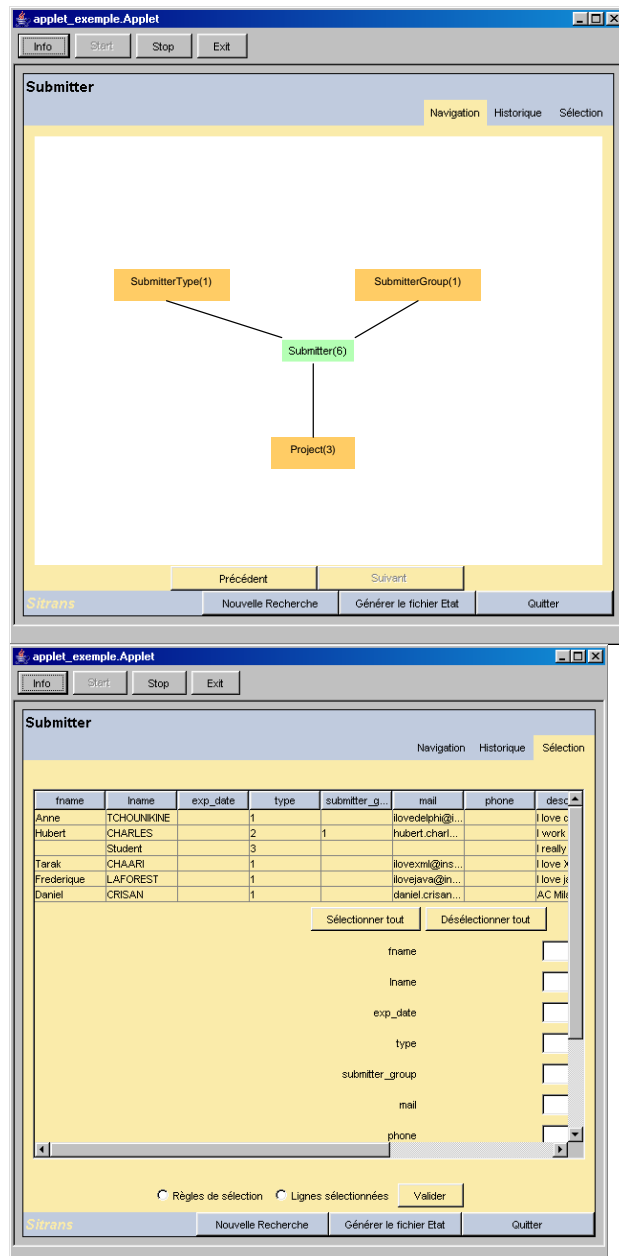


Figure 3: Applet for navigation in the data map

4 Conclusion

The SITRANS information system intends to help research workers, for them to organize, store and retrieve easily data on their microarray experiments. We have followed an agile

development process [7] (including the client very frequently and at all steps) to ensure the appropriateness of the resulting work. This information system is operational since May 2004 on the intranet of the Lyon I University campus. Since then, tests with large real experiments have been conducted by research workers and the tool is thus validated. At the moment, SITRANS is on the private network of the university campus and cannot be accessed by external users. We intend to provide access soon on the Internet, on <http://sitrans.insa-lyon.fr> (name under creation).

For a larger use on a more efficient computer, we intend to provide it on the PRABI web site soon [8]. But before, we have to finalize the navigational consultation user interface, whose first beta version is under testing procedures at the moment. It is intended to be provided to local end-users for usability tests before end of February 2005.

Another point we have to rapidly implement concerns the MAGE-ML extraction of data. We intend to couple this extraction with the navigational consultation interface, so that end-users can select data to be extracted during a navigation process.

5 Acknowledgments

This work has been supported by the french Ministry of Research by the ACI IMPBio program of September 2003. Thanks to the support of INRIA to the Rhône-Alpes microarray platform, a computer science engineer (Daniel Crisan) has also worked on the project during one year from may 2003 to April 2004 for the design and the implementation of the information system. Students from the BioInformatics Department of INSA Lyon and from the Telecommunications Department of INSA Lyon also participated in prototyping and developments. We also want to thank the biologists and technicians working on the Rhône-Alpes microarray platform for their help and advice.

6 References

- [1] Bramsa et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data, *Nature. Genetics*, 29(4), pp365-371, 2001
- [2] Paul T Spellman et al. Design and implementation of microarray gene expression markup language (MAGE-ML) *Genome Biology*, 3(9), pp. research0046.1-0046.9, 2002
- [3] Martin Fowler *Patterns of Enterprise Application Architecture*, Addison Wesley, 2002
- [4] Plate-forme microarrays/transcriptome de Nice/Sophia-Antipolis <http://medlab.ipmc.cnrs.fr/>. Last visited jan 11, 2005
- [5] ISO/IEC 13250:2000 Topic Maps: Information Technology -- Document Description and Markup Languages, Michel Biezunski, Martin Bryan, Steven R. Newcomb, ed., 3 Dec 1999
- [6] OUZIRI, Mourad. Utilisation des Topic Maps pour l'interrogation et la génération de documents virtuels : Application au domaine médical. Thèse : Institut National des Sciences Appliquées de Lyon, 2003, 220 p.
- [7] Scott W. Ambler, Ron Jeffries *Agile Modeling: Effective Practices for Extreme Programming and the Unified Process*, John Wiley & sons, 2002
- [8] Prabi web site <http://www.prabi.fr/>

7 Address for correspondence

Frederique Laforest
LIRIS, Bat Blaise Pascal
INSA de Lyon
F-69621 Villeurbanne Cedex
France
frederique.laforest@insa-lyon.fr