

NetBookingLIMS : automation of workflow within core genetics laboratories using a novel Internet-based information management system

Rifat Hamoudi, MSc, MBiol, CBiol, MBCS, CEng

Cancer Gene Cloning Centre,
Haddow Laboratories,
Institute of Cancer Research,
Sutton, SM2 5NG,
United Kingdom.
Tel : 0208-643 8901
Fax : 0208-770 7290
Email : rifat@icr.ac.uk

ABSTRACT

NetBookingLIMS is an Internet-based software used for the total automation of workflow within core diagnostic or genetics laboratories. NetBookingLIMS organises all data incoming into the laboratory in a flat file and converts it to a relational database that can be used to generate auditing, billing, troubleshooting and usage reports. Also, it generates instrument compatible data and organises the day-to-day management of the laboratory in an efficient and structured manner. It is written using Perl5 and C and designed to run solely using an Internet browser. NetBookingLIMS software as well as a user manual and documentation are available at <http://dcs.bbk.ac.uk/~ahamo01/nblfrontpage.cgi>.

INTRODUCTION

Pharmaceutical, commercial and academic scientific laboratories deal with enormous amounts of data. Whether this information is related to process, procedures, experimental design or analytical and clinical data, keeping it safe and organized - yet available to those who need it - is a formidable bioinformatics challenge. A core DNA sequencing and genetic analysis laboratory supports many research teams within an organisation and is responsible for the cloning and characterisation of genes. Usually a genetics laboratory employs many staff managed by one manager, has many instruments, supports many applications such as DNA sequencing, genotyping, fluorescent fingerprinting¹ and fluorescent mutation detection² and is used by many users (scientific researchers) with different requirements including high, medium and low throughput usage. Also it is difficult to carry out accurate auditing, billing and troubleshooting with such a complex working

model. Thus, within such a laboratory there are many real world issues such as managerial and organisational problems and algorithmically complex problems such as samples scheduling and relational database administration and maintenance.

Several commercial genetics laboratory LIMS (Laboratory Information Management System) already exist, such as BioLIMS (http://www.appliedbiosystems.com/informatics/productpages/bl_body.html). However, they are difficult to integrate and use for supporting a multi-user and multi-tasking laboratory and they are not user friendly. NetBookingLIMS (InterNet Booking Laboratory Information Management System) is a WWW-based software that automates the management of data within a laboratory. It includes many novel features, is dynamic and flexible, and can fully automate either a laboratory with one instrument and one staff or thousands of instruments and thousands of staff. It uses logical structure to ensure that booking, updating and running the samples is done in a logical sequential order eliminating human errors. This is done by displaying a month-ahead table of all booking on all instruments in the laboratory, as shown in figure 1. NetBookingLIMS contains many intelligent error checking algorithms and user-friendly forms for data entry. It allows the researcher to cancel bookings when experiments fail. It is platform independent and only requires an Internet browser to run. The system automatically generates instrument compatible sample sheets for the staff. It contains built in intelligent algorithms to modify the state of the web page according to the real world state. For example, it remembers the important information of a client (user, staff or manager) and displays the appropriate "view" according to the client's status. It has an automated scheduling module allowing users to schedule the booking of large numbers of samples. This is implemented using a hybrid first in first out



The screenshot shows a web browser window displaying a dynamic table for the months of February and March 2001. The table is organized as a grid with columns representing days of the week and rows representing different instruments or sample types. The interface includes a title bar, navigation buttons, and a main data area with a grid of cells containing text and numbers, representing the booking schedule.

Figure 1. Screenshot of the main dynamic table in NetBookingLIMS.

(FIFO)/priority scheduling algorithm.³ It has an auxiliary automatic sample generator software that caters to 96 and 384 microtitre plate formats. Its architecture is such that it uses internal centralised ASCII flat files to store booking records information which is then automatically converted to a relational database using an SQL (Structured Query Language) module to produce auditing, usage/billing and troubleshooting management reports. Examples of queries are shown in table 1. A booking invoice record consists of the following fields:

Invoice ID, Application Description, Instrument ID, Date of Run, User First Name, User Surname, Project Name, Team Leader, Email, Phone Number, Sample Sheet ID, Computer Results Sent to where Invoice ID is the primary key, Sample Sheet ID is a foreign key that derives the lane number and the corresponding sample name booked on a run and Application Description defines the type of application to execute on the instrument.

NetBookingLIMS is designed using SSADM (Structured System Analysis and Design Methodologies) and implemented as a 3-tiered architecture. Analysis includes a feasibility study which is carried out to determine the structure and relationship between various entities within the laboratory. An entity-relationship model (ERM) is produced as shown in figure 2.

In order to determine the workflow path and identify problems with the management of the laboratory, a data flow diagram is constructed as shown in figure 3.

Using the data flow diagram and entity relationship model as a basis, user requirements and specifications are determined and a 3-tier architectural design is conceived for the software. This consists of a front-end design which includes the data entry forms and the pages displayed to the end user via an Internet browser, a middle layer which is the main engine that takes the data from the front end and processes them according to the specification, and a back-end layer which deals with how the data are stored. This includes the centralised ASCII flat file and its link to the relational database that is queried by the manager in order to obtain various types of management and troubleshooting reports. An overview of the architecture of the software is shown in figure 4. Formal specification of the software was carried out using the formal specification language Z.⁴ The main engine is implemented in Perl5 using CGI.pm. This generates the front end in HTML (Hyper Text Markup Language) that is displayed to the end user via an Internet browser and its central ASCII data store can be easily linked to a RDBMS (Relational Database Management System) such as Oracle (<http://www.oracle.com>), MySQL (<http://www.mysql.com>) or Sybase (<http://www.sybase.com>).

It has a user-friendly interface that automatically generates SQL queries and displays the output in table format together with the SQL query that generated it. The NetBookingLIMS software is designed and implemented to run via Internet browser, thus posing minimal requirements on the end user. Although it is Internet-based, it can be easily ported to run locally on an intranet web server and since it stores the records in a flat file, it can function without the need for RDBMS on the same web server where it is installed. It has been successfully used to automate and manage the workflow in many diagnostic and genetics core laboratories.

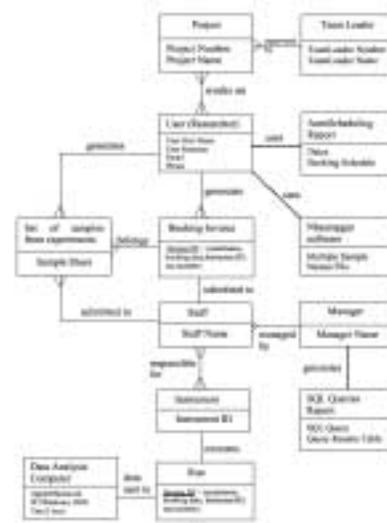
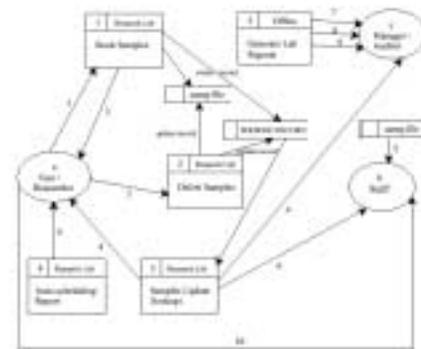


Figure 2. Entity-relationship model of a genetics core laboratory.

Figure 2. Entity-relationship model of a genetics core laboratory.



KEY

1. Book samples
2. Delete samples (cancel booking)
3. Booking Invoice
4. Update samples booked on specific run (lookup facility)
5. Automatic generation of instrument compatible sample sheet for run
6. Auto scheduling of samples report
7. Auditing report generation
8. Usage/billing report generation
9. Troubleshooting report generation
10. Experiment samples and invoice submission for a booked run

Figure 3. data flow diagram showing the workflow within a genetics core laboratory.

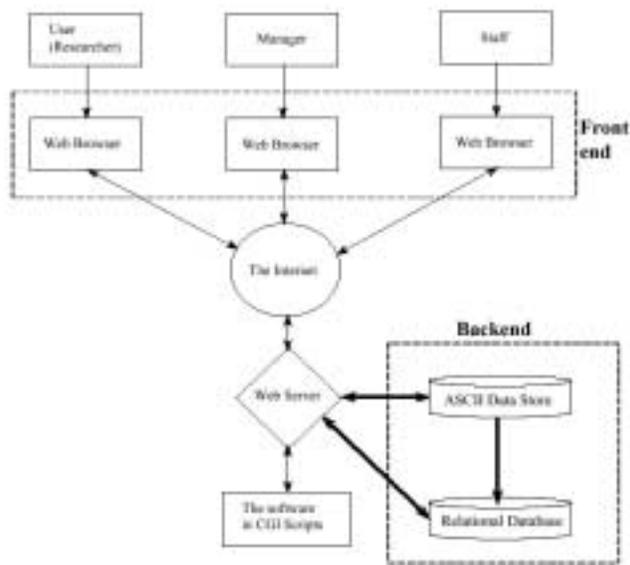


Figure 4. Overview of the architectural design of NetBookingLIMS software.

1.	Find the Application(s) booked on Machine(s) between specified dates?
2.	Count the number of Application(s) booked on Machine(s) between specified dates?
3.	Find the Team Leader bookings on Machine(s) between specified dates?
4.	Find all Project names managed (supervised) by Team Leader booked on Machine(s) between specified dates?
5.	Find all Project names and Application(s) booked by user with Firstname and Surname booked on Machine(s) between specified dates
6.	Find all attributes of booking record with specified NBLIMS ID (Invoice ID)?
7.	Find all booking records on Machine(s) between specified dates?

Table 1. Examples of NetBookingLIMS queries

ACKNOWLEDGEMENTS

This work was funded by the Institute of Cancer Research supported by the Cancer Research Campaign.

REFERENCES

1. Gregory S. G., Howell, G. R., Bentley, D. R., Genome Mapping by Fluorescent Fingerprinting, *Genome Research*, Vol. 7, No. 1, 1162, 1997.
2. Edwards S.M., Kote-Jarai, Z., Hamoudi, R., Eeles, R.A., An Improved High Throughput Heteroduplex Mutation Detection System for Screening BRCA2 Mutations - Fluorescent Mutation Detection (F-MD), *Human Mutation*, In Press.
3. Havender, J. W., Avoiding Deadlock in Multitasking Systems, *IBM Systems Journal*, Vol. 7, No. 2, 74, 1968.
4. Spivey, J. M., An Introduction to Z and Formal Specifications, *Software Engineering Journal*, Vol. 4, No. 1, 40, 1989.

Rifat Hamoudi has setup and is managing the Cancer Gene Cloning Centre, which is a core genetics laboratory whose main aim is to isolate, characterise and screen cancer causing genes. Mr. Hamoudi is also responsible for all the bioinformatics and robotics within the laboratory, he has obtained three Master of Science (MSc) degrees; MSc in Biochemistry from Kings College, University of London, MSc in Information Engineering from City University, London and MSc in Computing Science from Birkbeck College, University of London. He is a member of the Institute of Biology, British Computer Society and the Institute of Physical Sciences in Medicine. He has chartered biologist and chartered engineer status.

For more detailed information look at his web site at <http://www.users.totalise.co.uk/~rifat>

If you are interested in seeing more articles like this in JALA, please email jalanews@virginia.edu, and simply put the code (627) in the subject line.