SLIMS: a LIMS for handling next-generation sequencing workflows

Francesco Venco[1], **Arnaud Ceol**[2], **Heiko Muller**[2]✉

[1]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan, Italy
[2]Computational Research, Center for Genomic Science of IIT@SEMM, Istituto Italiano di Tecnologia (IIT), Milan, Italy

Received 16 September 2013; **Accepted** 18 September 2013; **Published** 14 October 2013

Competing interests: the authors have declared that no competing interests exist.

Abstract

Next-generation sequencing (NGS) is becoming a standard method in modern life-science laboratories for studying biomacromolecules and their interactions. Methods such as RNA-Seq and DNA resequencing are replacing array-based methods that dominated the last decade. A sequencing facility needs to keep track of requests, requester details, reagent barcodes, sample tracing and monitoring, quality controls, data delivery, creation of workflows for customised data analysis, privileges of access to the data, customised reports etc. An integrated software tool to handle these tasks helps to troubleshoot problems quickly, to maintain a high quality standard, and to reduce time and costs needed for data production. Commercial and non-commercial tools called LIMS (Laboratory Information Management Systems) are available for this purpose. However, they often come at prohibitive cost and/or lack the flexibility and scalability needed to adjust seamlessly to the frequently changing protocols employed. In order to manage the flow of sequencing data produced at the IIT Genomic Unit, we developed SLIMS (Sequencing LIMS).

Motivation and Objectives

Next-generation sequencing is becoming a standard method in modern life science laboratories for studying biomacromolecules and their interactions. Methods such as RNA-Seq and DNA resequencing are replacing array-based methods that dominated the last decade. A sequencing facility needs to keep track of requests, requester details, reagent barcodes, sample tracing and monitoring, quality controls, data delivery, creation of workflows for customised data analysis, privileges of access to the data, customised reports etc. An integrated software tool to handle these tasks helps to troubleshoot problems quickly, to maintain a high quality standard, and to reduce time and costs needed for data production. Commercial and non-commercial tools called LIMS (Laboratory Information Management Systems) are available for this purpose (Melo *et al.*, 2010; Stocker *et al.*, 2009; Triplet and Butler, 2012; Scholtalbers *et al.*, 2013). However, they often come at prohibitive cost and/or lack the flexibility and scalability needed to adjust seamlessly to the frequently changing protocols employed. In order to manage the flow of sequencing data produced at the IIT Genomic Unit, we developed SLIMS (Sequencing LIMS).

Methods

SLIMS is a web application with a MySQL backend that was developed in continuous interaction with the wet-lab scientists running the sequencing facility and with database experts from the University of Milan (Politecnico). SLIMS is written in the Java programming language, runs on a Glassfish web server, uses Hibernate for object-relational mapping, follows the Model-View-Controller model, and employs the Java-Server-Faces and the PrimeFaces framework for the frontend. Maven is employed to manage dependencies and to build the software. Message bundles are used for easy internationalisation. Workflows are provided to the system as XML files that list the protocol steps in detail. A script generator reads the workflow and generates all the commands needed to perform demultiplexing, quality control, alignments, and visualisation steps. SLIMS can run in a completely automated fashion, although human interaction is helpful at times, especially when non-standard tasks are to be performed. Access to the system is provided via an LDAP realm. LDAP authentication permits a single login policy for users of computational resources. Once authenticated, five roles are being assigned: admin, groupleader, user, analyst, and guest. Admins have full access and can submit or delete sequencing reagents, submit or delete sequencing requests, etc. Users can submit, view, and modify their requests. Supervisors can view the requests of their users. Analysts can create, modify, and launch workflows that ultimately lead to tracks viewable by the biologist in a web browser. Guests can view the general guidelines and the statistics regarding requests, waiting lists, and data delivery.

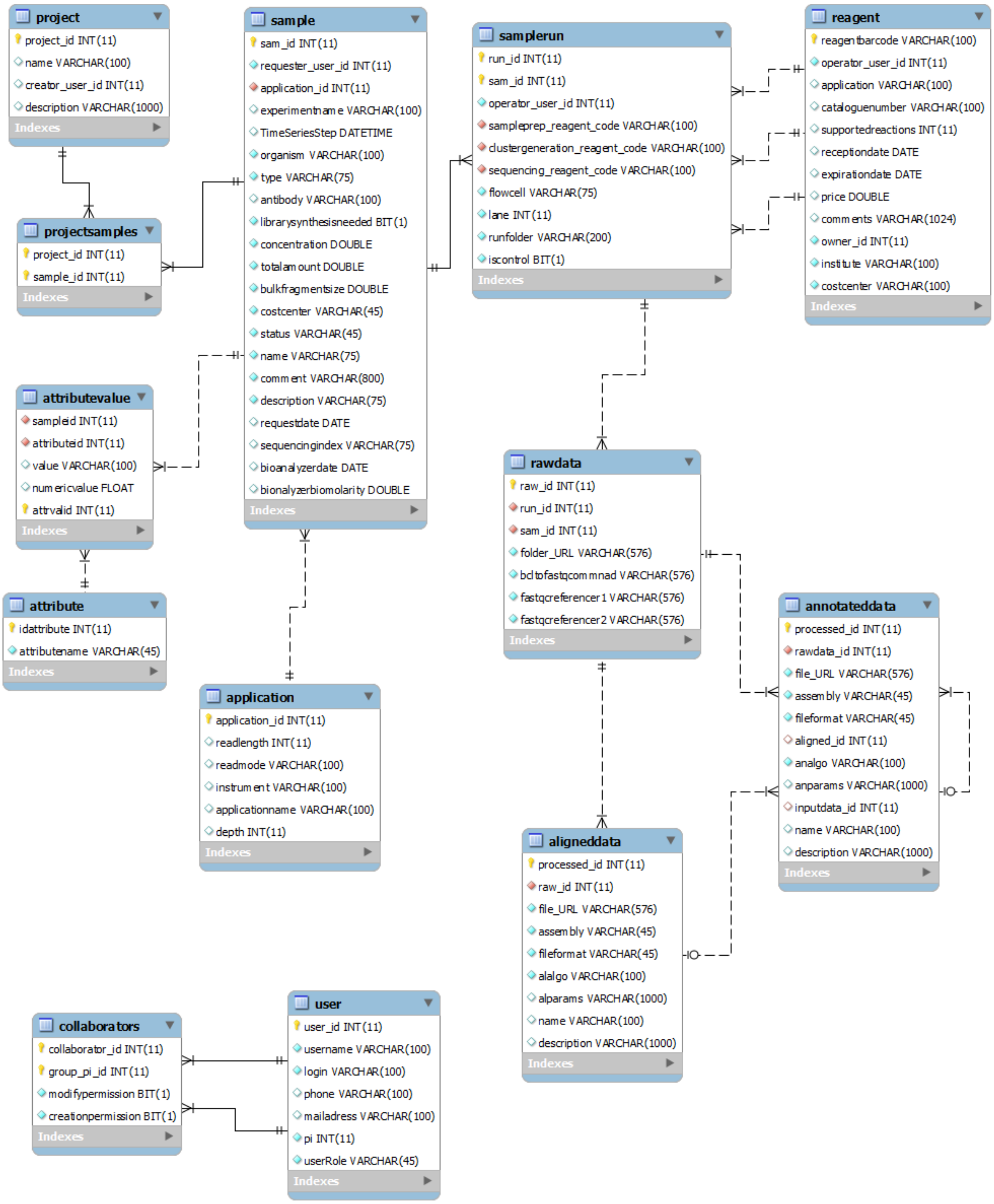

Figure 1. The data model of SLIMS.

Results and Discussion

SLIMS has been developed since August 2011. The first version has been deployed in September 2011. Since then, 2,700 samples have been processed. Samples processed before September 2011 were also inserted into the system. Since the use of SLIMS, data delivery procedures have been standardized making it easier for biologists and analysts to navigate the data. Furthermore, data delivery times have been drastically reduced. Currently, when the transfer of data from the sequencing machine to the data archive is complete, demultiplexing starts and provides data in Fastq format within hours, e.g. 4 hours for 50 base pair single read protocols. In summary, SLIMS is handling all the sequencing requests that are processed by the IIT Genomic Unit located at the IFOM-IEO-campus and has been adapted to necessities identified during the processing of numerous sequencing runs. SLIMS is available at http://cru.genomics.iit.it/SLIMS.

Acknowledgements

We are indebted to Prof. Stefano Ceri, Prof. Marco Masseroli, and Dr. Fernando Palluzzi for numerous discussions regarding the data model. This work was supported by a PRIN awarded to Prof. Stefano Ceri.

References

Melo A, Faria-Campos A, *et al.* (2010) SIGLa: an adaptable LIMS for multiple laboratories. *BMC Genomics* **11**(Suppl 5), S8. doi:10.1186/1471-2164-11-S5-S8

Scholtalbers J, Rößler J, *et al.* (2013) Galaxy LIMS for Next Generation Sequencing. *Bioinformatics* **29**(9), 1233-1234. doi:10.1093/bioinformatics/btt115

Stocker G, Fisher M, *et al.* (2009) iLAP: a workflow-driven software for experimental protocol development, data acquisition and analysis. *BMC Bioinformatics* **10**, 390. doi:10.1186/1471-2105-10-390

Triplet T, Butler G (2012) The EnzymeTracker: an open-source laboratory information management system for sample tracking. *BMC Bioinformatics* **13**, 15. doi:10.1186/1471-2105-13-15