MISO*: an open-source LIMS for small-to-large sequencing centres

Robert Davey, Xingdong Bian, Ani Thanki, Mario Caccamo

The Genome Analysis Centre, Norwich Research Park, UK

* "Managing Information for Sequencing Operations"

Abstract

Sequencing centres differ not only in their scale and output, but also their requirements for data management. The efficient storage of genomic metadata is vital for all sequencing centres. Off-the-shelf solutions are often very expensive and not cost-effective, especially for the smaller centre. Furthermore, paid customised support is often required, and the extensibility of these systems is rarely in the hands of the community. Alongside a desire to tailor an open information system in-house, data formats can change, new platforms are being developed and those platforms can evolve rapidly. These are valid concerns for both large centres characterised by high-throughput data production and smaller scale laboratories with constrained expenditure for IT solutions, and potentially project-specific requirements.

Hence, we present MISO ("Managing Information for Sequencing Operations"), a freely available open-source LIMS for recording next-generation sequencing (NGS) metadata. MISO can store relevant metadata for the most common NGS platforms (e.g. Illumina GA, HiSeq and MiSeq, Roche 454, ABI SOLiD and PacBio RS) and automatically generate public repository data submission schemas (e.g. EBI Sequence Read Archive). It comprises many essential features, such as secure authentication, fine-grained access control, barcode tracking, and reporting. We are also working on modules that expose the run information from these NGS platforms to allow automated tracking and generation of statistics to be an integral part of MISO, and an automated submission pipeline facilitating easy public data exposure.

Features

MISO provides the following elements:

- **Authentication** – user-centric access control to designated areas
- **Laboratory tracking** – project description, sample receipt, library preparation, run construction, and barcoding
- **Printing** – can directly connect to barcode printers and print barcodes for the relevant objects inside MISO
- **Bioinformatics pipelines** – monitoring and reporting of analytical processes, interacting directly with the computing cluster via a dedicated server module
- **Reporting** – accurate statistics from project status, sample processing, library preparation and sequencing throughput (Figure 1)
- **Data Visualisation** – through “traffic light” indicators, tree-structured status diagrams, and the sequencer resource calendar, MISO can easily show the status of each project and its elements (Figure 1)
- **Notification** – automated import of run metadata and notification of change of run status via a dedicated server module
- **Submission** – automated packaging of sequencing data and metadata, deployment of these data to remote public repositories. MISO is heavily modelled upon the submission schemas specified by the SRA (http://www.ebi.ac.uk/ena/) and is therefore able to automatically generate and deploy the required XML and data files (Figure 2)

We are working on new plugin modules for MISO, making it easier for the community to contribute features, improvements and integration tools.

Implementation

We have chosen a relational database model to underpin MISO. Currently this is MySQL, but can be replaced with any other RDBMS (data access objects will obviously have to be constructed for each persistence store). Object relational mapping (ORM) via data access objects (DAOs) is achieved through Spring’s JDBC Templates, which enable fine-grained control over queries over the underlying database. We are currently working on DB4O and document-based No-SQL DAO layers to provide object persistence.

We use simple Java beans, or Plain Old Java Objects (POJOs), to represent the objects in the database, and we construct concrete relationships between them in business logic space (Figure 2).

We have developed a RESTful web application from the Spring 3 MVC framework and, as such, MISO can support a number of view technologies. It comes with a user-friendly interface built on Java Server Pages (JSP) which is designed specifically for lab technicians, but can be queried programmatically using the REST URLs. This means helpful web services, such as remote reporting and custom integration, are available.

We use the D3 Javascript library (http://mbostock.github.com/d3/) to produce interactive visualisations of the tree project trees, reporting charts and run statistics (Figure 1).

Conclusions

Modular architecture and frameworks such as Spring enable us to supply a configurable extensible product, where functionality changes are possible without any code changes. For example, replacing a MySQL database model with DB4O or replacing the authentication mechanism is possible with minimal configuration.

Such elements, coupled with the open-source ethos of community development, can make for a cost-effective, highly modifiable and robust system that can benefit both the public knowledge repositories and sequencing centres of all sizes.

Availability and acknowledgements

MISO has been released to the community and is available in a variety of ways, including full source code and an off-the-shelf virtual image for getting MISO up and running easily. We would be pleased to help any developers or potential users interested in the project:

http://www.tgac.ac.uk/miso/

A demo instance is also currently available:

http://miso-demo.tgac.ac.uk/

---

**MISO**: an open-source LIMS for small-to-large sequencing centres

The Genome Analysis Centre, Norwich Research Park, UK

* "Managing Information for Sequencing Operations"

Abstract

Sequencing centres differ not only in their scale and output, but also their requirements for data management. The efficient storage of genomic metadata is vital for all sequencing centres. Off-the-shelf solutions are often very expensive and not cost-effective, especially for the smaller centre. Furthermore, paid customised support is often required, and the extensibility of these systems is rarely in the hands of the community. Alongside a desire to tailor an open information system in-house, data formats can change, new platforms are being developed and those platforms can evolve rapidly. These are valid concerns for both large centres characterised by high-throughput data production and smaller scale laboratories with constrained expenditure for IT solutions, and potentially project-specific requirements.

Hence, we present MISO ("Managing Information for Sequencing Operations"), a freely available open-source LIMS for recording next-generation sequencing (NGS) metadata. MISO can store relevant metadata for the most common NGS platforms (e.g. Illumina GA, HiSeq and MiSeq, Roche 454, ABI SOLiD and PacBio RS) and automatically generate public repository data submission schemas (e.g. EBI Sequence Read Archive). It comprises many essential features, such as secure authentication, fine-grained access control, barcode tracking, and reporting. We are also working on modules that expose the run information from these NGS platforms to allow automated tracking and generation of statistics to be an integral part of MISO, and an automated submission pipeline facilitating easy public data exposure.

Features

MISO provides the following elements:

- **Authentication** – user-centric access control to designated areas
- **Laboratory tracking** – project description, sample receipt, library preparation, run construction, and barcoding
- **Printing** – can directly connect to barcode printers and print barcodes for the relevant objects inside MISO
- **Bioinformatics pipelines** – monitoring and reporting of analytical processes, interacting directly with the computing cluster via a dedicated server module
- **Reporting** – accurate statistics from project status, sample processing, library preparation and sequencing throughput (Figure 1)
- **Data Visualisation** – through “traffic light” indicators, tree-structured status diagrams, and the sequencer resource calendar, MISO can easily show the status of each project and its elements (Figure 1)
- **Notification** – automated import of run metadata and notification of change of run status via a dedicated server module
- **Submission** – automated packaging of sequencing data and metadata, deployment of these data to remote public repositories. MISO is heavily modelled upon the submission schemas specified by the SRA (http://www.ebi.ac.uk/ena/) and is therefore able to automatically generate and deploy the required XML and data files (Figure 2)

We are working on new plugin modules for MISO, making it easier for the community to contribute features, improvements and integration tools.

Implementation

We have chosen a relational database model to underpin MISO. Currently this is MySQL, but can be replaced with any other RDBMS (data access objects will obviously have to be constructed for each persistence store). Object relational mapping (ORM) via data access objects (DAOs) is achieved through Spring’s JDBC Templates, which enable fine-grained control over queries over the underlying database. We are currently working on DB4O and document-based No-SQL DAO layers to provide object persistence.

We use simple Java beans, or Plain Old Java Objects (POJOs), to represent the objects in the database, and we construct concrete relationships between them in business logic space (Figure 2).

We have developed a RESTful web application from the Spring 3 MVC framework and, as such, MISO can support a number of view technologies. It comes with a user-friendly interface built on Java Server Pages (JSP) which is designed specifically for lab technicians, but can be queried programmatically using the REST URLs. This means helpful web services, such as remote reporting and custom integration, are available.

We use the D3 Javascript library (http://mbostock.github.com/d3/) to produce interactive visualisations of the tree project trees, reporting charts and run statistics (Figure 1).

Conclusions

Modular architecture and frameworks such as Spring enable us to supply a configurable extensible product, where functionality changes are possible without any code changes. For example, replacing a MySQL database model with DB4O or replacing the authentication mechanism is possible with minimal configuration.

Such elements, coupled with the open-source ethos of community development, can make for a cost-effective, highly modifiable and robust system that can benefit both the public knowledge repositories and sequencing centres of all sizes.

Availability and acknowledgements

MISO has been released to the community and is available in a variety of ways, including full source code and an off-the-shelf virtual image for getting MISO up and running easily. We would be pleased to help any developers or potential users interested in the project:

http://www.tgac.ac.uk/miso/

A demo instance is also currently available:

http://miso-demo.tgac.ac.uk/