
Software Requirements Specification

for

Molecular Integration Database

Version 0.1

Prepared by Allan Kamau

SANBI

September 15, 2004

Table of Contents

1. Introduction.....	1
1.1 Purpose	1
1.2 Document Conventions.....	1
1.3 Intended Audience and Reading Suggestions.....	1
1.4 Project Scope.....	1
2. Overall Description.....	2
2.1 Product Perspective.....	2
2.2 Product Features.....	2
2.3 User Classes and Characteristics.....	2
2.4 Operating Environment.....	2
2.5 Design and Implementation Constraints.....	2
2.6 User Documentation.....	2
2.7 Assumptions and Dependencies.....	3
3. System Features.....	3
3.1 Data Integrity.....	3
3.2 Data Security.....	4

4.External Interface Requirements.....	4
4.1User Interfaces.....	4
4.2Hardware Interfaces.....	4
4.3Software Interfaces.....	4
4.4Communications Interfaces.....	4
5.Other Nonfunctional Requirements.....	5
5.1Performance Requirements.....	5
5.2Safety Requirements.....	5
5.3Security Requirements.....	5
5.4Software Quality Attributes.....	5
6.Other Requirements	

Revision History

Name	Date	Reason For Changes	Version

Introduction

Purpose

The database sequencing pipeline application makes use of a database to store, manage and retrieve data associated with the overall sequencing process. Also to be included in the database are users credentials data, process management data, logging and auditing data.

Document Conventions

In this document, ERD is used to represent Entity Relationship analysis Diagram. The contents contained herewith are subject to change as the pipeline application evolves during development as more requirements are identified. RDBMS stands for Relational DataBase Management System. Molecular Integration Database is represented by MID.

Intended Audience and Reading Suggestions

This document is intended for use by the developer(s) developing the sequencing pipeline or those involved in the sequencing pipeline maintenance. Application assemblers and deployers and system and database administrators will find this document useful in order to successfully deploy and manage the sequencing pipeline application.

Project Scope

The proposed database will be relational in nature and will be implemented in an RDBMS server

software. The objective of the database design and implementation is to develop a Molecular Integration Database that will provide the sequencing pipeline software with the necessary data requirements for its functioning and the functioning of the processes and procedures the sequencing pipeline addresses.

The data to be stored in the MID can be used by other applications some of which have not been identified at a later stage.

Overall Description

Product Perspective

The Molecular Integration database is a crucial component of the sequencing pipeline software. The proposed MID is a replacement of the current data store implementation that makes use of the operating systems file system. A database driver is required by the sequencing pipeline application in order to communicate with the MID.

Product Features

The MID will benefit from the RDBMS server software's features that it will be implemented on. This will include data security, data integrity, scalability and speed, data backup and restore facilities offered by the RDBMS.

User Objectives

The users: the sequencing application developer(s), the scientists submitting the samples for sequencing would like to have data posted and retrieved from the MID should be consistent.

Operating Environment

The MID will be implemented on a PostgreSQL 7.x database. The proposed operating system and hardware will be Linux or any flavour of UNIX on any processor architecture that can run the operating system.

Design and Implementation Constraints

The MID should be implemented on Open Source software. And the data should be stored on an industry standard implementation and not on proprietary data. This ensures that data is easily understandable accessible with the necessary security authentication and authorization to the

appropriate applications like sequencing pipeline application and other applications through the use of industry standard data access drivers.

User Documentation

A process (DFD) diagram, ERD document and a table layout showing the data types has been included.

Assumptions and Dependencies

The MID will not directly access remote or external data sources. Data from such sources will be dumped in to the appropriate table(s) in the MID by a third party application.

System Features

The MID is required to manage the data stored in it and ensure data correctness, security and provide fast and efficient data retrieval and storage.

Data integrity

3.1.1 Description and Priority

The MID will take advantage of the RDBMs server software data referential integrity which insures that related data will not be partially stored, deleted or modified. This is important because incorrect data will lead to incorrect findings which obviously is not recommended in the scientific community.

3.1.2 Stimulus/Response Sequences

Any attempt to store or delete data in the MID database will need to be verified to ensure that if the data being stored should belong to (be associated with) preexisting data in the database then the preexisting data should really exist in the database. The same holds true of any attempt to delete data that the data been deleted may not be required by other existing data. An example of this is an attempt to delete a scientists who already has chromatograms associated with them.

3.1.3 Functional Requirements

The MID table structures will define the necessary integrity rules. An attempt to violate any of this rules during insert, update or deletion will result in an error been thrown.

Data Security

3.2.1 Description and Priority

This functionality will be provided by the RDBMS server software.

Access to the MID will be controlled. The database administrator will create users and groups for the applications and individuals who will be required to have access to the data.

For these users, login credentials will be defined along with permissions detailing what they can and may not.

It is important to ensure that the data is not lost or deleted by unidentified users and hackers.

3.2.2 Stimulus/Response Sequences

All access to the MID will be restricted to logged in users. Any action on the data contained in the MID will be checked against the user's privileges.

3.2.3 Functional Requirements

The MID database will need to be configured to only allow certain users to access it.

External Interface Requirements

User Interfaces

Client access to the MID is either by the RDBMS client tools or by third party applications like the sequencing pipeline. The RDBMS internal client tools connect the database in a native and proprietary way, the third party applications will make use of database connectivity drivers to connect to and access the MID.

Hardware Interfaces

The MID will not require a hardware interface as it will be implemented inside a pre-installed RDBMS.

Software Interfaces

Client applications that will require access to the MID will require a language and RDBMS specific database driver or libraries. For example if the client application is written in Java, then a JDBC driver for PostgreSQL 7.x is required.

Communications Interfaces

Communication to the MID will be by TCP/IP. Initially the database will reside behind a firewall.

Other Nonfunctional Requirements

Performance Requirements

The MID should be as scalable enough to satisfy the growing needs of the users and user requirement some of which are yet to be identified, and should gracefully degrade in performance under high load.

Performance priority should be given to the data associated with user interaction, such as the uploading of the chromatograms.

Safety Requirements

Yet to be identified.

Security Requirements

The MID data should be protected from hackers and eavesdroppers, this can be achieved by restricting access to the MID by electronic means by use of a firewall and by use of user authentication and authorization.

Data backup is also essential in order to minimize the data loss in the event of hardware or software crash, mischievous user activities, or loss of equipment by means of fire, burglary or by other means.

Other Requirements

Yet to be identified.

Appendix A: Glossary

RDBMS stands for Relational DataBase Management System

MID stands for Molecular Integration Database

Appendix B: Analysis Models

ERD, DFD and table layout document are included.

Appendix C: Issues List

None.