Development of a Framework for Laboratory Information Tracking in Support of Research and Clinical Operations

CIDR Center for Inherited Disease Research Johns Hopkins University

B. Craig, B. Tibbils, K. Topper, S. Griffith, L. Watkins, K. Doheny

Center for Inherited Disease Research (CIDR), Johns Hopkins Genomics (JHG), IGM, JHU-SOM, Baltimore, MD

JOHNS HOPKINS UNIVERSITY

Introduction

The Center for Inherited Disease Research (CIDR) part of Johns Hopkins Genomic (JHG) provides high quality next generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. CIDR's current genotyping Laboratory Information Management System (LIMS) was designed to handle the tracking of detailed information for large numbers of samples going through specific genotyping assays. This LIMS has over time become difficult to maintain and extend as newer resources and NGS services have emerged. A third party LIMS was selected for NGS applications that addressed the widely different set of tasks and information to be stored for sequencing services. Annual licensing costs and requirements to re-write custom code and upgrade to newer releases to maintain support add to the difficulty of sustaining two LIMS deployments. These deficiencies combined with the expansion of research and clinical genomic services at JHG (jhgenomics.jhmi.edu) have highlighted the need to create a new LIMS to support the breadth of current and future protocols. These rapidly changing laboratory protocols require reliable tracking of large amounts of workflow information including a record of sample movements, identification of robot resources, reagent association and user comment tracking. Here we describe the development of a MySQL database backed Java framework consisting of tools to reduce disruption to lab processing with module hot swapping, limit database schema complexity and increase reusable graphical user interface (GUI) components to flexibly build and extend workflows for any protocol driven technology.

Motivation for LIMS Infrastructure replacement

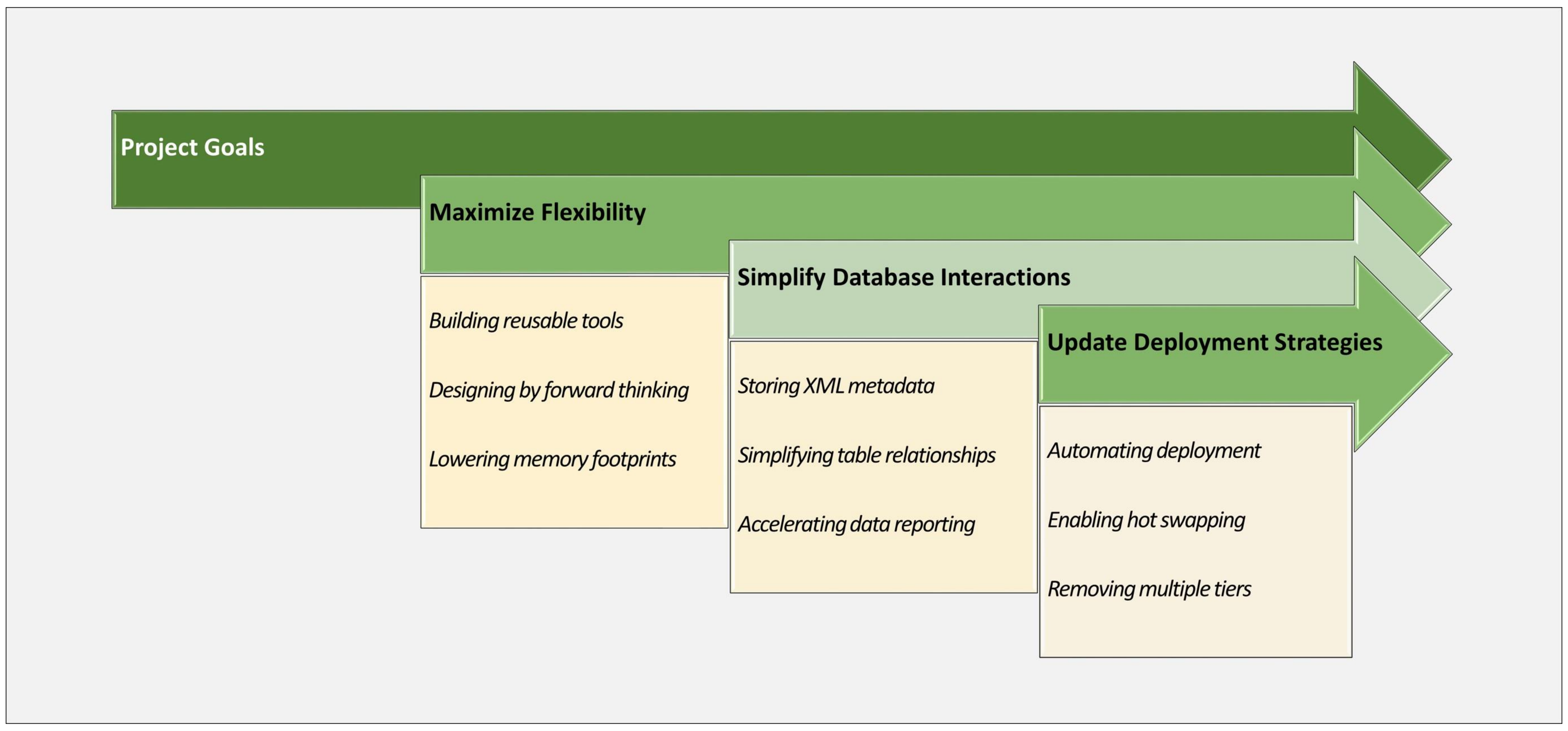
- Genotyping Service LIMS
  - Developed with 1 Illumina Infinium workflow in mind
  - Complex code management
  - Uses outdated software and hardware resources
- Sequencing Service LIMS
  - Pay for annual user license
  - Version updates require extensive time commitments
- Difficult to maintain two LIMS deployment strategies

Current LIMS deployment strategies

Topic	Current In-house Genotyping LIMS	Current 3rd Party Sequencing LIMS
Age (years)	12+	7
Database design	Highly normalized MySQL schema	Proprietary MySQL schema
Application Logic	HTTP requests -> Java code validation -> JDBC -> HTTP responses	Proprietary business logic software
Client Interface	JavaScript web interface	Proprietary JavaScript web interface
Downtime	Any modification requires code redeployment	Business logic software updates Modifications to UI fields

Project Goals

General LIMS Framework Goals



Project Goals

Maximize Flexibility

Building reusable tools

Designing by forward thinking

Lowering memory footprints

Simplify Database Interactions

Storing XML metadata

Simplifying table relationships

Accelerating data reporting

Update Deployment Strategies

Automating deployment

Enabling hot swapping

Removing multiple tiers

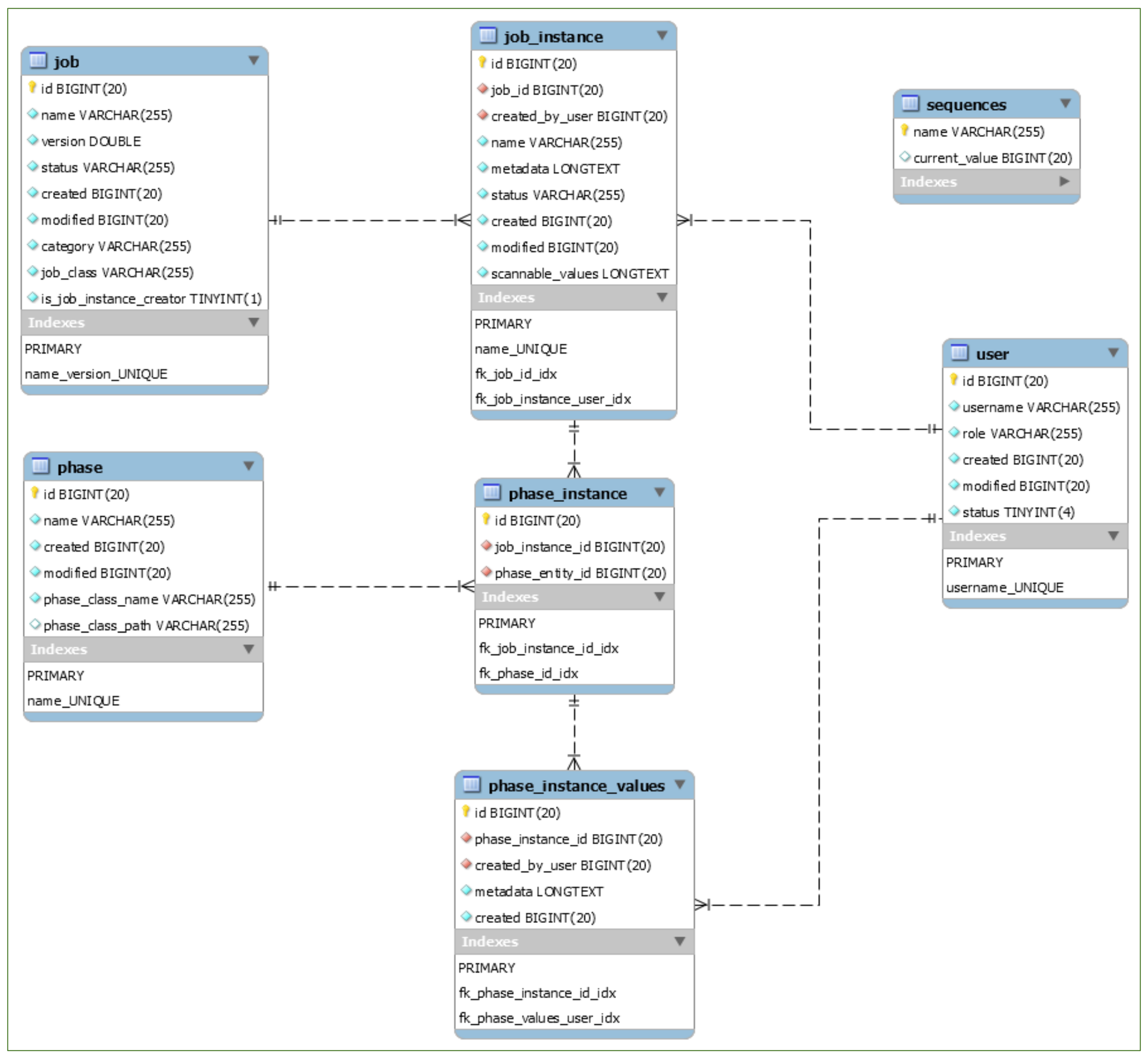
Specific LIMS Framework Goals

Increasing	**Decreasing**
Turnaround time for new LIMS jobs	Complexity of code maintenance
Robustness of audit tracking	Deployment errors and bugs
Scope of the LIMS to include clinical and R&D operations	User frustrations

New LIMS Framework Decisions

- Database Design
  - MySQL 5.7 (Figure 1)
    - Use of LONGTEXT datatype
      - Store submission data in XML format
    - Simplified table relationships
- Application Logic and Client Interface combined
  - Java 8 with Java FX GUI (Figure 2)
    - Customized and reusable components
    - Consistent phase navigation look and feel
      - "You are here" visualization
    - Real time and submit only validation notifications
  - Java Database Connectivity (JDBC)
  - Serialization of Java objects into XML
  - Automating code deployment
    - Gradle scripting
      - Compiles and validates source code
      - Creates Java Archive (JAR) files
      - Creates deployment directories
- Downtime
  - "Hot swapping" JAR files
    - Reflection and database validation
    - Eliminates downtime

Figure 1: MySQL Database schema



job
id BIGINT(20)
name VARCHAR(255)
version DOUBLE
status VARCHAR(255)
created BIGINT(20)
modified BIGINT(20)
category VARCHAR(255)
job_class VARCHAR(255)
is_job_instance_creator TINYINT(1)
Indexes
PRIMARY
name_version_UNIQUE

job_instance
id BIGINT(20)
job_id BIGINT(20)
created_by_user BIGINT(20)
name VARCHAR(255)
metadata LONGTEXT
status VARCHAR(255)
created BIGINT(20)
modified BIGINT(20)
scannable_values LONGTEXT
Indexes
PRIMARY
name_UNIQUE
fk_job_id_idx
fk_job_instance_user_idx

sequences
name VARCHAR(255)
current_value BIGINT(20)
Indexes

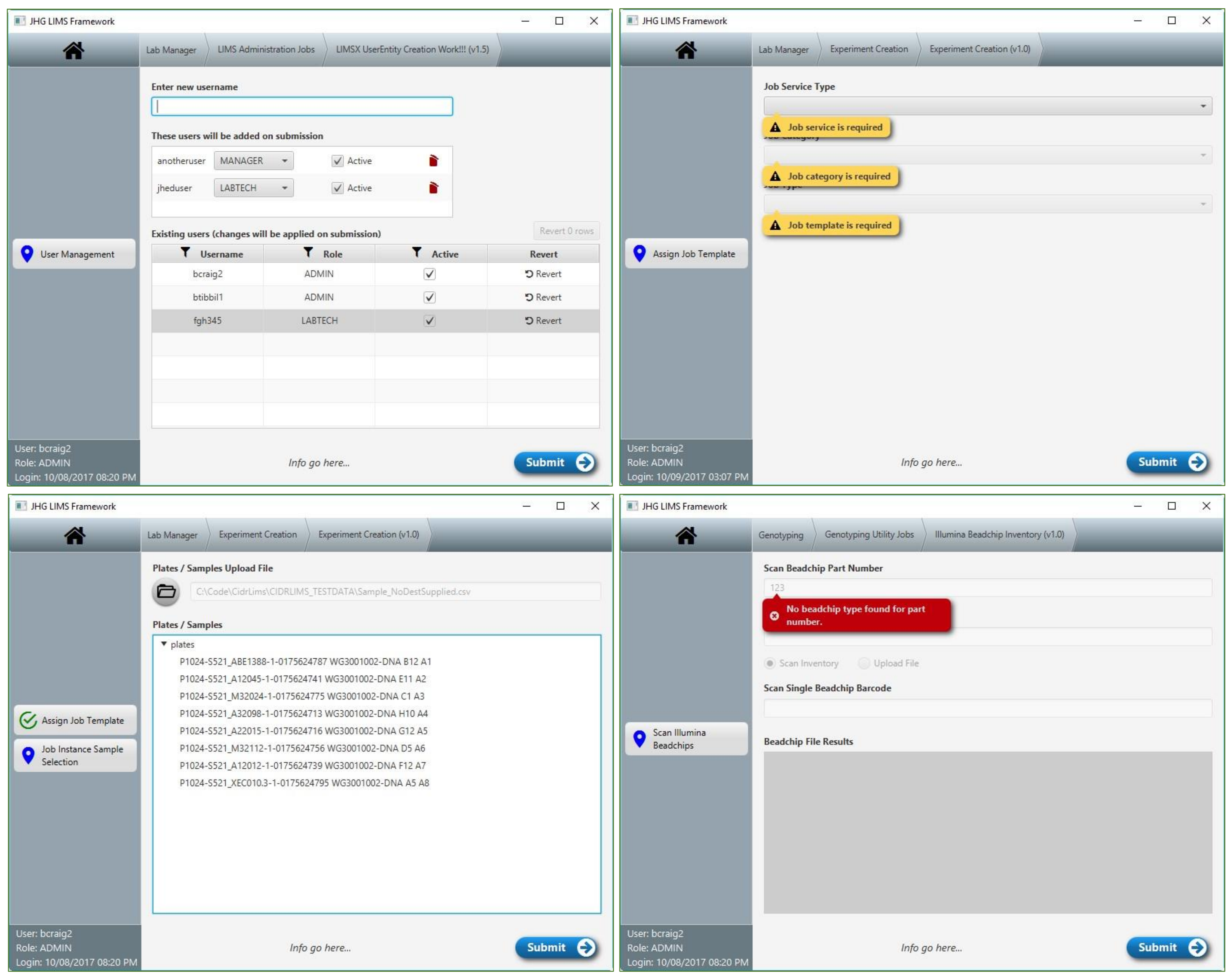
user
id BIGINT(20)
username VARCHAR(255)
role VARCHAR(255)
created BIGINT(20)
modified BIGINT(20)
status TINYINT(4)
Indexes
PRIMARY
username_UNIQUE

phase
id BIGINT(20)
name VARCHAR(255)
created BIGINT(20)
modified BIGINT(20)
phase_class_name VARCHAR(255)
phase_class_path VARCHAR(255)
Indexes
PRIMARY
name_UNIQUE

phase_instance
id BIGINT(20)
job_instance_id BIGINT(20)
phase_entity_id BIGINT(20)
Indexes
PRIMARY
fk_job_instance_id_idx
fk_phase_id_idx

phase_instance_values
id BIGINT(20)
phase_instance_id BIGINT(20)
created_by_user BIGINT(20)
metadata LONGTEXT
created BIGINT(20)
Indexes
PRIMARY
fk_phase_instance_id_idx
fk_phase_values_user_idx

Figure 2: Representations of GUI components



JHG LIMS Framework
Lab Manager > LIMS Administration Jobs > LIMSX UserEntity Creation Work!!! (v1.5)
User Management
Enter new username
These users will be added on submission
anotheruser MANAGER ☑ Active
jheduser LABTECH ☑ Active
Existing users (changes will be applied on submission)
Revert 0 rows

Username	Role	Active	Revert
bcraig2	ADMIN	☑	Revert
btibbil1	ADMIN	☑	Revert
fgh345	LABTECH	☑	Revert

User: bcraig2
Role: ADMIN
Login: 10/08/2017 08:20 PM
Info go here...
Submit

JHG LIMS Framework
Lab Manager > Experiment Creation > Experiment Creation (v1.0)
Assign Job Template
Job Service Type
Job service is required
Job category is required
Job template is required
User: bcraig2
Role: ADMIN
Login: 10/09/2017 03:07 PM
Info go here...
Submit

JHG LIMS Framework
Lab Manager > Experiment Creation > Experiment Creation (v1.0)
Assign Job Template
Job Instance Sample Selection
Plates / Samples Upload File
C:\Code\CidrLims\CIDRLIMS_TESTDATA\Sample_NoDestSupplied.csv
Plates / Samples
▾ plates
P1024-S521_ABE1388-1-0175624787 WG3001002-DNA B12 A1
P1024-S521_A12045-1-0175624741 WG3001002-DNA E11 A2
P1024-S521_M32024-1-0175624775 WG3001002-DNA C1 A3
P1024-S521_A32098-1-0175624713 WG3001002-DNA H10 A4
P1024-S521_A22015-1-0175624716 WG3001002-DNA G12 A5
P1024-S521_M32112-1-0175624756 WG3001002-DNA D5 A6
P1024-S521_A12012-1-0175624739 WG3001002-DNA F12 A7
P1024-S521_XEC010.3-1-0175624795 WG3001002-DNA A5 A8
User: bcraig2
Role: ADMIN
Login: 10/08/2017 08:20 PM
Info go here...
Submit

JHG LIMS Framework
Genotyping > Genotyping Utility Jobs > Illumina Beadchip Inventory (v1.0)
Scan Illumina Beadchips
Scan Beadchip Part Number
123
No beadchip type found for part number.
● Scan Inventory ○ Upload File
Scan Single Beadchip Barcode
Beadchip File Results
User: bcraig2
Role: ADMIN
Login: 10/08/2017 08:20 PM
Info go here...
Submit

Discussion

The missions of CIDR and JHG are to map a genetic path to better health by integrating expertise, enabling data driven discovery and informing patient care. To accomplish this mission, services offered are constantly evolving requiring accelerated development of information tracking mechanisms. Originally designed for very specific workflows, the current LIMS infrastructure can no longer evolve at the pace needed to accomplish this goal. Building a new LIMS from the ground up gives the opportunity to evaluate new database types, deficiencies in the current LIMS and software development practices. Leveraging extensive database experience aided in the decision to use MySQL to make use of XML data types to store submission data in a graph like interconnected way (Figure 1). Conversion of user supplied data from Java objects to XML for database insertion, GUI repopulation and reporting is supported using the XStream Java library (x-stream.github.io). Each submission is treated as a new timestamped record in the database to extend audit tracking mechanisms. User feedback strongly shows that wet bench protocols can be optimized if GUI and validation code were more flexible. To address this deficiency dynamic phase GUI elements with methods for global and lab customized validations have been created. Validation failures are presented to users by contextual pop-ups (Figure 2). Review of current LIMS implementations reveal now obsolete Java library versions and code duplication to support lab and vendor protocol changes over time. These shortcomings are addressed by using a current JDK version, creating and updating internal cross project Java utilities and when possible using Maven (search.maven.org) to manage external library dependencies. Development work now focuses on creation of jobs and phases to mirror current genotyping production protocols. Through the use of updated technologies, forward thinking and inclusion of highly technical individuals this project will be integral to driving research and clinical operations into the next decade.

www.cidr.jhmi.edu