



MHI Internship Report:
Designing and Building an Internal Database at
AGADA Biosciences

By:
Xiaoqing Zhuang
B00716255
XiaoqingZhuang@Dal.Ca

Performed at:
AGADA Biosciences Inc.
1344 Summer Street, Dalhousie university,
Halifax, NS B3H 0A8

In Partial Fulfillment of the Requirements of the Master of Health Informatics Program,
Dalhousie University, Halifax, NS

Report of Internship for the Period January 3rd – April 28th, 2017
Date Submitted: April 24th, 2017

Acknowledgment and Endorsement

This report has been written by me and has not received any previous academic credit at this or any other institution.

I would like to thank Drs. Kanneboyina Nagaraju and Eric Hoffman for providing me with advice related to the Internal Database design. I would like to thank my supervisor, Amanda Mullen, Director of Operations, for providing me with feedback on my work and support. I would also thank all team members at AGADA Biosciences for the guidance, support and encouragement throughout the internship period.

Finally, I would like to thank Dr. Raza Abidi and other faculty members in the Masters of Health Informatics program for helping me to prepare for this internship.

(Electronically signed)

Xiaoqing Zhuang

Executive Summary

This report keeps track of the author's internship at AGADA Biosciences in Halifax, Nova Scotia. AGADA Biosciences is a company who specializes in performing pre-clinical studies using murine models. They focus on facilitating orphan drug development for industry, non-profits, and academic-based groups.

During the internship period, the author was assigned a role as software developer to design and build an internal database for AGADA Biosciences with the purpose of efficiently retrieving data from the database. Tasks performed by the author were managing data, designing database models, recommending information systems, building the database, installing and implementing systems, and reporting on progress.

This internship is highly related to the Health Informatics program. Knowledge learned from Health Informatics can be applied to all of the required tasks. Most of the problems, such as vague requirements, data quality issues, and time conflicts during the project can be solved with Health Informatics solutions such as drawing ERD and BPMN diagrams.

Objectives were met by the end of the internship. The author also learned how to put Health Informatics knowledge into practice and present results to employers. In addition, the author experienced teamwork, working with database management systems, and data processing.

Several recommendations are made for the future works.

Table of Contents

Acknowledgment and Endorsement	1
Executive Summary	2
Table of Contents	3
1. Introduction.....	4
1.1 Project Objectives.....	4
1.2 Relevance	4
1.3 Outcome	5
2. Description of the Organization.....	5
2.1 AGADA Biosciences Background	5
2.2 AGADA’s Business Services	6
2.3 AGADA Biosciences and Dalhousie University.....	6
3. Description of the Performed Work.....	7
3.1 Internship Description	7
3.2 Internship Role and Responsibilities	8
3.3 Internship achievements	10
4. Relation to Health Informatics.....	11
4.1 Relation to Health Informatics Courses	11
4.2 Relation to the Concepts of Health Informatics	13
5. Discussion on a Problem and Proposed Solution	14
5.1 Problem Description.....	14
5.2 Proposed Solution.....	16
6. Conclusion	18
7. Recommendations.....	18
Reference	19
Appendix.....	19

1. Introduction

1.1 Project Objectives

The core objectives of the project assigned to author throughout the internship period are to design and build an AGADA Biosciences Database to store data that is generated from pre-clinical studies in order to facilitate easy retrieval of the preclinical studies' outcomes.

Other objectives are

- Managing existing pre-clinical studies' data.
- Building relational database models for developing an internal database.
- Recommending a relational database management system based on employer's requirements.
- Developing working relationship with a dynamic, fun, tight-knit and committed team of researchers
- Working with AGADA Biosciences' management team to develop existing AGADA website by adding features and improving contents.

1.2 Relevance

This project involved managing pre-clinical study data, researching and using information management systems, communication, coding and health statistics which are related to both medical and health informatics.

- Building a database for AGADA Biosciences requires understanding their pre-clinical data. Knowledge from course Clinical Care Fundamentals can be applied.
- The course, Health Informatics Systems and Issues provided me the knowledge required to research and manage information systems.

- Creating the database model and the information flow diagram requires knowledge from the course Health Informatics Flow and Use.
- Knowledge from course, project management helps the author to manage the whole project in an efficient and effective way.
- Organizing and getting useful information from AGADA Biosciences pre-clinical studies raw data requires using knowledge from health informatics statistics and data mining for health informatics courses.

After data processing, AGADA Biosciences can get more accurate and useful data from database that can improve their future pre-clinical study outcomes and help them facilitate orphan drug development.

1.3 Outcomes

When this project is completed, an AGADA database will be created for internal use, with a database management system which was recommended and installed. AGADA Biosciences can access its own database in an easy and secure way. It allows them to leverage useful information from the database instead of digging into several spreadsheets. The entire AGADA Biosciences team was trained to use this database to retrieve data and generate customized reports by using a simple database user interface.

2. Description of the Organization

2.1 AGADA Biosciences Background

AGADA Biosciences is a privately-owned company founded by Drs. Nagaraju and Hoffman. The company was established in 2013 in Halifax, Nova Scotia. Its offices and laboratories are located at the newly complete Innovacorp building in Halifax, Nova Scotia. Its labs at the

Summer Street location are shared with Dalhousie University. AGADA's new murine facilities are maintained by Dalhousie University expert staff. AGADA focuses on facilitating orphan drug development for industry, non-profits, and academic-based groups. AGADA's founders Dr. Nagaraju and Hoffman are key opinion leaders for preclinical efficacy studies. They helped lead international efforts to develop standard operating procedures (SOPs) for pre-clinical outcome measures in muscular dystrophy mouse models with the trans-EU TREAT-NMD group ("About Us | AGADA Biosciences", 2017).

AGADA's mission is "Meeting a rapidly growing need for services that accelerate drug development for orphan diseases, with specific expertise in murine models of Duchenne muscular dystrophy and other muscle disease" ("About Us | AGADA Biosciences", 2017).

2.2 AGADA's Business Services

AGADA provides following services ("Services| AGADA Biosciences", 2017):

- *In vivo* functional testing. For example; treadmill exercise, treadmill exhaustion assay, and Rota rod Test.
- Muscle physiology. For example; fatigue resistance measurements.
- Biochemical and molecular testing; For example, western blotting and RT-PCR.
- Histological analysis. For example; analysis of inflammation with hematoxylin and eosin staining.
- Muscle cell-based *in vitro* assays. For example, cytotoxicity and activity screening.

2.3 AGADA Biosciences and Dalhousie University

AGADA Biosciences shares labs with Dalhousie University and some facilities are maintained by animal care staff from Dalhousie University. Drs. Nagaraju and Hoffman, the co-founders of AGADA also work closely with Dalhousie University animal facility staff and research faculty.

They hold Adjunct Faculty appointments in the Department of Pharmacology at Dalhousie University ("About Us | AGADA Biosciences", 2017). The Sr. Director of Research, Dr. Kitipong Uaesoontrachoon also holds an adjunct faculty appointment in the Department of Pharmacology at Dalhousie University.

AGADA works with the Dalhousie University Co-op office to provide internship opportunities for Dalhousie students. The author got this job through the MyCareer web page created by Dalhousie university.

3. Description of the Performed Work

3.1 Internship Description

This internship took place at AGADA Biosciences in Halifax, Nova Scotia. The targeted faculty and disciplines for this internship based on the description were Computer Science and Masters of Health Informatics.

The main objectives of this project were to design and build an AGADA Biosciences Database to store data that was generated from pre-clinical studies in order to facilitate easy retrieval of the preclinical study outcomes.

The requirements for the intern were:

- Working independently with little supervision.
- Experience with database development
- Familiarity with both front and back end web development.
- Having the ability to research and recommend technology and technical tools.
- Working within a multi-disciplinary team environment.

- Willingness to perform and execute any other company and research activities that are assigned by the company management in a timely manner.
- Having the ability to be proactive, highly organized, and capable of multi-tasking, along with qualities to handle a fast-paced environment.

3.2 Internship Role and Responsibilities

As an intern from the Masters of Health Informatics program, the author was assigned a role as a software developer to create an internally used database for AGADA Biosciences.

The author was also responsible for:

- **Collecting requirements and specifications from the employer.**

The first step for building the database was to create data requirements document (David McCaldin, 2015). Through discussions with the employer and team members, the author recorded all their requirements for the database and created a requirements document for AGADA Biosciences.

- **Creating a database design plan.**

After analyzing those requirements and doing some research, the author created an informal database design plan with 10 steps. The database design plan described how the author would design and build the database for AGADA Biosciences. In addition, author also created slides to present this plan to the employer.

- **Creating data models for building database.**

There are three data models that were built by author; a conceptual model (displayed in Appendix 1), logical model and physical model (displayed in Appendix 2). Based on the data

requirements, author also drew a ERD (entity-relationship diagram) which is used to demonstrate the relationships between AGADA's data.

- **Researching and recommending relational database management system.**

The author did some researches and created some comparison slides for AGADA. In those slides, the author evaluated the current state of AGADA's data and compared three database management systems; Oracle SQL, Microsoft SQL and Mysql.

- **Researching and recommending a database server.**

After deciding on database management system, author was assigned the task of researching and recommending a database server. Based on the specifications of the selected database management system and discussions with AGADA's IT service person, the author recommended a database server to install the database management system.

- **Installing and implementing the relational database management system.**

After buying a database server, the author got the remote administrator account for installing the selected database management system onto the remote server.

- **Creating tables and importing existing AGADA's data into the database.**

Based on the created physical database model, the author created all of the tables in the database and imported all of the organized spreadsheets into the database.

- **Designing and creating a database application.**

The author created a web application for AGADA Biosciences that was designed to facilitate use by AGADA staff without prior SQL knowledge. (displayed in Appendix 3)

- **Training team members to use the database management system.**

Getting information from the database and generating customized reports required that the users to learn how to use the database management system. The author wrote a set of database

instructions (displayed in Appendix 4) and set up two training sessions for the AGADA members.

- **Meeting weekly with employer to report the work progress.**

The author was required to meet with the supervisor, Amanda Mullen weekly to report on the progress of the database project.

- **Adding features to existing website.**

AGADA has a new existing website which has not been launched. The author made some changes to its style and wrote some website instructions for AGADA members.

3.3 Internship achievements

The key achievements the author has gained from the internship can be summarized into 5 categories.

1. Put knowledge into practice.

During the internship period, the author got the chance to use knowledge learned from the health informatics program into a real job. For example, lessons learned from course, health informatics flow, was applied to creating database models and knowledge from course, health informatics statistics, was used to analyze AGADA's data.

2. Communication skills.

Communication was a very important tool for understanding the employer's requirements and for expressing the author's thoughts. The author could not build a database to satisfy the employer without understanding their requirements. In AGADA, the author was the only employee who had a background in Health Informatics. Good communication between personnel with different backgrounds was vital. In order to meet the employer's needs, the author had a lot of communications with AGADA staff members which improved author's communication skills.

3. Teamwork skills.

Building a database required a lot of teamwork. The author and the AGADA team set a goal of building an internal database. Everyone in the AGADA lab was proactive in helping to build this database and was willing to help each other to fix any problems generated in the building process.

4. Time management skills.

On the first work day, the author was assigned a google calendar account by the supervisor to schedule daily work. Following that schedule, the author made full use of time and completed each scheduled task on time. This internship helped the author form good habits for time management.

5. Database management skills.

This internship was aimed to design and build a database for AGADA, so author used knowledge learned from the health informatics program. It helped refresh the author's database management knowledge. In addition, the author learned more through research and solving database problems during the internship period.

4. Relation to Health Informatics

4.1 Relation to Health Informatics Courses

Knowledge learned from health informatics program prepared the author for designing and building the database in a proper way for AGADA Biosciences.

- **Understanding AGADA' data.**

AGADA's data is generated from pre-clinical studies. Designing a database for the data required that the author understand AGADA's data and how it was collected. AGADA's data can be

divided into four types: clients' information, mouse information, trial information and outcomes. Understanding the data helped the author organize it into different categories and identify the relationships. Knowledge from course HINF6120 Clinical Care Fundamentals was useful for helping the author to understand AGADA's data. For example, the author was required to recognize muscle names such as, quadriceps or triceps.

- **Creating database models.**

Designing the database consisted of three models; the conceptual model, logical model, and physical model. The database conceptual model was the easiest, and included the entity names and the relationships between those entities. The logical model was more complex than the conceptual model and added the data type to each attribute listed under the entities. The database physical model represented the final look of the relational database, more constraints were added to this model. The entity Relationship Diagram (ERD) was used to demonstrate those models. ERD knowledge learned from course HINF 6101 Health Informatics Flow and Use was applied. Other techniques learned from this course were also very useful, such as creating clinic workflows.

- **Managing the project.**

Designing and building a database was a small project. In the course HINF 6300 IT Project Management, the author learned how to use available information technology in the best way for a project. Methods learned from this course were very useful for designing, preparing, carrying out and completing the project. For example, triple constraints (scope, time, cost) needed to be considered before starting the project planning, team building, communication, and time management.

- **Analyzing AGADA's data in a statistical manner.**

Analyzing AGADA's pre-clinical study data for calculating the standard deviation and p-value required knowledge in statistics and the capability of using statistical software. The course HINF 6030 health informatics statistics trained the author on how to use R Studio (statistic software) to do statistical calculations. It also provided knowledge for analyzing data and drawing conclusions on the calculated results.

- **Data processing.**

Prior to importing AGADA's pre-clinical studies' data into the created internal database, the data needed to be organized and processed. Taking out the meaningless data and keeping data consistent by correcting inaccurate data and checking data completeness improved the data quality in the database. The course HINF 6210 Data Mining for health informatics, discussed how perform data processing which was applied to this internship.

4.2 Relation to the Concepts of Health Informatics

Health informatics is an interdisciplinary program which is designed with the vision of "health as the focus and technology as the enabler" ("Program details | health informatics program", 2017). AGADA Biosciences is focused on facilitating orphan drug development with robust pre-clinical efficacy studies and biomarker support ("Home | AGADA Biosciences", 2017). Building an internal database helped AGADA members to leverage high quality data in an efficient way. Using a database management system enabled them to analyze their studies' data and generate useful information from the data in order to improve the accuracy of their drug efficacy testing in the future. Furthermore, if the accuracy of drug efficacy testing can be improved, the more effective drugs can be propelled through the drug development pipeline faster in order to improve human health.

5. Discussion on a Problem and Proposed Solution

During the internship period, the author encountered some problems which increased the difficulty of accomplishing the project. In this section, author will list all encountered problems and corresponding solutions for each problem.

5.1 Problem Description

Problems happened in the project that can be summarized into following points:

- **Vague requirements**

Most employees in AGADA Biosciences studies were related to biology. They are professionals in the biology area but they are not familiar with the databases. Before author started designing the database, there was a need to collect requirements from AGADA's members. Due to AGADA's members unfamiliarity with database concepts, they used different methods to describe their requirements for the database which lead to some misunderstanding of the requirements. On the one hand, the author misunderstood their requirements. On the other hand, the database terms used by author to describe the database made them more confused.

Author designed the first draft of database for AGADA with those vague requirements which could not be satisfied by employer. More changes were made to fix the draft which took a lot of time at the first stage of designing the database.

- **Data quality**

The quality of pre-clinical study data impacts every conclusion and decision made after a trial ends. Accurate and reliable data is always required in analyzing data to get accurate results. In order to get accurate, meaningful information from the database, data should be processed before

being loaded into the database. The problems of AGADA's pre-clinical data were: data inconsistency, missing data, data inaccuracy, data incompleteness, and vague data relationships.

- **Data inconsistency:** Tables in the database have columns. Those columns have been assigned a fixed data type such as NUMBER, VARIABLE and FLOAT. The data type of a column has been decided in the database physical model. If one column's type is NUMBER, data in this column cannot be a character or other symbols. However, in AGADA's data which is stored in the spreadsheets, data type in one column could be inconsistent. For example, in a NUMBER type column, they may put 'NA' into that column which results in a loading error.
- **Data missing:** There was some data missing in the existing AGADA pre-clinical data, and a need to find them from the raw data which increased workload. Some data is important for AGADA, but for some reason they may not have recorded some data into their spreadsheets causing data to be missing or hard to find.
- **Data inaccuracy:** Data inaccuracy happened because the AGADA did not define the data in a proper way. There is one column of data named "treatment week", it means the start week for administering treatments to the test subjects. Some people in AGADA defined the first week of treatment as week=0, but others defined the first week of treatment as week=1 which led to data inaccuracy.
- **Data incompleteness:** Some trials are still in process, so some spreadsheets are not complete and some data is incomplete.
- **Vague data relationships:** In order to build a database, the author needs to understand the relationships between data. In AGADA's existing spreadsheets,

the relationship between the data is vague which increased the difficulty of the database design.

- **Time conflict**

Usually AGADA Biosciences co-founders, Drs. Kanneboyina Nagaraju and Eric P Hoffman are not in Halifax, Nova Scotia. AGADA employees report their work through an online software.

Each time some decisions should be made for the database project, the author needed the supervisor, Amanda Mullen to set up a meeting to report to the co-founders. When they have other tasks on hand, the meeting times were postponed.

Each member in AGADA Biosciences also have their own roles to play, so some tasks related to database could not be finished on time.

5.2 Proposed Solution

Health Informatics Solutions for each problem are:

- **Vague requirements**

Working within a multi-disciplinary team environment requires strong communication skills.

Misunderstanding comes from miscommunication and lack of communication. The solution for this problem should focus on the methods of communication. Some methods the author thinks would be useful for communicating with AGADA members:

- **Using easy words:** Most of them are not familiar with database terms, so changing terms into easily understood words is beneficial. For example, change the term “foreign key” into “an id link to another table”.
- **Draw information flow diagrams:** If the author could not describe an idea clearly to AGADA staff members, it was better to illustrate it. For example, the

relationships between data was hard to describe in words, author drew the entity relationship diagrams (ERD) to show it to others.

- **Creating planning list:** From course IT project management, the author learned that planning is a very important step. Creating a list of planning steps for building a database can help AGADA members understand the database better.
- **Draw BPMN diagram:** BPMN diagram can help the author learn how they get their studies done and where is the data generated. Knowing more about the data improves author's understanding of their requirements.
- **Data quality**
 - **Document data quality issues:** Documenting all encountered data quality issues is good method for AGADA members to avoid creating those issues again.
 - **Design data entry form:** Data quality problems usually take place at the data entry phase, so a well-designed data entry form could be helpful. A data entry form can guide the user to enter their data in a defined order and in a good format. It will keep user from entering inconsistent type of data and missing data.
 - **Pre-processing data:** Pre-processing data before loading it into the database to avoid data quality issues in the database.

- **Time conflict**

Solving the time conflict problem is not easy. AGADA Biosciences is a company provides pre-clinical testing services, so pre-clinical testing tasks are more important. The author thinks creating priority scheduling is useful. Priority scheduling is scheduling a task based on its priority which can help people manage time better.

6. Conclusion

As a student from Master of Health Informatics, the internship at AGADA Biosciences was a great experience. It provided the author the opportunity to design and build a database in the real world. The author achieved following things through this internship:

- Main objective of this project was achieved, an internal database is built.
- Knowledge learned from Health Informatics program has been applied to the internship.
- Experience working with a dynamic, fun, tight-knit and committed team of researchers.
- Experience working with front and back end web development.
- Using Health Informatics methods to solve problems.

7. Recommendations

Lessons learned from this great internship experience will be beneficial for the future.

- Using system thinking and breaking a complex project down into small tasks can increase project success.
- A good teamwork is vital for a project success and for the future work.
- Health Informatics solutions are very useful to solve some projects' problems. So, do not waste the knowledge learned in the class and apply them into the real world.
- Health Informatics can improve the healthcare delivery with technology. However, Health Informatics concept is not accepted by all companies' related to health, more work should be done to add this concept into healthcare.

Reference

About Us. AGADA Biosciences. (n.d.). Retrieved on April 20th, 2017 from: <http://www.agadabio.com/about-us/>

David McCalidin. (2015). Tutorial: Step by Step Database Design in SQL. Retrieved April 20, 2017, from <https://www.linkedin.com/pulse/tutorial-step-database-design-sql-david-mccalidin>

Home. AGADA Biosciences. (n.d.). Retrieved on April 20th, 2017 from: <http://www.agadabio.com/>

Program details. Master of Health Informatics. 2017. Retrieved on April 20th, 2017 from: <https://www.dal.ca/academics/programs/graduate/health-informatics/program-details.html>

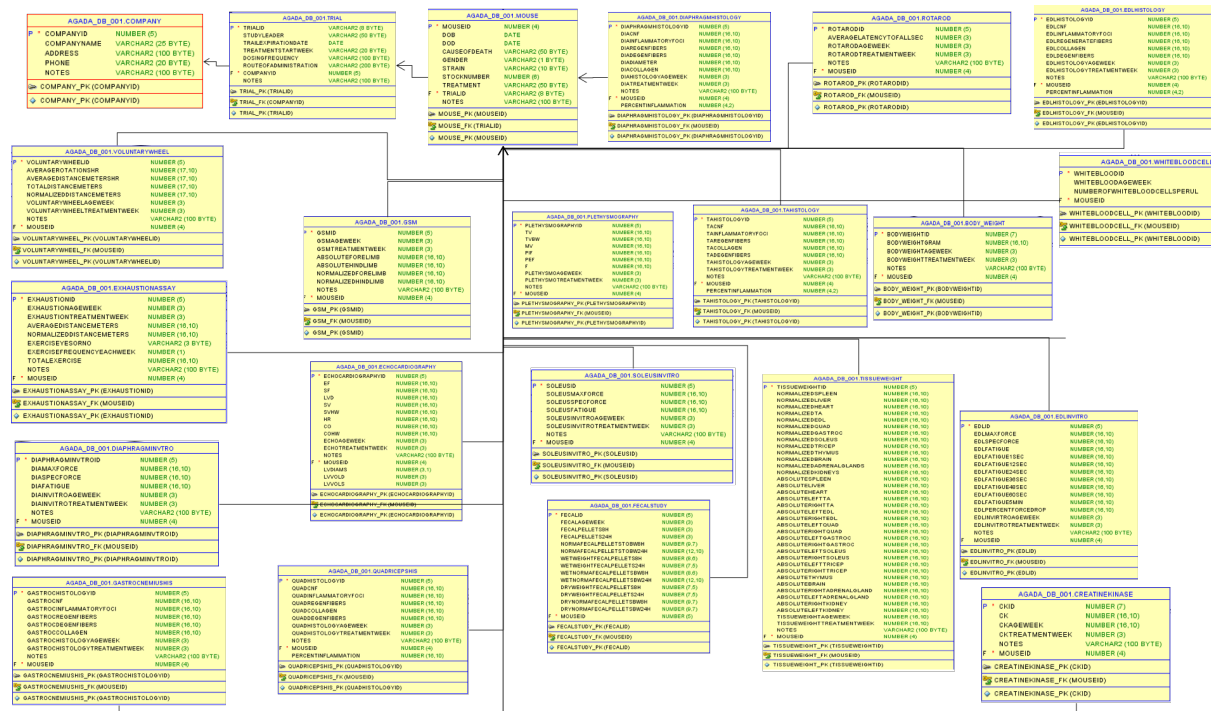
Services. AGADA Biosciences. (n.d.). Retrieved on April 20th, 2017 from: <http://www.agadabio.com/services/>

Appendix

1. Conceptual model designed by the author



2. Physical model designed by the author



3. AGADA web application designed by the author

agadauser											Log Out
Home	Search		Go	Actions							Create
Company											
Trial											
Mouse		36	11-MAY-15	16-JUL-15	Endpoint Dissection	M	mdx	1801	SMT022357 - 30mg/kg	2015-012	Breed at AGADA biosciences
Body Weight		39	11-MAY-15	16-JUL-15	Endpoint Dissection	M	mdx	1801	SMT022357 - 30mg/kg	2015-012	Breed at AGADA biosciences
GSM		41	11-MAY-15	17-JUL-15	Endpoint Dissection	M	mdx	1801	SMT022357 - 30mg/kg	2015-012	Breed at AGADA biosciences
Echocardiography		42	08-MAY-15	15-JUL-15	Endpoint Dissection	M	mdx	1801	SMT022357 - 30mg/kg	2015-012	Breed at AGADA biosciences
Plethysmography		34	08-MAY-15	14-JUL-15	Endpoint Dissection	M	mdx	1801	SMT022357 - 30mg/kg	2015-012	Breed at AGADA biosciences
Rotarod		38	11-MAY-15	16-JUL-15	Endpoint Dissection	M	mdx	1801	SMT022357 - 30mg/kg	2015-012	Breed at AGADA biosciences
Voluntary Wheel		45	08-MAY-15	16-JUL-15	Endpoint Dissection	M	mdx	1801	SMT022357 - 30mg/kg	2015-012	Breed at AGADA biosciences
Exhaustion Assay		46	11-MAY-15	17-JUL-15	Endpoint Dissection	M	mdx	1801	SMT022357 - 30mg/kg	2015-012	Breed at AGADA biosciences
Creatine Kinase		50	11-MAY-15	15-JUL-15	Endpoint Dissection	M	mdx	1801	SMT022357 - 30mg/kg	2015-012	Breed at AGADA biosciences
Tissue Weight		201	14-SEP-17	03-SEP-15	-	M	mdx	1801	Group Q	2015-013	-
EDL in vitro		202	15-SEP-17	04-SEP-	-	M	mdx	1801	Group Q	2015-	-

4. Database instructions created by the author

AGADA Biosciences Oracle Database Instructions

Xiaoqing Zhuang

March 22, 2017

Contents

- Login Oracle APEX 2
- Create New Table in the Database 2
- Import spreadsheets into AGADA Database..... 5
- Export spreadsheets from AGADA Database..... 8
- Insert, Delete, Modify Rows in Database through Oracle APEX..... 10
- Insert, Delete, Modify Columns in Database through Oracle APEX..... 12
- Create, Select Columns, Filter, Sort, Modify, Delete in AGADAUSER Application..... 14
- How to Query the Database and Create Reports 18
- How to export the whole database for backup 26