Bioinformatic Interface - Rat Maze Database

By

Mark Nelson

Submitted to
the Faculty of the Information Engineering Technology Program
in Partial Fulfillment of the Requirements for
the Degree of Bachelor of Science
in Information Engineering Technology

University of Cincinnati
College of Applied Science

June 2006

Bioinformatic Interface - Rat Maze Database

by

Mark Nelson

Submitted to
the Faculty of the Information Engineering Technology Program
in Partial Fulfillment of the Requirements
for
the Degree of Bachelor of Science
in Information Engineering Technology

_____	_____
Mark Nelson	Date
_____	_____
John Nyland, Field Svc Asst Professor	Date
_____	_____
Patrick C. Kumpf, Ed.D. Interim Department Head	Date

Table of Contents

1. Project Description and Intended Use

Dr. Jeffery Johnson, a researcher in the College of Engineering, has a facility at the Genomic Research Institute (GRI) which houses an animal maze which is able to collect large quantities of information through laser sensors. This maze organizes the data in comma delineated files produced by proprietary software (1). He also has a set of servers in the College of Engineering which are ready to house the data. He wanted the data to be easily accessibly by multiple facilities.

The two major areas of this project considered were the functionality and the quality of the solution. Each of these areas were further separated into sub-areas. For functionality, the sub-areas of features, capabilities, and security were addressed. For quality, the sub-areas of usability, reliability, performance, and supportability were addressed.

To start addressing functionality, the project looked at the features that will be required for this project. To start with, the data which is collected at GRI needs to be easily inserted into the servers found in the Engineering College. This data needs to be safe, secure, and accessible for data mining. The data also needs to be accessed from multiple platforms. In short, the data needs to be organized and available to Dr. Johnson's research team and to others who are granted access by Dr. Johnson.

To address capabilities, the system created to fulfill Dr. Johnson's desires needed to organize the data. The system needed to allow Dr. Johnson's team to access the data through programming (allowing them to create their own custom made data mining tools). The system needed to have a system for doing simple queries on the data. The system needed to be available to researchers, granted access by Dr. Johnson, to access the data from their own computer system. Of course it will need to be written in English and have a set of help files to guide the user through the system. It needed to be user friendly, for those researchers already familiar to the data being stored in the servers. In other words, the system needed to be a database which will have the accessibility through the Internet.

The last factor in functionality is security. This system needed to be open to two different users. The first group of users will already be behind the network security of the University of Cincinnati, such as their firewall, routers, and viral and mal-ware

protection. The second group will be users found outside of the University of Cincinnati. For the first group of researchers, the system needed to be available so they can design their own custom project and access the data. This means the data needed to be available on Campus network systems, but the physical location needed to be in a secured location. The second group of users will be located outside of the university network, so a secure system of accessing the data needed to be available, such as a secure login with server side Web pages being delivered to the privilege-granted researchers outside of the university.

The first aspect of quality is usability. The system setup needed to be user friendly. The group of users will already be comfortable with the use of Web browsers. They will all have access to Internet Explorer as a browser. They will all have operating systems with the .Net Framework. Their knowledge-background will already encompass the data being stored in the system. So, the system just needed to have a logical login and directory system to allow the user to use it. The system needed to have a set of help pages to explain the contents of the interface. Also, the system needed to have proper data management documentation for the researchers in Dr. Johnson's team to be able to create their custom data mining tools.

In addition, the solution needed to be reliable. It needed to be available to the researchers at all times. It needsed to have proper backups and services. The navigation system for the system needed to be uniform, so the researcher doesn't have to spend too much time learning a new system. The system needed to be reliable when it comes to keeping the data organized.

Performance also needed to be addressed. The number of researchers accessing the data will not be high volume. The performance needed to be inline with all other online systems. There needed to be reasonably short wait time for system responses. The system needed to be able to process the large amounts of data on the server-side and through put the results to the access terminal. The result from simple queries and data inserts needed to be accurate. The data needed to be available all the time (allowing downtime for routine maintenance). And as mentioned before, the resource usage will not be very high; only a couple of researchers are expected to access the system a day.

The last quality factor is supportability. The system needed to have an organization system which will be able to adapt if future data needs to be added. The system needed to be maintainable by a single person. The system needed to remain inside the University of Cincinnati network system, but allow for limited access by researchers granted permission. The system needed to configure correctly to multiple OS platforms, but always through Internet Explorer. All of these functions will need to be maintained by the system administrator through the engineering servers themselves, which means other than granting researchers usernames and passwords, the system admin will address problems through the programming and server, not through a special login.

By addressing all of these factors for Dr. Johnson's needs, the problem statement elucidated the needs of the system. Areas such as Functionality and Quality were addressed and boundaries were established for the scope of the project. The next step will be addressing these needs.

2. Design Protocols

2.1 Description of the Solution

My solution to Dr. Johnson's problem was to design a database for the storage and simple querying of the data, called Johnson Lab Database. It was developed in such a way that the learning curve will be small. The support will be handled by me for the first five years and proper documentation will be generated, such as ERD's and original comma delineated data files. The user will be accustomed to Web pages, so as not to have a problem with the simple navigation homepage. The programs will also have a simple step-by-step process for entering data.

Some of the key features of this solution were be:

- Easy data dump for the comma delineated files generated by the radial arm maze
- Log in to establish the roles granted
- Simple searches of already existing data
- Ability to log in outside of the University of Cincinnati
- Ability for Dr. Johnson's lab to write specific algorithms for the database

- Ability for users to request a change in the data
- Users without usernames and passwords will be able to request a username and password
- The Admin will be able to grant a log in account with the correct permissions

Johnson Lab Database has a data entry program which will be located on the computers of specific Johnson lab employees. These computers will be within the UC network and will be protected by Windows login, locked doors, and UC security guards. This will control the system the data entry and help promote security.

Some of the key concerns:

- *Increased Security*

  Computers with the program are locked down with windows usernames and passwords. They will also be located in secure location.
- *Data Entry*

  The program allows the user to search for previous records before adding a new one. .
- Data Modification

  The users have the ability to request modification for data already entered.
- *Easy Navigation System*

  The user will have and easy to use navigation page.
- *Maintainability*

  The programming language which was agreed upon was C#. This is the language which is taught at the College of Applied Science and was agreed upon to lessen the learning curve involved in development of the database.

Johnson Lab Database has a Web interface for analysis accessibility within the University and outside the University with a unique username and password. With this capability, Dr. Johnson and his lab group are able to analyze their data from the main campus or at GRI. This means if they need to run an ad-hoc query from the UC campus or outside, they won't have to return to the Biomedical engineering network to run the query.

Some of the key concerns:

- *Increased Security*

  The access to different functions are limited to the privileges granted to each user.
- *Easy Navigation System*

  After logging in, the user is directed to their granted home page. Each group of users have the same standard home page and all navigation will occur through this home page. This allows for easy use of the back button, if the user is comfortable with it, and will control the navigation.
- *Web Access*

  As mentioned before, the major interface for outside users for this database is Web based. Because Internet Explorer is the major Web browser used by Dr. Johnson's lab, this will be the browser the Web pages are designed to be viewed in.
- *Maintainability*

  The programming language agreed upon was C# and ASP.net. These are the languages which are taught at the College of Applied Science and were agreed upon to lessen the learning curve involved in development of the database.

2.2 User Profile

There are four user profiles based on the needs of the user.

2.2.1 Data Generator

The Data Generator is the technician at GRI hired to manage the lab rat experiments. They are experts in areas other than computers and data-mining, such as animal behavior and data creation. Their use of the database will be the least of the four groups. They will need to be able to log in. They will need to be able to add new records. They will be able to find old records. They will be able to add data to old records. They will be able to review the data they added, not the comma delineated data which will needed to be mined, but the simple record keeping records. And they will be able to request changes to the previously created records. They will not be able to permanently

change or delete records themselves, because, as stated before, their expertise lies outside the areas of data management and manipulation.

2.2.2 Johnson Lab Data Analyst

This will be the researchers who have a background in bioinformatics and data mining. They will have the same abilities as the Data Generator, but will have some extra analytical uses for the database. They will be able to create new records and add data to them. They will be able to find old records, review them, and add data to them. They will be able to request changes to the already existing data. This group will also be able to do two types of data-mining. They will have a simple menu system to compare experiments. Lastly, they will have usernames and passwords to the MS SQL Server so they can create their own custom programmed algorithms when they need to.

2.2.3 Data Analyst Outside of the University

These users will be identical to the UC Data Analysts except they will not be given access to the database directly. This is due the security required for the database. One of the aspects of security is that the system will remain within the UC network. Giving outside users access directly to the database would be a security risk.

2.2.4 Administrator

This is the user with the greatest range of tools on the database. This role will only be granted to trained database administrators who will be specifically chosen by Dr. Johnson. This user will be granted all of the previously mentioned privileges. But this user will be the person who grants usernames, passwords, and changes to the database. This adds more work to this user group but adds an extra line of security to protect the integrity of the data.

** Table changes will not be managed through this Web interface. Changes will only be allowed by Administrators who have access to the locked server room.

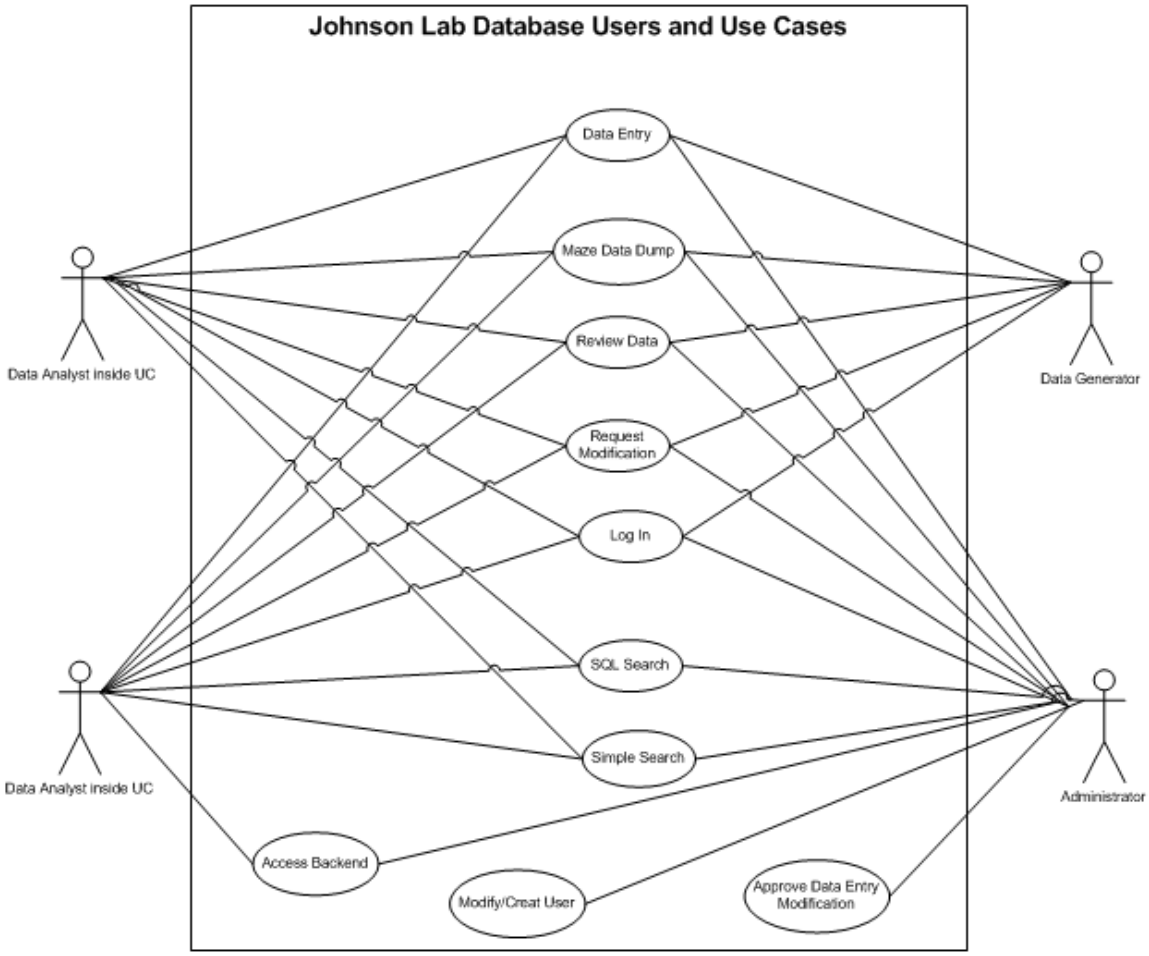
2.3 Organizational Scheme

Johnson Lab Database Web pages are divided into three different home pages depending on the level of access the username and password grants. The three groups are:

- Data Generator
- Data Analyst (both UC and outside-UC users)
- Administrators

The home page is the base of all navigation. From here, the user is able to move to their different functions, such as Create, Add, Review (includes Request Change), Simple Search, Modify/Add User, and Approve changes. A more detailed description of the Use Cases is documented in Figure 1.

Figure 1: Johnson Lab Database Use Case Diagram

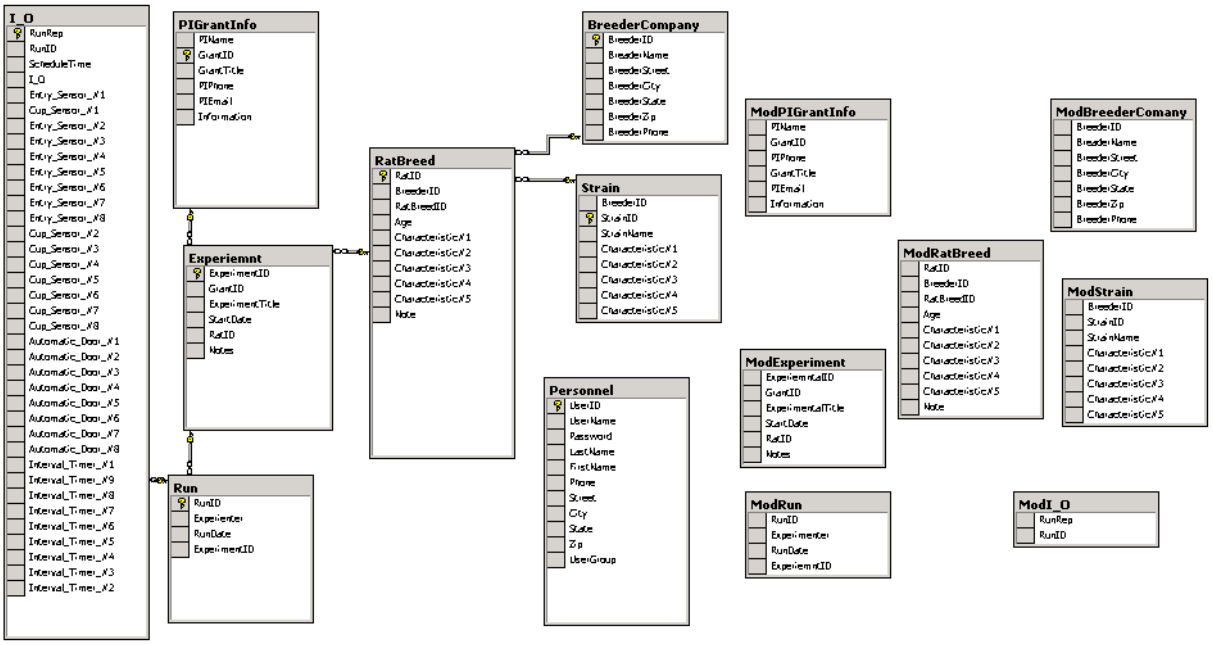


2.4 Database Design

The database design is not a complex system, but does incorporate the use of normalized relational tables. These tables have one-to-many and one-to-one record relationships. There is some redundancy. Dr. Johnson has asked that the animal maze movement data (comma delineated flat file generated by radial arm maze) is stored in the database in its original form as well as in relational tables. Most of the tables have indexed, auto-incrementing primary keys, linking PI/Grant information, Experimental Information, Experimental conditions, and Animal Information. A redundancy is created to hold submitted changes to the already existing tables. This will usually be empty except for the occasional request for changes to the data entry portion of the database. An ERD of the database is shown in Figure 2, Johnson Lab Database ERD. Here you can see that the data collected by the radial arm maze (table I_O) is related the Grant and Primary

Investigator, the experiment, the replicate, and the Rat involved. The Personnel table records who has accessibility to the interfaces.

Figure2: Johnson Lab Database ERD



The Web pages have a primary login page which will access the personnel tables to access if the login is correct and what type of user is logging in. Different links are available to the different types of users. The links are as follows:

Data Generator

- Data Entry – the user will be able to enter in new experimental descriptions, animal types, and experimental runs.
- Review Data – the user will have the ability to review data entry data through the use of drop down lists.
- Request Data Modification – the user will have the ability to request data entry data corrections when typos occur.

Data Analysis (both UC and non-UC Analysts)

- Data Entry – the user will be able to enter in new experimental descriptions, animal types, and experimental runs.
- Review Data – the user will have the ability to review data entry data through the use of drop down lists.
- Request Data Modification – the user will have the ability to request data entry data corrections when typos occur
- Simple Data Search – the user will be able to access maze generated data through the use of a series of drop down boxes.
- Structured Data Search – the user will be able to use structured comples queries.

Administrator

- Data Entry – the user will be able to enter in new experimental descriptions, animal types, and experimental runs.
- Review Data – the user will have the ability to review data entry data through the use of drop down lists.
- Request Data Modification – the user will have the ability to request data entry data corrections when typos occur
- Simple Data Search – the user will be able to access maze generated data through the use of a series of drop down boxes.
- Structured Data Search – the user will be able to use structured comples queries.
- Modify/Create User – This link will allow the administrator to change the user's profile in the personnel table or add a new record.
- Approve Data Modifications – this link will allow the Administrator to review modification requests and approve their changes.

2.5 User Interface

The interface of choice is the use of Internet Explorer designed Web pages and C#.Net to manage the data in the database. From conversations with Dr. Johnson, the

users will be using Internet Explorer and Windows .Net Frame Work Windows operating systems (4).

2.5.1 Interface Design/Navigation

The overall method of moving through Web site will be through the use of each user's homepage. This is where the links mentioned in the previous section are located. Each link leads to a new page and then will return to the home page when completed. This allows for the use of the back button if the user is comfortable using it. It also keeps from having a complicated navigation scheme.

The overall movement through the data entry program uses a straight forward procedure. This is well documented to allow the user to skip steps if the data is redundant.

2.5.2 Color Scheme

The color scheme for this project is red, black, and white. These colors were chosen to match the schemes in the University of Cincinnati Web pages. All of the regular text will be in black for ease of reading and all of the important text will be in red to indicate its importance.

2.5.3 Help

Each Web page and program form has a help link in its top right corner. This page describes the links available on the current page.

3. Deliverables

To provide a well designed and easy-to-use system, a wide range of deliverables were laid out. The following deliverables were the result of the design phase of the Johnson Lab Database:

1. A back-end RDBMS on MS SQL Server 2000.
2. Web interfaces developed in HTML, C##, and ASP.Net
3. Authentication for Data Generators, Data Analysts, and the Administrators.
4. Role- specific navigation home page for ease of navigation of the application.

5. Abilities for Data Generators:
    - Add records into the database for Animals, Experiment Descriptions, and Experiments.
    - Review the data entered manually.
    - Request modifications to the manually entered data.
6. Abilities for Data Analysts:
    - Add records into the database for Animals, Experiment Descriptions, and Experiments.
    - Review the data entered manually.
    - Request modifications to the manually entered data.
    - Do simple data searches.
    - Do Structured data searches.
7. Abilities for Administrators:
    - Add records into the database.
    - Review the data entered manually.
    - Request modifications to the manually entered data.
    - Do simple data searches.
    - Do structured data searches.
    - Create or modify users.
    - Approve modifications to data entry data.

4. Design and Development

The next section describes the project's timeline, the project resources, overall budget including hardware and software resources, and personnel resources.

4.1 Timeline

The project involved several challenges, learning mistakes/opportunities, and accomplishments. Below are my accomplishments of the Senior Design sequence

4.1.1 Senior Design I Accomplishments

During Senior Design I, I accomplished the following:

- Analyzed the needs of the research group
- Researched what research databases were available for the RAM
- Considered the specific needs of the Dr. Johnson's Lab
- Increased my knowledge on Web authoring tools and languages
- Researched Database connection tools
- Setup Operating System on given server.
- Setup OS on said server
- Established network connectivity for server
- Setup IIS and ASP on server
- Installed and setup MS SQL Server
- Prepared Design Freeze documentation and oral presentation

During the RAM database research, I interviewed multiple people, Dr. Johnson4 and Jerrod Bruce1, about existing technology to organize the data. Unfortunately, they confirmed what I found in the line of web searches, nothing existed.

Originally this project was to be set up using non Microsoft products. But after trying for a year to get the Linux, Apache, MySQL to connect to the UC network and allow for the access over the network, the decision was made to move to MS products. This was a major set which required me to start over for studying the database connectivity.

4.1.2 Senior Design II Accomplishments

During Senior Design II, I accomplished the following:

- Database design
- Secure login
- Homepages for both desktop program and Web Interface
- Data entry pages for the desktop program
- Simple search data grids to review the data entry
- Unit testing was begun
- Prepared Design Freeze documentation and oral presentation

As mentioned above, the OS for developing the interfaces was changed during SD I. Developing the database connections took a lot more time than expected. Therefore a lot

of the graphical interfaces were created during SD II and left for SD III to for connectivity.

4.1.3 Senior Design III Accomplishments

During Senior Design III, I accomplished the following:

[To be added later]

4.2 Project Resources

A list of the resources used to complete this project can be found under the Budget Table, Table 1. Additional resources include "Programming ASP.Net" by Jesse Liberty and Dan Hurwitz (5), "Microsoft Visual C# .NET" by John Sharp and Jon Jagger (7), "Beginning C# databases" by James Huddleson (2), and "Programming ADO.NET" by Richard Hundhausen and Steve Borg (3).

4.3 Project Budget

Table 1 is a proposed budget for the project.

Table 1: Budget

Budget

Resource Name	Type		Single Cost	Note
Dr. Johnson Server	Hardware		$2,500.00	Purchased Previously Through Grant
Internet Connection	Hardware		$0.00	Purchased Previously Through Grant
Visual Studio	Software		$900.00	Purchased Previously Through Grant
MS Server 2003	Software		$800.00	Purchased Previously Through Grant
MS SQL Server 2000 Enterprise	Software		$10,000.00	Purchased Previously Through Grant
Photoshop	Software		$350.00	Purchased Previously

						Through Grant
Worker Name	**Hourly**	**Overtime**	**Hours**	**Overtime**	**Cost**	**Note**
Mark Nelson	$5.00	$7.50	140	$0.00	$700.00	Volunteer Work
Prof John Nyland	$80.00	$120.00	8	$0.00	$640.00	Volunteer Work
Prof Robert Schlimmer	$80.00	$120.00	3	$0.00	$240.00	Volunteer Work
				Total	$16,130.00	Free

5. Proof of Concept

As a proof of concept, some of the Web page interfaces have been included. More shall be added as the design progresses. Figure 5 is the login page and Figures 6 is menu page for the administrator type. Figure 7 is the menu page for the data entry program. More interfaces plus their navigation scheme can be seen in Appendix A and B.

Figure 5: Log in

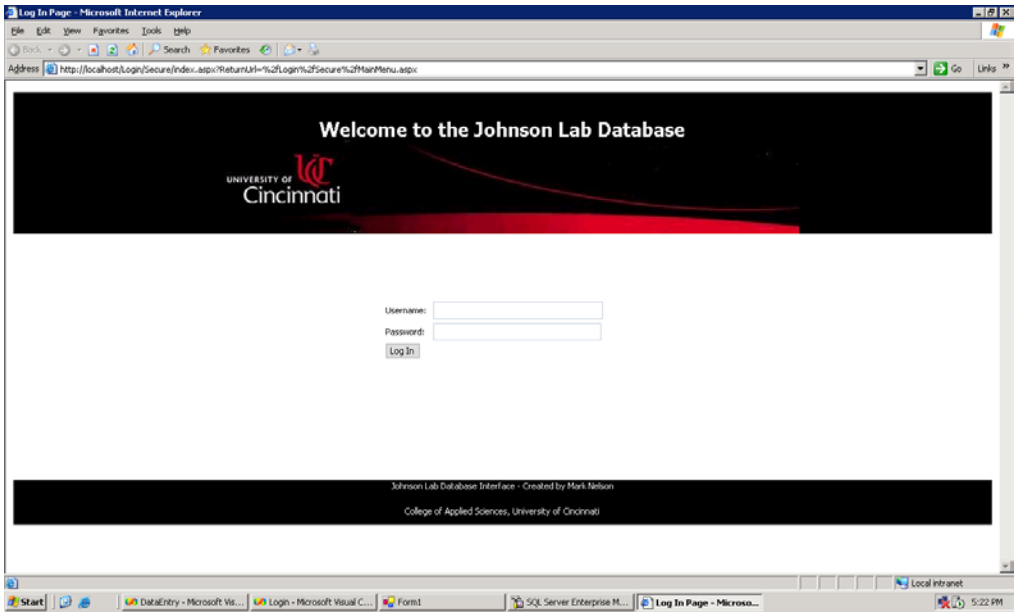
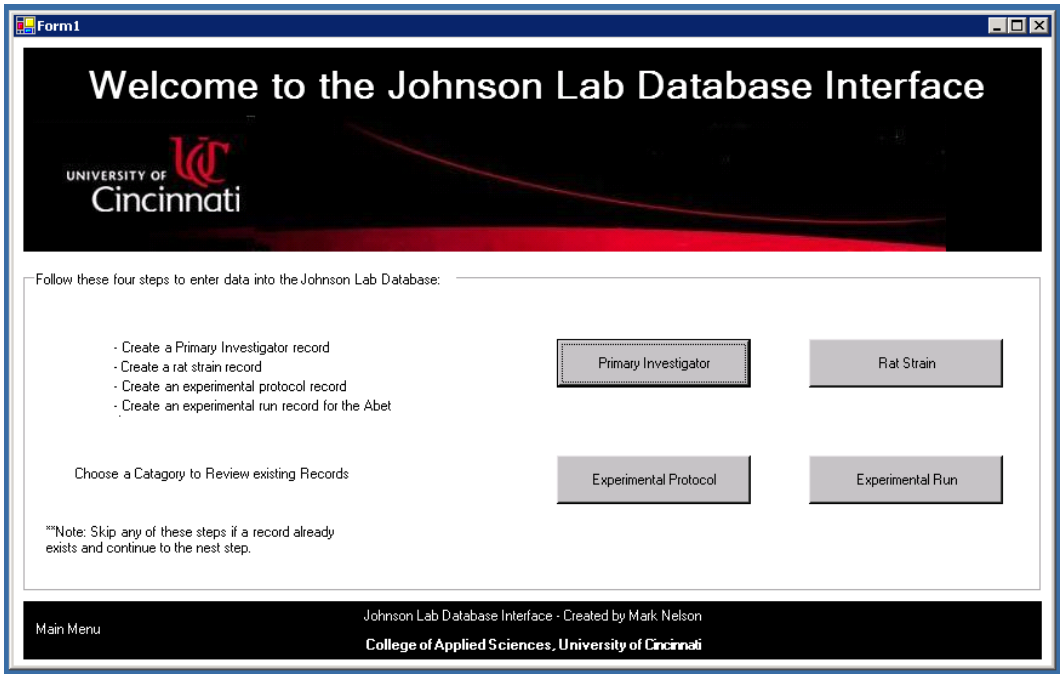


Figure 6: Administrator Menu

Figure 7: Data Entry Menu



6. Testing

The interfaces were developed using an iterative process consisting of HTML code, ASP.Net code, and C#/ADO.Net. The Web site were broken down to separate Web pages. Each Web page was developed first testing the HTML coding, second testing and client-side scripting, third testing the server-side ASP, and lastly testing the ADO.Net

coding. The program was broken down to separate forms. Each form was tested for both correct data entry and for error catching. These tests were completed before deployment. Once the pre-deployment testing was completed the Web pages and program were deployed. The same iterative process was completed to test the post-deployment Web site and program. The Web pages were also tested against World Wide Web Consortium validation, as was suggested by Prof Nyland in conversations with him (6).

To test each section of code, a correct procedure was employed and the result was examined. If the code passed the correct procedure, next a series of incorrect procedures were employed to test the code's robustness. When transferring data to and from the database, the MS SQL server was used to verify the accuracy of the data transferred. To test the search options, a set of test data was entered into the database and a search with a known result was run. The results given by the Web page and program were compared to the known results.

Personnel testing was also performed using the iterative process. The GUI's were first discussed with the personnel who will be using the system. The workflow and business procedures were used to design the structure of the interfaces. At the completion of each page a user from each group was asked to try the interface and gave their opinion of its usability. In the end the users were almost was knowledgable about the interface and database as the developer.

7. Conclusions and Recommendations

7.1 Conclusions

This project was created in response the Dr. Johnson's need for a method of organizing his lab data. I created a user friendly database and interface for the research data created through the use of the Lafayette Biological Instrumentaion's Radial Arm Maze. To prepare the project, I used MS Server 2003, MS SQL Server 2000, Visual Studio 2003, Adobe Photoshop CS2, C#, and ASP.Net. The project was completed over the three quarter Senior Design sequence. The budget of approximately $16,000 would be a reasonable estimate for the completion of this project. The project fulfilled all of the Design Freeze deliverables. Testing was performed to ensure the product's usability.

7.2 Recommendations

As mentioned in a previous section, the platform for development was changed from non-Microsoft to Microsoft products. This was a major challenge for me, I had learned MS product development through the IET program at University of Cincinnati, but never learned any languages outside of MS. I tried for a year to set up the linux and MySQL software. In the middle of Fall 2005, we changed the scope of the project to a Microsoft project. At this time, I had to start over with the reading I needed to do. One suggestion for future students is to have a couple of classes on a product before attempting to develop in it.

I had the same problems when it came to ADO.Net and ASP.Net. Both technologies were only taught for a single quarter. I had a really hard time reading a book and getting everything the writer was trying to say from the static pages. I learn more easily with a dynamic teacher, leading me through the procedures. My second recommendation is to take as many classes in the subject area you are going to use before the end of SD I (before you start your design phase).

8. References

1) Bruce, Jerrod. Field Service Technician and Developer, Lafayette Bioinstrumentations. Personal interview. October 21, 2005.

2) Huddleson, J. (2005). Beginning C# databases. Berkeley, CA: America Springer-Verlag.

3) Hundhausen, R., & Borg, S. (2002). Programming ADO.NET. New York, New York: America Wiley.

4) Johnson, Jeffery. Associate Professor, College of Engineering, The University of Cincinnati. Personal interview. October 14 and 28, 2005, November 11, 2005, and January 13, 2006.

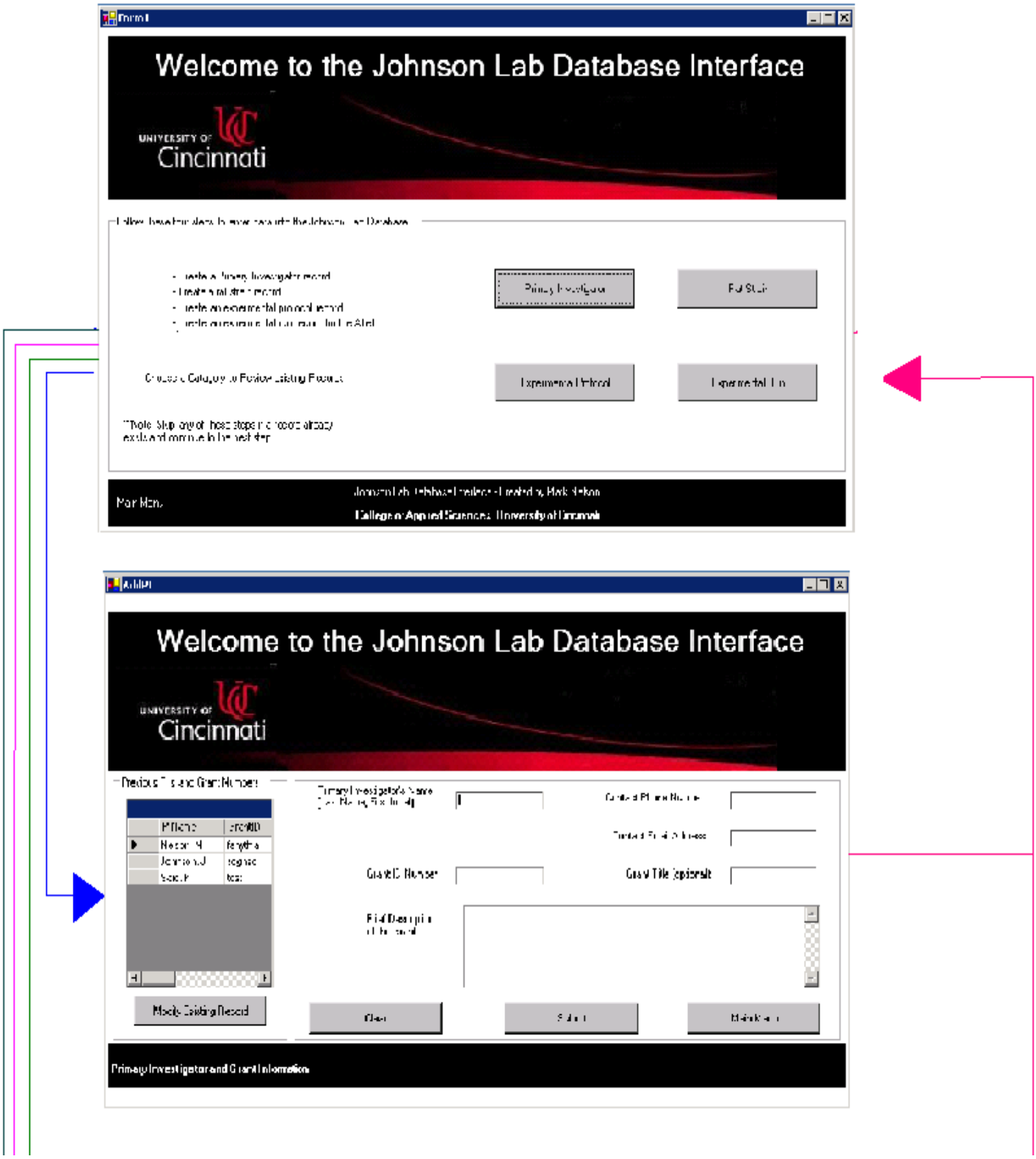
5) Liberty, J., & Hurwitz, D. (2002). Programming ASP.Net. Sebastopol, CA: American O'Reilly & Associates, Inc.

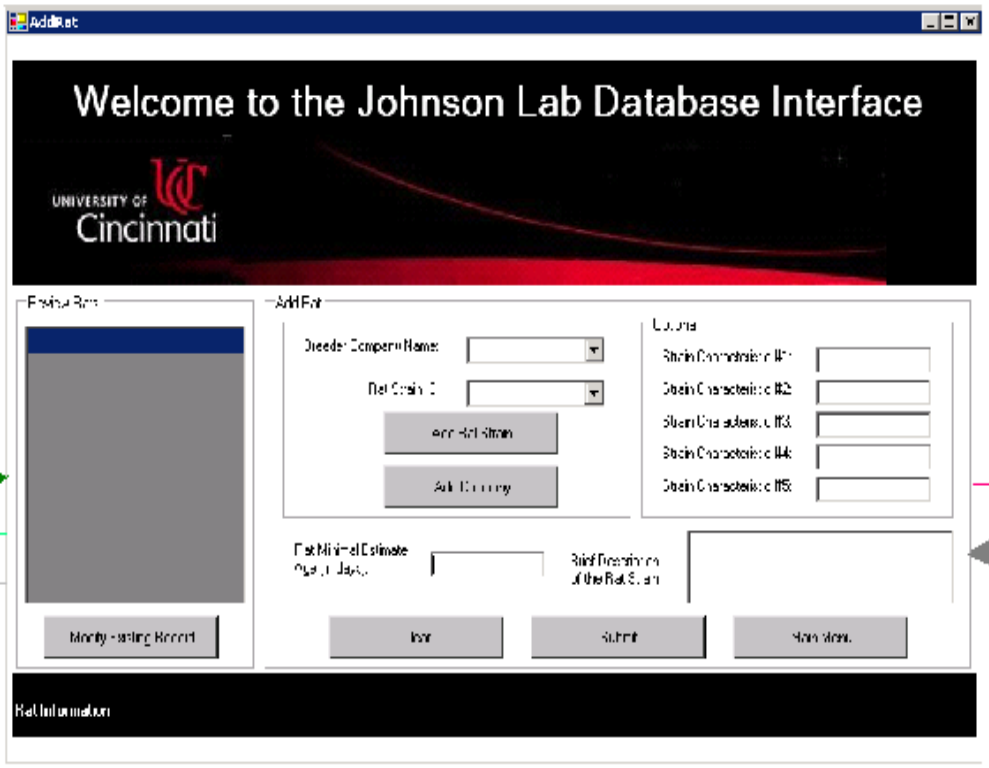
6) Nyland, John, Field Service Professor, College of Applied Sciences, the University of Cincinnati. Personal interview. February 3, 2006.

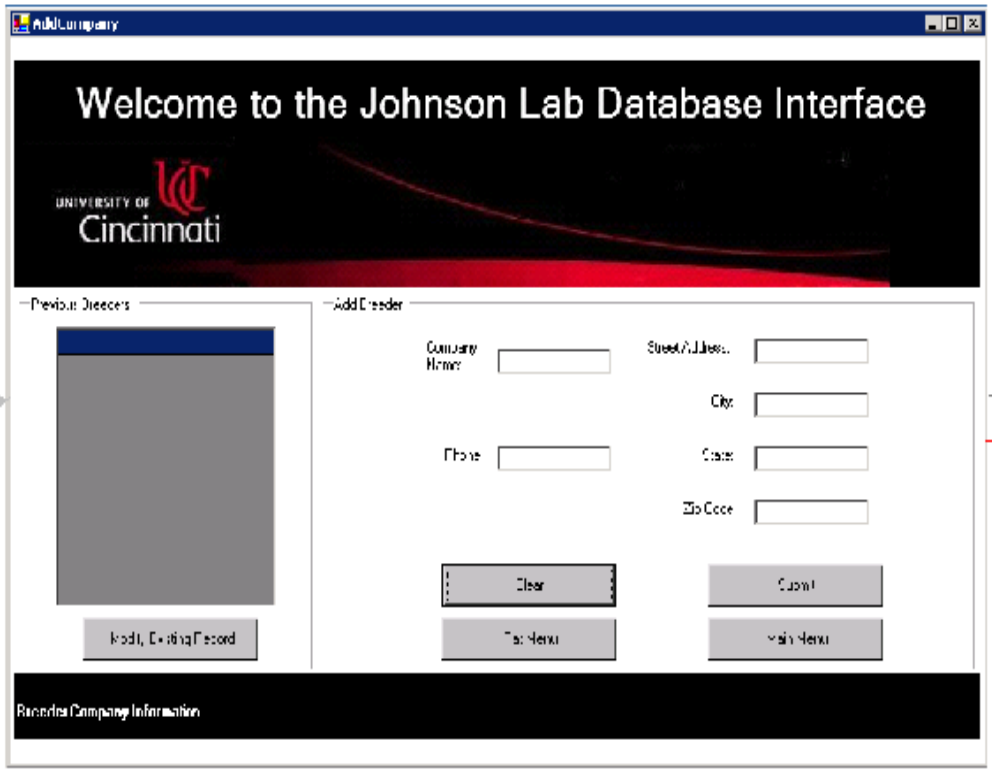
7) Sharp, J., & Jagger, J. (2003). Microsoft Visual C# .NET. Redmond, Washington: America Microsoft Press.

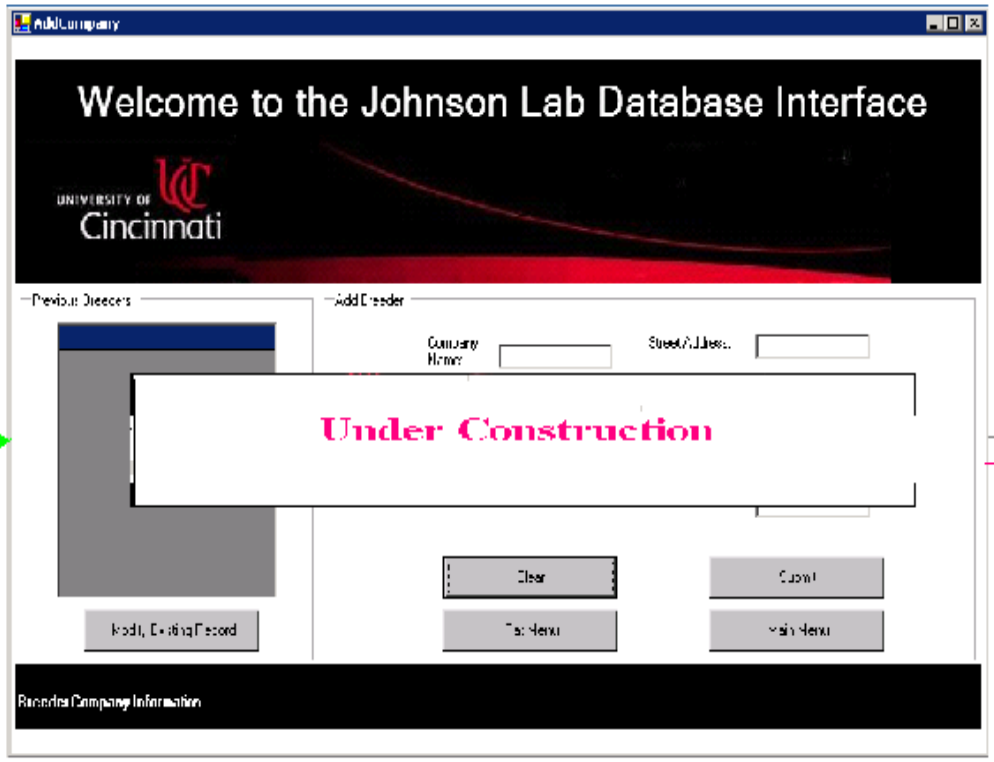
8. Appendix

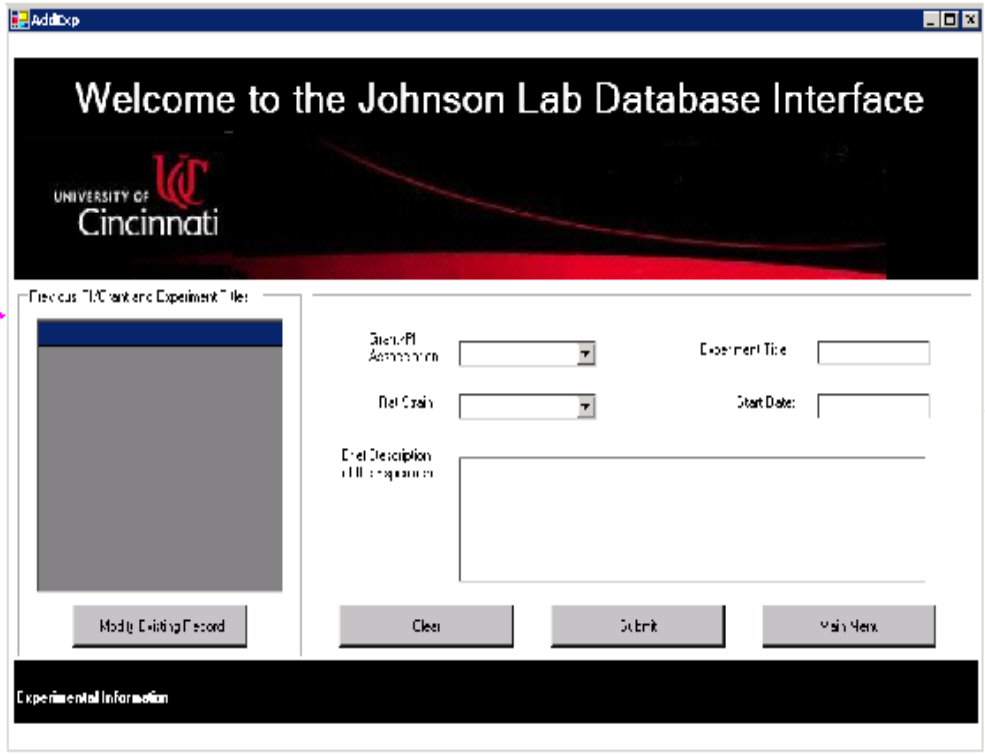
8.1 Appendix A Data Entry Navigation

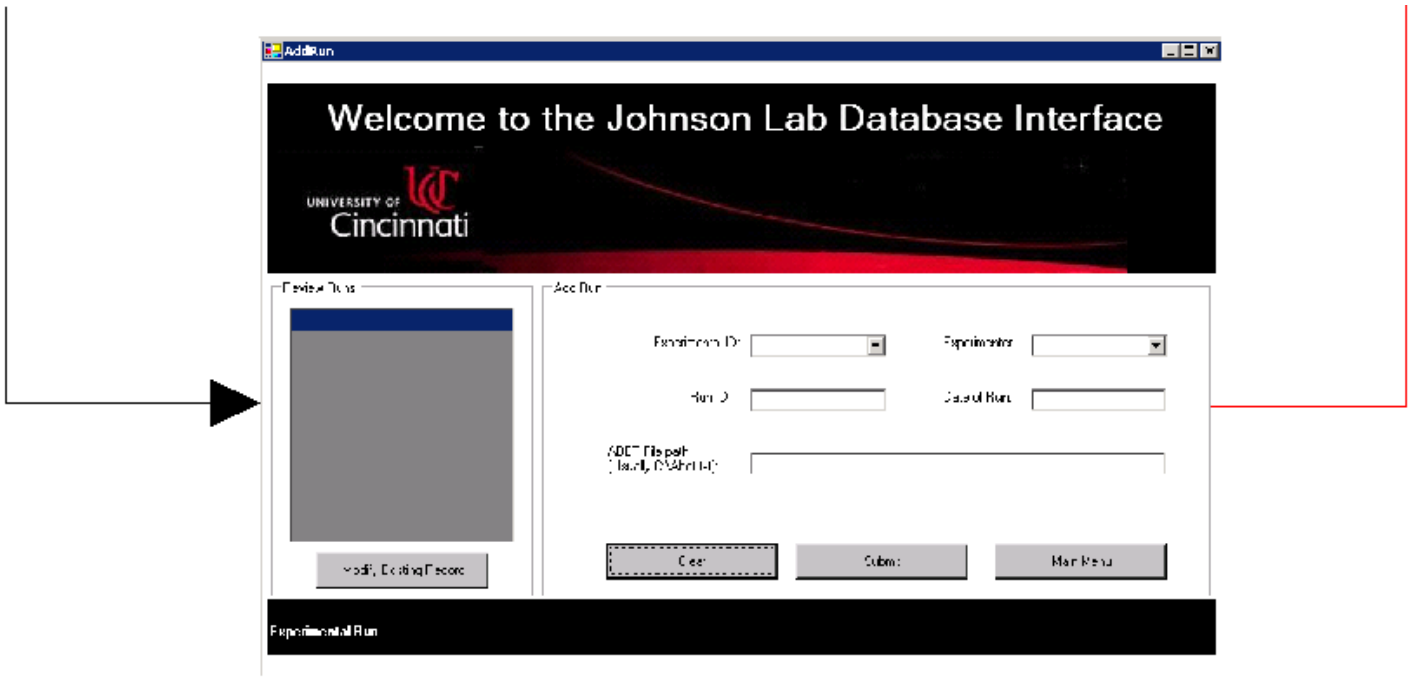

Welcome to the Johnson Lab Database Interface
UNIVERSITY OF Cincinnati
Main Menu
Welcome to the Johnson Lab Database Interface
UNIVERSITY OF Cincinnati
Modify Existing Record
Primary Investigator and Grant Information


Welcome to the Johnson Lab Database Interface
UNIVERSITY OF
Cincinnati

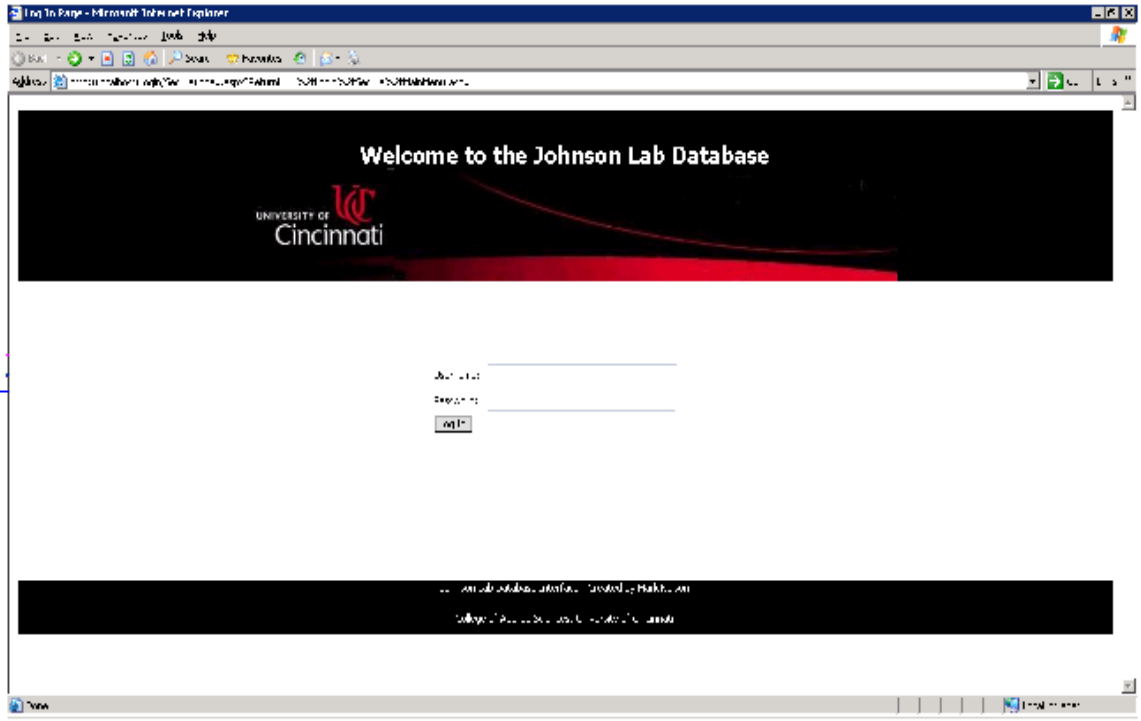

Welcome to the Johnson Lab Database Interface
UNIVERSITY OF
Cincinnati


Welcome to the Johnson Lab Database Interface
Under Construction

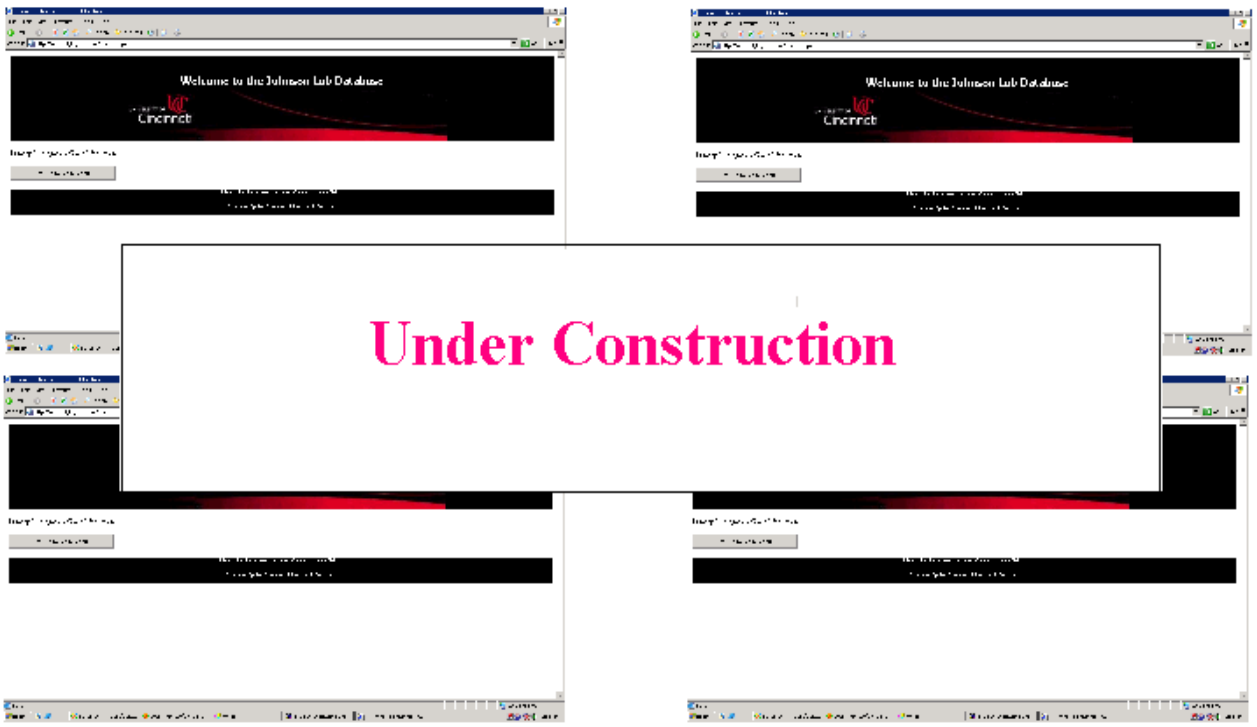

Welcome to the Johnson Lab Database Interface


AddRun
Welcome to the Johnson Lab Database Interface
UNIVERSITY OF
Cincinnati
Add Run
Experimenter
Run ID
Clear
Submit
Main Menu

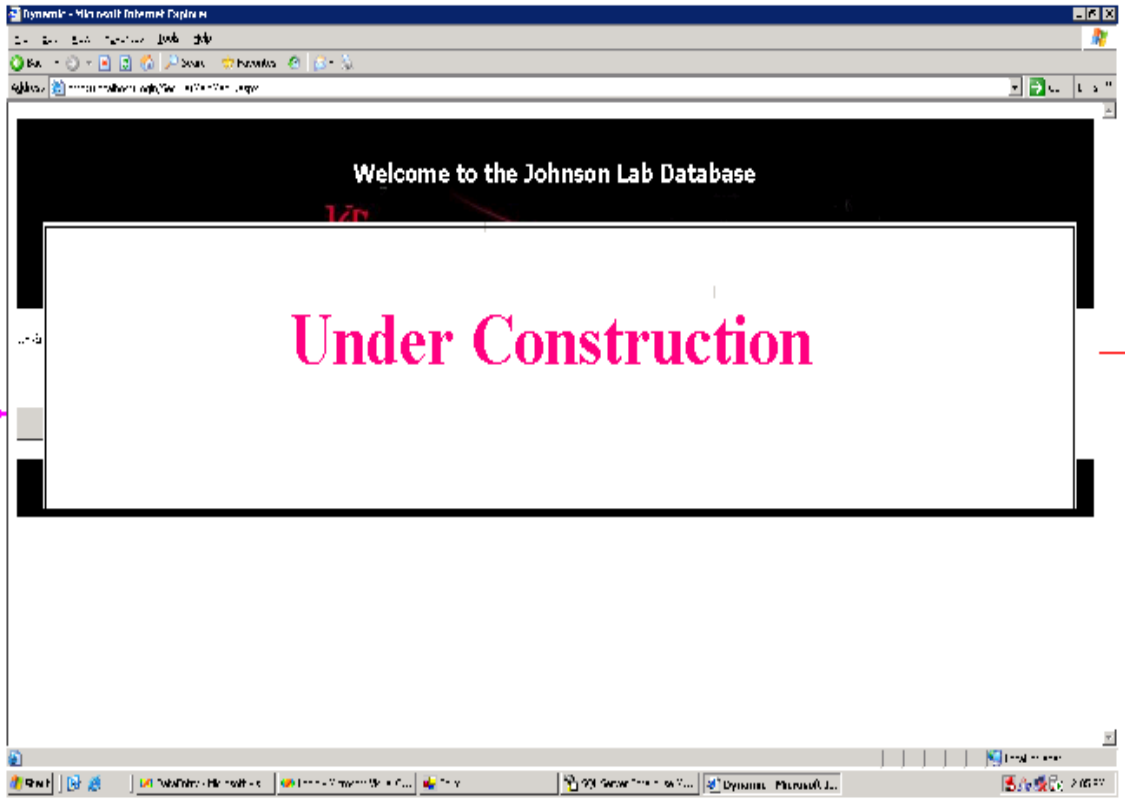
8.2 Appendix B Web Navigation

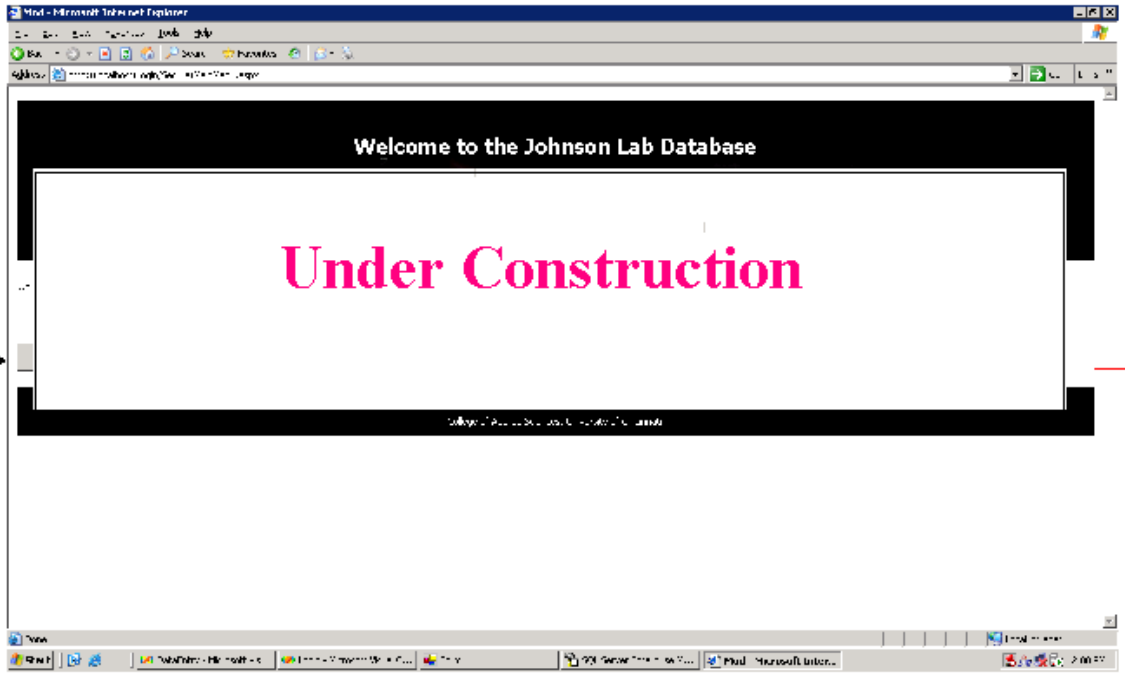



Welcome to the Johnson Lab Database
UNIVERSITY OF
Cincinnati


Under Construction


Welcome to the Johnson Lab Database
UNIVERSITY OF
Cincinnati


Welcome to the Johnson Lab Database
Under Construction


Welcome to the Johnson Lab Database
Under Construction