Applied Information Technology And Computer Science Vol.4 No. 1 (2023) 997-1011

AITCS

Homepage: http://publisher.uthm.edu.my/periodicals/index.php/aitcs
e-ISSN :2773-5141

Design and Development of SDR System using PHP and HTML

Chung Shue Li[1], Salama A. Mostafa [1]*,

[1]Fakulti Sains Komputer dan Teknologi Maklumat,
Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, 86400, MALAYSIA

*Corresponding Author Designation

DOI: https://doi.org/10.30880/aitcs.2023.04.01.057
Received 16 June 2023; Accepted 11 June 2023; Available online 30 June 2023

Abstract: Research Management Centre (RMC) UTHM manages and enhances all the research cultures and activities in the university. The purpose to develop the Scientific Data Repository (SDR), a web-based scientific data repository, is to provide a platform of research data management and archives. The system allows RMC to manage the research data and the researchers can donate or request the research data. This system is developed using C# programming language and based on Prototyping methodology. State the user and modules in the system… Therefore, this system can assist RMC to organize the research data more efficiently and help the researchers to obtain the data they need.

Keywords: Data Repository, Research Data, Web-based Application

1. Introduction

Scientific research is a systematic study that investigates and observes scientific theories and hypotheses. It is an organized and planned process that involves multiple steps to advance and develop knowledge. In university, scientific research is significant for students. It is because by involving them in the research, it enables them to deepen their knowledge and to explore their creative potential [1]. Most of the scientific research is based on collection and analysis of data. Research data is defined as information gathered and analyzed to develop a research result. There are different formats of research data such as documents, image, video, and audio. To manage and store the research data, a data repository is needed. In simpler words, a data repository is a library that collects, manages, and stores data sets. In university, several scientific research of different subjects will be carried out thus the research data collected are huge in number and size. To organise all the data, a data repository is the best option.

In UTHM, all the research activities are enhanced by the Research Management Centre (RMC). Since there is no platform for the storage of research data, the data are stored by the researchers themselves. Most of them are kept in portable devices like CD and disk or cloud service like Google

Corresponding author: salama@uthm.edu.my

Drive and Microsoft One Drive. Some of the researchers did not keep the research data after finishing the research because there was no scientific data repository for them to store the data. There are mainly four types of research data which are observational data, experimental data, simulation data and derived data [2]. All of them need a period of time to be collected and captured. For example, observational data, which can be gathered using surveys, are captured in real time thus it is unable to be recovered if the data is lost.

Portable devices and cloud services are good for storing research data but they are not easily organized and stored. First, most of the data is kept by the researchers themselves thus the RMC cannot collect and manage all the data. Next, the preservation of research data is hard without a data repository because portable devices and cloud services may be damaged and not permanently accessible. Besides, the research data is not open and public to other researchers that may need help in collecting data for their research studies [3]. Therefore, the RMC should have its own platform to organize and store the research data.

Thus, in this project, a web-based scientific data repository, Scientific Data Repository (SDR) is designed to allow the RMC to publish and store the scientific research data. Researchers in UTHM can upload their research data to this repository to enable others to view the data. Using this system, UTHM can manage and archive all the research data efficiently. The data repository will store the data based on different subjects, authors and years of publication. By this repository, the maintenance and preservation of research data can be ensured thus the research data is permanently accessible. Besides, other researchers can also access this repository to view and download the data sets for their research.

The purpose of this project is to develop a web-based scientific data repository for UTHM. Thus, the following objectives are set to achieve the goal mentioned: (i) To design a web-based scientific data repository for UTHM using an object-oriented approach; (ii) To develop a web-based scientific data repository for UTHM using C# programming language; (iii) To conduct testing on (i) that has been built.

This paper is organized into five sections. Section 1 introduces the project background and explains the problem statement, objectives, scope, expected result and significance of this project. Section 2 discusses a review of the literature examining the topics and systems that have been done. While in Section 3, the works and deliverables for each phase are explained. The outputs of each phase will be shown in this section. Next, Section 4 discusses the results and finding of this project. Lastly, Section 5 presents the outcome of this project.

2. Related Work

2.1 Data Repository

A data repository is defined as a library that allows researchers to store their research data sets. It is a type of large database infrastructure to collect, organize and store data sets to carry out the analysis and sharing of research data. Data sharing is important nowaday as it encourages research by providing new questions and sources of data. The research data can be used for second analysis by other researchers thus data sharing is essential in scientific work [4]. Thus, the widespread use of research data should be enhanced by the research institutions. There are a few examples of data repositories such as data lake, data warehouse, data cube, metadata repositories and data marts. A good quality data repository should meet the requirements. First, the data must be preserved for a long-term period, at least 5 years from publications. Next, the data repository should be supported by a research institution or organization. The terms of using an There are some advantages of using a data repository. First, data analysis can be carried out faster and more effectively. It is because all the data is stored and preserved in one location. Researchers only need to analyze the data from one source rather than from multiple sources. Duplicated data can be avoided if researchers can focus on analyzing from one source. Without a data repository,

cross system analysis is hard to be carried out and burdens the process of researching. Besides, research data is easier to organize. A data repository is managed by a research institution or organization. They will manage and maintain the data repository to track any error or problem occurred in the data repository. Digital data is easier to manage compared to data stored in manual ways such as papers or storage devices. Besides, with a data repository, the communication between researchers is better thus the data sharing is easier [5].

2.2 Web-Based Application

Technology used in the SDR system which is developed in this project is the web-based application. Web-based applications can be explained as a software package that can be used through a web browser. Most of the cases, web-based applications are processed over the internet. The applications interact with the users through the Hypertext Transfer Protocol (HTTP). It acts as a medium to allow users to communicate to a remote server using a web browser. Unlike traditional applications, the installation is not required for web-based applications. Thus, usage of web-based applications also known as Software as a Service (SaaS). There are a few examples of web-based applications such as online ticket booking system, Customer Relationship Management (CRM) system and online payment system. In recent years, development and usage of web-based applications rose in a great number and became an essential tool in the business and management department. Web-based can serve many users in the same time better than desktop application [6]. However, there are some places that have poor internet connection. Users from those places do not have a chance to access to the contents of web applications [7].

2.3 Study of Existing System in UTHM

Organizational Research Knowledge Experts Dashboard (ORKED) is an online system to allow researchers to organize their research activities. Before starting carry out any research, they need to submit the form including the details of their research and wait for the verification from the Research Management Centre (RMC). Besides, researchers need to record the activities and expenses during the research. They can make the payment and claim for research fund in the system too. Using this system, the researchers and the RMC can easily track the progress and activities of the research.

However, the research data gathered and submitted to the ORKED system is not available for viewing and download by other users. The sharing of research data is become harder as the data submitted can only view by the uploaders themselves. Sharing of research data is important as it avoids repeated research and allow deeper research toward the topic.

Besides, if there is a platform for research data sharing, other users from other universities or institutions can make use of our research data. As the name of UTHM will be cited in their research report, the exposure of our university is increased among the research area.

2.4 Study of Kaggle

Kaggle offers an online and public platform that allows users to find and publish their data sets. Besides, Kaggle lets the machine learning practitioner and data scientists to discuss on an online community. It offers machine learning competitions that attract many teams and users to participate in the contests. There are a few features developed in Kaggle which are login and register, forum, profile, newsfeed, bookmark and search history.

The first feature is the login and register feature. Kaggle allows new users to register and login their accounts by email and password. To reset users' password, users have to request an email before changing their password. Next is the forum feature which allows users to have discussions on different topics. Inside the discussion, users can post their questions and others can comment below to solve the questions. Users can also bookmark the discussion topics they are interested in and find them under the bookmark tab. Besides, Kaggle users will have their own profile page. In the profile page, users can add and edit their information such as jobs, location, bio and organization. Users can view the

discussions they had participated in and the datasets bookmarked in the profile page. In addition, Kaggle has the newsfeed feature which displays the latest activity. Users can check the latest post or discussion on the newsfeed page. For the bookmark feature, users can bookmark the datasets or discussions and view them in the profile page. Thus, users do not need to search for the datasets or discussions again for the next time. Lastly, users can check their search history under the "recently viewed" tab. Any datasets or discussions viewed will be displayed here.

2.5 Study of UCI Machine Learning Repository

The UCI Machine Learning Repository is a collection of datasets which is created by the students of University of California Irvine. It is well-known as the primary source of machine learning datasets (re3data.org, 2021). It offers features such as login and register, profile, newsfeed and display of paper citing datasets.

For the new account registration, users need to fill in their name, email and password. While for the login, users need to fill in the email and password. If the password is forgotten, users can request an email to reset the password. Next feature is the profile feature. Users can add and edit their name, email, institution and address. Besides, users can delete their account using the "Delete Account" button in the profile page. Users can check the status of the donated datasets such as approved, pending and rejected in the profile page. The news feed feature of this repository displays the most popular and latest datasets in the homepage. Users can view the top six of the popular and latest datasets. Last feature is the display of paper citing datasets. Under the datasets page, the papers which are citing the datasets will be displayed and users who are interested can click the link provided.

2.6 Study of Google Dataset Search

Google developed a search engine which is Google Dataset Search to allow users to locate the data which is freely for use. It will provide a platform to let users search the datasets and the links of the website (Natasha, 2020). The features offered are login and register, multi-language, feedback, preview of result and saved datasets.

First is the login and register feature. Users can register and login using their Google accounts. Next, Dataset Search supports multi languages such as English, Chinese, Korean and others. Users can change the language and the results will be displayed in that language too. For the feedback feature, users can directly send the feedback message without sending email. Users can attach the screenshot and highlight the problem area. Besides, users can view the preview of each result. Users can check the name of the websites, issue date, creator and the description of the datasets. Lastly, users can bookmark the datasets and view them again under the "Saved datasets" tab.

2.7 Proposed Application: Scientific Data Repository (SDR)

The proposed application in this project is known as the Scientific Data Repository (SDR). In this system, there are new and adapted features implemented compared to other existing systems. First is the registration feature. The registration is open for all users including non-staff users. They can register new accounts using the first name, last name, email address and password. For the login feature, users need to fill in the email address and password to login. Invalid username, email address and password will trigger the error message to remind the users. The password can be changed by the users. Next is the user profile. Each user has their own profile, displaying their information such as name, email address, phone number, faculty and others. Users can upload their social media links on the profile page too. Besides, the data collection feature allows users to bookmark the data they are interested in.

The data bookmarked will be stored at the profile page. Thus users can check the data again on the profile page and no repeated searching is required. Users can always remove the bookmarked data from the collection list. There are five category types offered in the browsing feature of the SDR system. Users can browse the data by the categories which are author, data type, data name, year of publication

and subject. Apart from that, users can perform simple search or advanced search. For simple search, a personalized search engine will be developed in the SDR system. The result of the search displayed will be related to the users' faculty and history search.

While for the advanced search, users can search the data based on the keywords and the system will return accurate results to them. Besides, users can submit the feedback form directly in the SDR system. Users need to fill in their email address and comment in the text box provided. In addition, access control features are offered in this system. Data providers can limit the users' access to their uploaded data by setting the data as public or private. However, a change control feature is developed in which the non-staff users still can request access to the private data. A form will be provided to the users to fill in the details and reason of requesting the access. The data providers can check the request and decide to approve or reject the request.

2.8 Comparison with the Existing Applications

All systems are missing some features that are implemented in other systems. The basic features like login and advanced search are offered by all applications. The proposed application will include the aforementioned features as well. Besides, the features will be modified in the proposed application. For example, the registration for staff users is omitted because the administrator will register the staff user in the database. Next, some features such as data collection, five browse categories types, feedback form and access control are implemented in some of the applications. In the proposed application, all these features will be implemented. In addition, a personalized search engine will be developed to allow the users to get a more accurate result based on their faculty and course. All three existing applications do not have the feature change control. Thus, the proposed application provides the change control feature in which the users can send a form to the data provider to approve the request to view the private data. The other pros and cons of the existing applications are analyzed and documented too. The proposed application will use the result as reference to develop the features in the system.

Table 1: System's Comparison

Features	ORKED	Kaggle	UCI Machine Learning Repository	Google Dataset Search	SDR
Login	✓	✓	. ✓	✓	✓
Registration for New Account	✗	✓	✓	✓	✓
User Profile	✓	✓	✓	✓	✓
Five or More Browse Category Types	✗	✓	✓	✗	✓
Advanced Search	✗	✓	✓	✓	✓
Feedback Form	✗	✗	✗	✓	✓
Access Control	✓	✗	✓	✗	✓
Change Control	✗	✗	✗	✗	✓

3. Methodology

In this project, prototyping model is selected as the methodology. The Prototyping model is one of the types of system development methods. Using this method, a prototype is developed, tested and refined until a satisfied outcome is accomplished. It is suitable for the scenario that the requirements of the system are not known. The Prototyping model is an iterative process that takes place between developers and users. There are six phases in the prototyping model which are Identify and Analyse Requirement phase, Quick Design phase, Construct Prototype phase, User Evaluation phase, Refine Prototype phase and Implement and Maintain System phase.

3.1 Identify and Analyse Requirement Phase

In the Identify and Analyse Requirement phase, to gather the system requirements from users, a meeting with the stakeholder, RMC is carried out. During the meeting, requirements of users and their expected outcome of this system are collected. All the requirements and expected outcome are analysed and documented. Next, the problem statements and objectives are determined, and the project schedule is planned. The output of this phase will be the proposal, Gantt Chart, System Requirement Document, Requirements Traceability Matrix (RTM), use case diagram, activity diagram, class diagram and sequence diagram.

Table 2: Functional Requirements

Modules	Functional Requirements
Login and Registration Module	- The system should allow new users to register account. - The system should allow administrator to register other administrator and UTHM users. - The system should allow users and administrator to login using username and password.
Data Management Module	- The system should allow data uploaders to add, edit and delete the research data. - The system should allow administrator to reject or approve the upload request of UTHM users. - The system should allow administrator to add, edit and delete the research data.
Data Upload Module	- The system should allow UTHM users to upload research data with different formats. - The system should allow data uploader to limit access of non-UTHM users towards their data. - The system should allow data uploader to change the access of others towards their data.
Browse and Search Module	- The system should allow users to browse the research data based on five category types. - The system should allow users to simple search the research data by typing keywords in the search engine. - The system should allow users to advanced search the research data. - The system should provides a personalized search engine for users based on their faculties and course studied. - The system should allow users to add research data interested into collection. - The system should allow non-UTHM users to request access towards the private research data.
Feedback Module	- The system should allow users to send feedback using the feedback form. - The system should allow administrator to view the feedback.
Profile Module	- The system should allow users to edit personal information. - The system should allow users to view the data uploaded in the profile.

Table 3: Non-Functional Requirements

Non-Functional Requirement	Description
Usability	- The system should display help message to allow users learn to use the system. - The system should allow user to reach their goal using a simple and clear interface.
Security	- The system should limit the access of non-UTHM users towards the private research data. - The system should ask for the password when users trying to edit their password. - The system should allow data uploader and administrators only to edit and delete the research data.
Performance	- The system should provide response time which is 5 seconds and below.
Reliability	- The system should allow users to use the system 85% of the time without bugs and failure.
Availability	- The system should allow users to access the system at anytime.

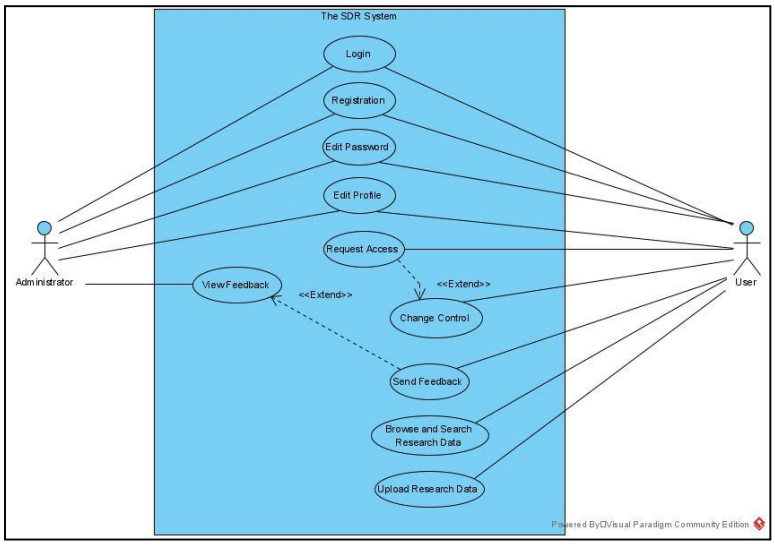

Figure 1: Use Case Diagram of Proposed System

Use case diagram is used to present the behaviour of a system. The requirements of the system and the relationships between requirements and users are clearly shown in the diagram. Figure 4.1 shows the use case diagram of this project. There are 2 actors involved in this use case diagram which are administrator and user.

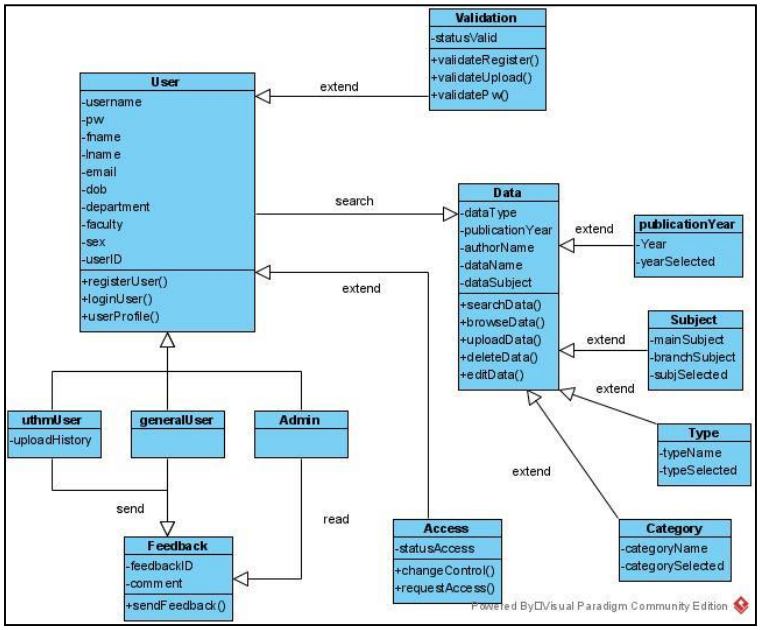

Figure 2: Class Diagram

Class diagram can be known as structural diagram, which is a type of static diagram. Class diagram shows the static view of the system and describe the functionalities offered by the system. The class diagram, including classes, attributes, functions and relationships between classes are shown above.

3.2 Quick Design Phase

In the Quick Design phase, the design of the process, interfaces and database will be carried out and done quickly. The system architecture will be designed in this phase. Besides, the user interfaces and database are designed for the next phase. The output of this phase is the system architecture, user interface design and database design.

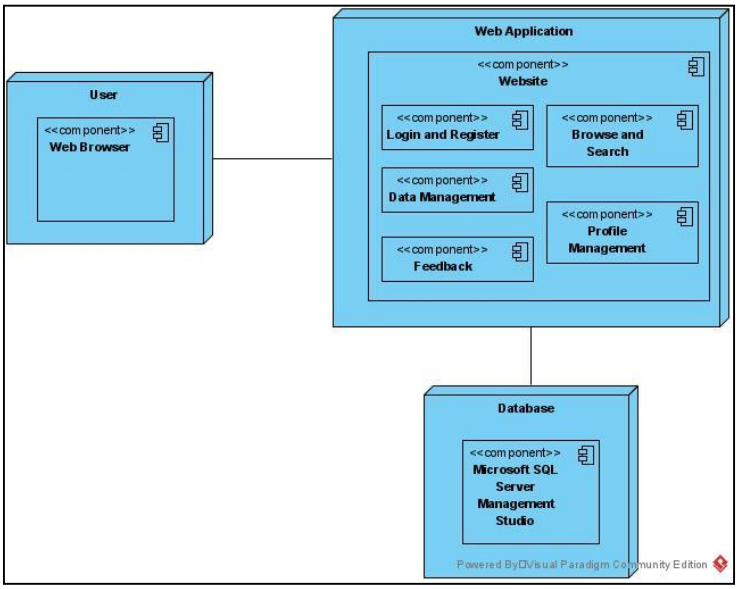

Figure 3: System Architecture of the Proposed System

System Architecture is used to model the structures and behaviour of the system. Figure 3 shows the system architecture of the SDR system.

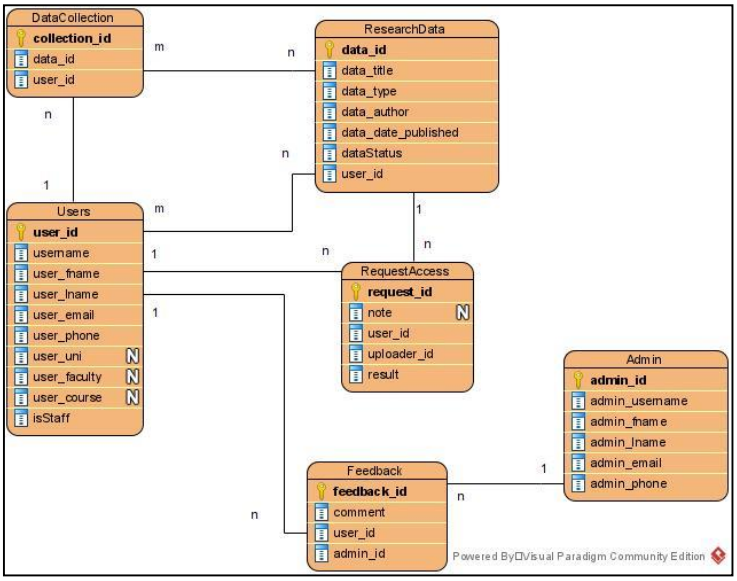

Figure 4: Entity Relationship Diagram

A database management system is designed, implemented and maintained in the phase of database design. A good database design is essential to ensure the accuracy and consistency of the data. The Entity Relationship Diagram (ERD) that shows the relationship between table in the database are shown above.

Interface design is the process that design and develop the looks and styles of the system. The goal of interface design is to create a user-friendly interface to allow users use it easily. The interface design of each module in the SDR system will be shown.

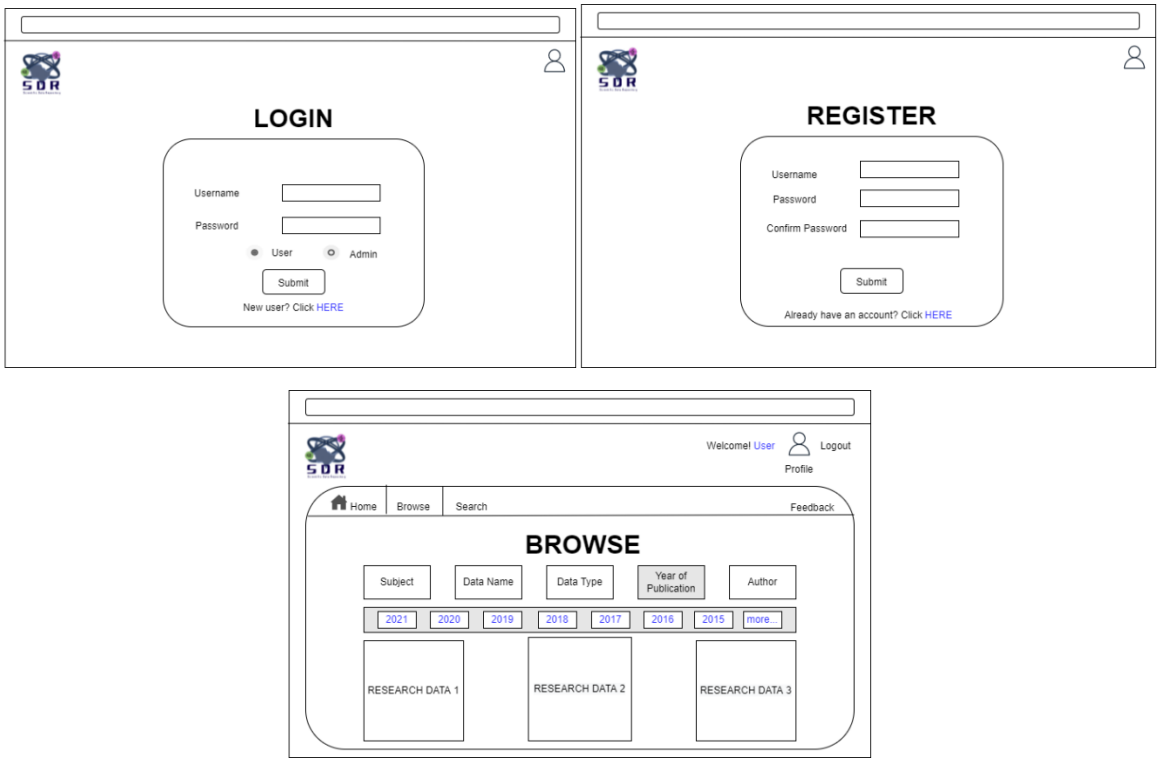

Figure 5: Interface Design

3.3 Construct Prototype Phase

In the Construct Prototype phase, the initial prototype of Scientific Data Repository is developed, and only basic requirements and interfaces are implemented. The user interfaces and database design in the previous phase will be implemented in the initial prototype. The output of this phase is the programming code and initial prototype.

3.4 User Evaluation Phase

In the User Evaluation phase, the initial prototype is firstly tested by the users to discover errors. This phase is divided into component testing, modules testing, sub-system testing, system testing, and the acceptance testing. The testing will be performed on the prototype and all the results from the testing will be collected. The output of this phase is the feedback from users.

3.5 Refine Prototype Phase

In the Refine Prototype phase, the prototype is improved based on the feedback from users in the previous phase. The prototype is refined repeatedly until all the requirements of users are fulfilled. The User Evaluation phase and Refine Prototype phase will be carried out until a final prototype is developed. The output of this phase is the final prototype.

3.6 Implement and Maintain System Phase

In the Implement and Maintain System phase, the system is ready to implement after the final prototype is developed. The final system will be documented. After the system is deployed, maintenance of the system is carried out from time to time to ensure the system is updated. The output of this phase is the final report and complete system.

4. Results and Discussion

4.1 Implementation Result

In this part, a few examples of modules will be shown. The Scientific Data Repository is developed using programming language PHP and HTML, while the database used is MySql. The coding for this project is created during the implementation phase. Figure 6 shows some examples of the interfaces.

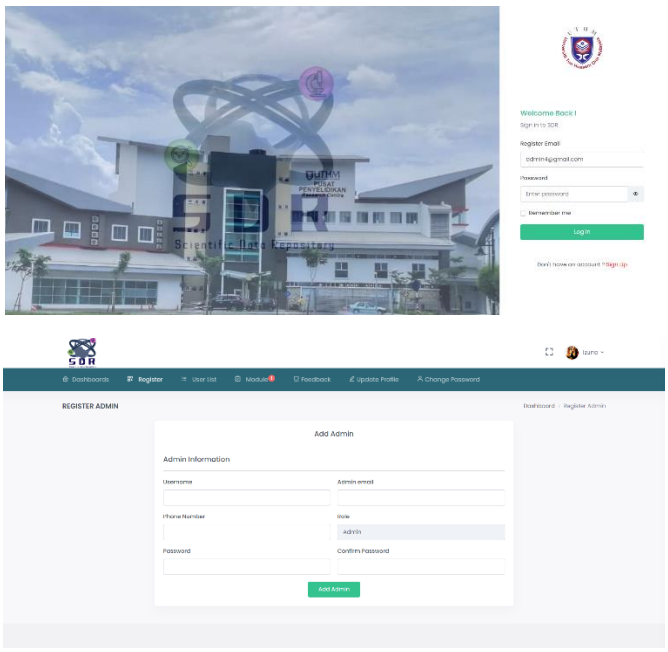

Figure 6: Example of Implemented Interfaces

4.2 Functional Testing Result

Before the system is released, testing is important to make sure that all the functionalities of the system can be carried out. In this part, a testing will be implemented on the function of each module developed in the SDR. There is total six modules in this system which are login and registration, data management, data upload, browse and search, feedback and profile. Table 4 shows the testing result for each module functionality.

Table 4: Testing Result of System Modules

Modules	Test Case	Expected Output	Actual Output
Login and Registration Module	Display all textbox and button.	All textbox and button are displayed.	Success
	Display error message when registration is failed.	Error message is displayed.	Success
	Display success message when user successfully register.	Success message is displayed.	Success
	Users leave blank input during registration.	Registration is failed.	Success
	Users enter valid input during registration.	Registration is success.	Success
	Display error message when wrong input during login.	Error message is displayed.	Success
	Display success message when login is successful.	Success message is displayed.	Success
	Users enter wrong input during login.	Login is failed.	Success
	Users enter correct input during login.	Login is success.	Success
Data Management Module	Display all data uploaded by UTHM users.	All data uploaded is displayed.	Success
	Approve data uploaded.	Data is displayed and status show approved.	Success
	Reject data uploaded.	Data is not displayed, and status show rejected.	Success
	User edit data uploaded.	Data is edited.	Success
	User delete data uploaded.	Data is deleted.	Success
Data Upload Module	Display all textbox and button.	All textbox and button are displayed.	Success
	Form inputs are not complete.	Data upload failed.	Success
	Form inputs are complete and valid.	Data upload success.	Success
	Data uploaded is displayed to administrator for approval.	Data approval is displayed at administrator page.	Success
	Display approved or rejected status of data.	Status is displayed in user profile.	Success
	Users upload data as private.	Data cannot access by non-UTHM users.	Success
	Users upload data as public.	Data can access by non-UTHM users.	Success
	Users accept change control request.	Private data can access by requester.	Success
Browse and Search Module	Display all data uploaded.	All data uploaded is displayed.	Success
	Display all sub menus.	All sub menus are displayed.	Success
	Display data based on categories selected.	Related data based on the category are displayed.	Success
	Users download the data.	Data is downloaded.	Success
	Non-UTHM users request access to private data.	Uploader will receive request.	Success
	Users enter keyword in search engine.	Related data is returned.	Success
Feedback Module	Display all textbox and button.	All textbox and button are displayed.	Success

Modules	Test Case	Expected Output	Actual Output
	Users leave blank input in feedback form.	Feedback submission failed.	Success
	Users complete the feedback form.		
	Display feedback form submitted to administrator.	Feedback submission success.	Success
		Administrator can view the feedback.	Success
Profile Module	Display user information.	All user information is displayed.	Success
	Users edit profile information.	Users is redirected to edit page.	Success
		Edit failed.	
	Invalid input during edit profile.	Edit success.	Success
	Valid input during edit profile.	Users is redirected to change	Success
	Users change password.	password page.	Success
		Status approved or rejected will	
	Display status of data uploaded.	be displayed.	Success

4.3 User Acceptance Testing

After the functional testing is conducted, the system is next implemented with the user acceptance testing. The testing is conducted by sharing the website link to a group of users and their review is collected using Google Form. There are 10 respondents selected to participate in the user acceptance testing. They will review the system based on functional and non-functional requirements. The result is converted to a chart shown in Figure 7 and Figure 8.

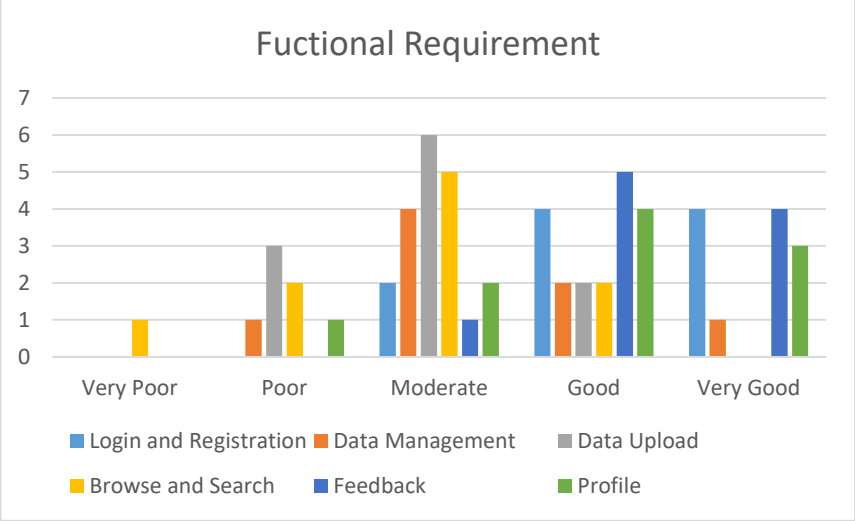

Figure 7: User Acceptance Testing Result of Functional Requirements.

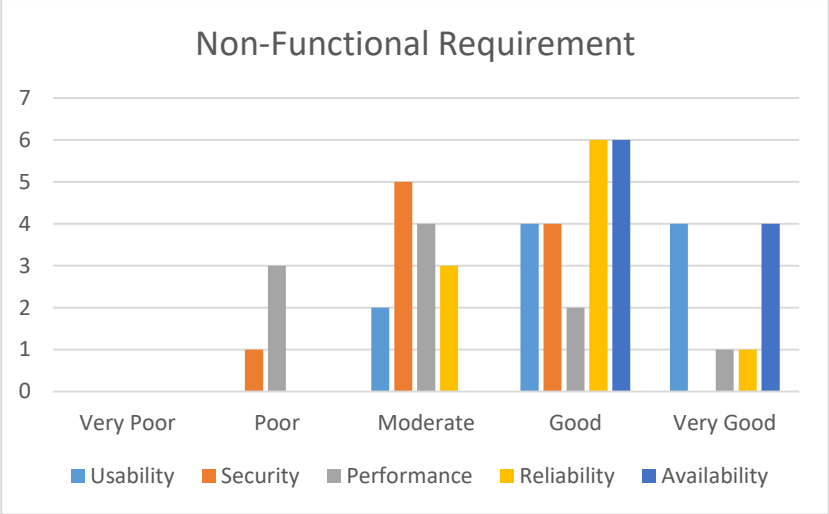

Figure 8: User Acceptance Testing Result of Non-Functional Requirements.

4.4 Discussion

In conclusion, the implementation of the system is successful since the functionalities can be worked properly. However, there is room for improvement of the system since some ratings are below "good" rating. For example, the Browse and Search module should be more efficient to return more accurate result to users. Performance of the system should be improved by make sure the loading time not taking too long to response.

5. Conclusion

At the end of this project, a website for this scientific data repository was successfully developed for UTHM to manage and store the research data. Users can upload or download the data sets from this repository.

Acknowledgment

First of all, I would like to thank my supervisor, Dr. Salama A Mostafa, for all the comments and guide to complete the Final Year Project from Semester 1 till the end. Due to guidance from Dr. Salama, I was able to finish the project and report on time. During the period, I learnt a lot of knowledge and skill that are very useful in the future. Next, I would like to thank all my friends who helped and encouraged me in this project. Finally, I would like to thank to my family who always supported me to complete this project.

References

[1] Ikrami M (2018). The Significance of Scientific Research in the Professional Development of Students. Retrieved on September 30, 2021, from doi: 10.1007/978-3-319-73093-6_18

[2] Çaparlar CÖ, Dönmez A (2016). What is Scientific Research and How Can it be Done?. Turk J Anaesthesiol Reanim. 2016;44(4):212-218. Retrieved on September 30, 2021, from doi:10.5152/TJAR.2016.34711

[3] Chigwada, J., Chiparausha, B. and Kasiroori, J. (2017). Research Data Management in Research Institutions in Zimbabwe. Data Science Journal, 16, p.31. Retrieved on September 30, 2021, from doi: 10.5334/dsj-2017-031

[4] Trisovic A., Mika K., Boyd C., Feger S, & Crosas M. (2021). Repository Approaches to Improving the Quality of Shared Data and Code. Data, 6(2), p.15. Retrieved on October 25, 2021, from doi: 10.3390/data6020015

[5] Posavec, K., Celjak, D., & Musap, L. J. (2020). Role of a Croatian National Repository Infrastructure in Promotion and Support of Research Data Management. Data Science Journal, 19(1), p.48. Retrieved on October 25,2021, from doi: 10.5334/dsj-2020-048

[6] Dissanayake, N. & Dias, K. (2017). Web-based Applications: Extending the General Perspective of the Service of Web [Paper Presentation]. Conference: 10th International Research Conference of KDU (KDU-IRC 2017) on Changing Dynamics in the Global Environment: Challenges and Opportunities. Rathmalana, Sri Lanka.

[7] Rashid, Z., Kadiman, S., Zulkifli, Z., Selamat, J., & Hashim, M. Review of Web-Based Learning in TVET: History, Advantages and Disadvantages. International Journal of Vocational Education and Training Research. Vol. 2, No. 2, 2016, pp. 7-17. Retrieved on October 25, 2021, from doi: 10.11648/j.ijvetr.20160202.11