NCF Repo-C-tory: A Research Repository for Naga College Foundation, Inc.

A Thesis/Capstone Project

Presented to the Faculty of College of

Computer Studies

Naga College Foundation, Inc.

Naga City

In partial fulfillment of

the Requirement for the Degree

Bachelor of Science in Information Systems and

Bachelor of Science in Computer Science

Arnel E. Almario Jr.

John Rey S. Tolosa

November 2023

I. INTRODUCTION

The importance of a centralized repository for research articles cannot be stressed in the fast developing field of academic studies. This research project is devoted to the thorough exploration and refining of such a repository—a virtual nexus meant to house the collective expertise of many scientific activities. As our global knowledge base grows, the demand for a uniform platform becomes more pressing. The purpose of this project is to delve into the complexities of repository design, functioning, and accessibility, with the overriding goal of creating an optimized and inclusive place for all research articles. By studying repository dynamics, we want to uncover ways that improve not only the preservation and retrieval of research works, but also the collaborative potential inherent in a repository built to foster collaboration.

With more research that is added every year of every school calendar, there is no doubt the research fell in a file or stack of hardbound that resides in the library. Unfortunately, this research should not only settle on the shelves. Making them electronically available as references, or to be cited are the ones it truly deserves. This paper emphasizes the need of a cloud-based research repository to be implemented in every university that can be utilized to serve its purpose. This research repository is based on an online publication and subscription model. Online publication provides reading sources via the internet which is accessible and more convenient to most people. The repository will also adapt the concept of configurability as

the users may have their own preferences with regards to how they publish or subscribe a paper. Research that is within the repository that will be referenced, cited, or downloaded has the corresponding remuneration based on the approval of the School. In this way, more researchers will continue to provide more scholarly output to be published and to gain more citing, downloads, and eventually more remuneration. The repository has the potential to expand as more researchers will be turned in its service and would be beneficial to stakeholders. The respondents in this paper show the acceptability of the process, making them more likely to work in any educational institution.(The Development of a Cloud-based University Research Repository Software Using a Configurable Subscription Model, 2023).

According to Okon, R., Eleberi, E.L. and Uka, K.K. (2020) ,Institutional repositories are essential research infrastructure for research-based universities. A properly dimensioned institutional repository has the potential to increase research impact and enhance the visibility of an institution through its scholarly outputs.

The advances of technology resulting in the proliferation of mobile devices have changed the way we live and have necessitated the restructuring of the educational system. This can be employed to aid student's participation in research studies(Development of a Research Project Repository, n.d.).

Scientific data repositories are essential infrastructure for promoting open research data and enabling data sharing. They serve as a bridge between policies and researchers. Jiang et al. (2023) says that Building a trustworthy and

FAIR-compliant repository is a challenge. Thus guidelines on repository selection and requirements for trustworthy repositories have been proposed. The paper analyzes Repositories registered on the re3data website and summarizes their current development. It also focuses on featured practices in domain-specific and generalist repositories.

A study about the Current Developments in the Research Data Repository RADAR (Bach et al., 2022) is a cross-disciplinary research data repository that focuses on archiving and publishing digital research data from disciplines without specific research data management infrastructures. It ensures long-term availability of datasets according to FAIR criteria and promotes the implementation of the FAIR Principles. RADAR follows a modular system architecture and allows integration with existing systems through APIs. It uses a two-step process for data transfer and archiving. RADAR participates in the National Research Data Infrastructure and plays a relevant role as a generic infrastructure service in several NFDI consortia. The paper highlights the need for further development in terms of interoperability and discusses the potential of the FAIR Digital Object Framework to enhance data interoperability. RADAR aims to enhance its services based on the needs of different research communities, such as integrating subject-specific terminologies and providing annotation options with subject-specific metadata.

Synthesis of the State-of-the-Art

The traditional model of research dissemination through physical libraries faces limitations in accessibility and reach. Centralized research repositories offer a compelling alternative, providing online access to research articles, promoting knowledge exchange, and fostering collaboration across institutions and disciplines. This paper examines the current landscape of centralized research repositories, highlighting key trends, challenges, and promising future directions. With the emergence of open access initiatives is challenging the traditional model of research dissemination through journals that require payment. Repositories play a crucial role in enabling open access by providing free and unrestricted access to research outputs. This has the potential to accelerate scientific progress and foster collaboration on a global scale. The rise of collaborative research is evident in modern-day studies, which increasingly involve collaborations between multiple institutions and across international borders. Repositories serve as facilitators for seamless sharing and co-ownership of research data, supporting collaborative research efforts and promoting interdisciplinary studies.

Repositories are implementing robust search functionalities and adhering to metadata standards in order to ensure that research data is easily discoverable and accessible to researchers, regardless of their location or institutional affiliation. Additionally, repositories are adopting standardized data formats and application programming interfaces (APIs) to enable smooth data exchange between different platforms and promote data reuse in new research projects. To ensure the long-term preservation of research data and metadata, repositories employ strong data storage and backup mechanisms, making them accessible for future

generations of researchers. To cater to the specific needs and access preferences of researchers, repositories are offering various subscription options. These options may include tiered access based on usage, author affiliations, or research disciplines. Repositories are also incorporating user-centric features, such as personalized dashboards, customizable search filters, and annotation tools, to enhance the user experience and improve research workflows. Some repositories are even experimenting with reward systems that are based on citations, downloads, and other metrics. These reward systems aim to incentivize researchers to contribute their work and actively engage with the platform. By doing so, high-quality research can be encouraged and the visibility and impact of research outputs can be increased. In addition to traditional subscription fees, repositories are exploring alternative funding models, such as grants, partnerships with research institutions, and in-kind contributions, to ensure long-term sustainability and avoid financial barriers for researchers while Cloud-based repositories offer scalability to accommodate the ever-growing volume of research data and ensure flexible access for researchers from anywhere with an internet connection. Compared to on-premise solutions, cloud-based repositories can be more cost-effective for institutions as they eliminate the need for expensive hardware and infrastructure maintenance. Furthermore, cloud platforms provide advanced security features and comply with data privacy regulations, ensuring the safe and secure storage of sensitive research data.

Centralized research repositories offer a transformative approach to research dissemination and collaboration. By addressing the identified challenges and embracing emerging opportunities, **NCF Archives** can play a critical role in

accelerating scientific progress and fostering a more open and collaborative research ecosystem

Gap Bridge of the Study

This study would also bridge the gap between the portability of the thesis/documents and the students' ease of accessibility. As many research papers often lack basis and related literatures, the proposed repository would resolve through a centralized and standardized system. Also, academic institutions' prevalent issues such as effective organization and searchability would be resolved, bridging the gap of the research body from the knowledge sources.

Theoretical Framework

The proposed system is anchored into one algorithm: Fuzzy Search/Matching Algorithm for the content searching algorithm. This algorithm returns a list of results based on likely relevance even though search argument words and spellings may not be an exact match. It consists of several algorithms that work together to produce accurate and fast results. Examples are N-gram, and Levenshtein distance. Overall, the algorithm browses datasets looking for strings matched to the user query and marks each with relevance score. This relevance score would determine the sorting/ranking of results. However, the one that would be used in the study is a hybrid approach and compiled into one library called FuzzyWuzzy. This type of algorithm would be suitable for one of the key features of NCF Archives, which is

the full-text typographical error-safe search engine. It also displays any suggested relevant results based on the user's query.

Conceptual Framework

The Input-Process-Output (IPO) Framework, added with Feedback as the fourth component, is the approach utilized in the conceptualization of this study.

Input:

a) **Research Data:** Raw data collected from various research projects conducted by students, faculty, and researchers.

b.) **Metadata:** Information about the research data, including author details, date of publication, keywords, etc.

c.) **External Data**: Relevant external data sources, literature, or datasets that contribute to the research repository.

Process:

a.) **Data Ingestion and Storage:** Import and store research data in a standardized format within the repository.

b.) **Metadata Tagging:** Assign metadata tags to classify and categorize research data for efficient retrieval.

c.) **Access Control:** Implement security measures to control access to the repository based on roles and permissions.

d.) **Search and Retrieval:** Develop a robust search functionality for users to easily find and retrieve relevant research data.

e.) **Analytics and Reporting:** Implement tools for analyzing usage patterns, trends, and generating reports for stakeholders.

Output. The study's outcome is the NCF Repo-C-tory: A Research Repository for Naga College Foundation, a digital platform where all the thesis or capstone projects can be restored and retrieved.

Feedback. A feedback mechanism was added to the Input-Process-Output (IPO) framework. The process of modifying software functionalities and features will be based on feedback, demand, and criticism from the users.

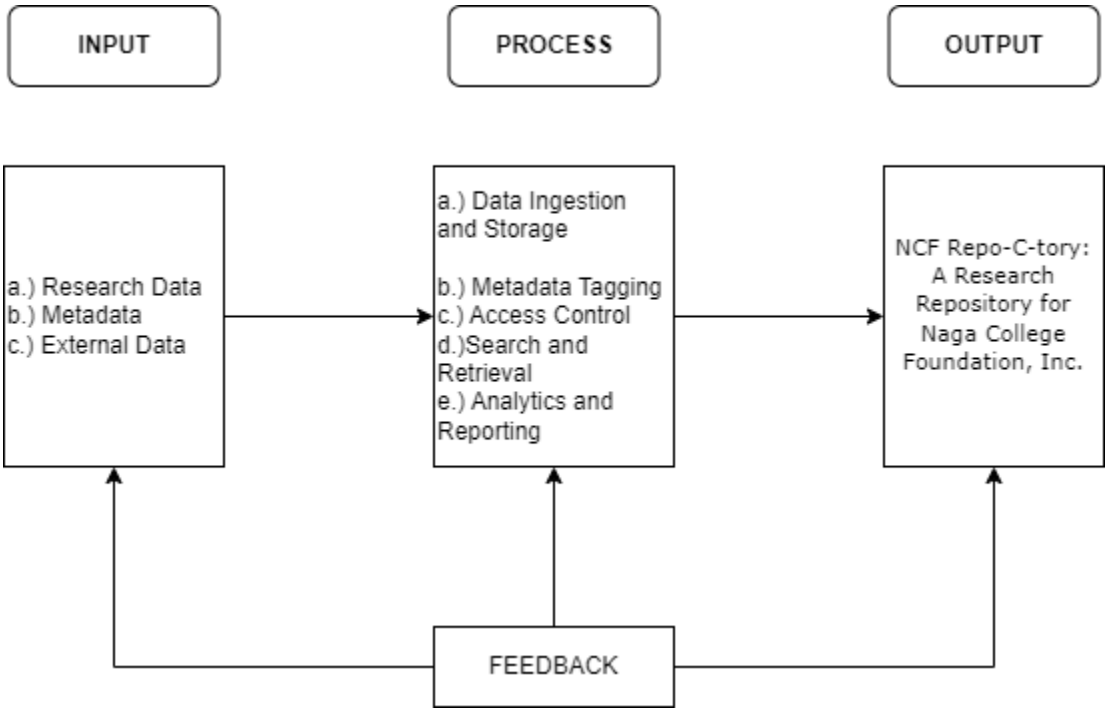

Figure 2 Conceptual Paradigm

Objectives of the Study

The main objective of the study is to design and develop the NCF Archives: A Research Repository for Naga College Foundation, as well as to determine if a repository will be helpful in storing and retrieving research papers.

1. To determine the current state of the system used in the storage, management, and retrieval of research documents in the research center of Naga College Foundation, Inc.
2. To design and develop system modules for the storage, management, and retrieval of the research documents in the research center of Naga College Foundation, Inc.
3. To evaluate the level of post-study system usability utilizing the International Business Machines (IBM) - Post-Study System Usability Questionnaire (PSSUQ).

Assumptions

The study was anchored on the following assumptions:

1. Faculty, researchers, students, and other stakeholders at Naga College Foundation will actively engage with NCF Archives, contributing to the repository and utilizing its resources.

2. Relevant and accurate research data, including papers, theses, and scholarly works, will be consistently available for inclusion in NCF Archives.

3. The repository's user interface and design are intuitive and user-friendly, minimizing the learning curve for faculty, researchers, and students.

4. Users of NCF Archives will find the repository's features and functionalities satisfactory, contributing to a positive user experience.

5. NCF Archives is designed and implemented with a long-term vision, ensuring its sustainability, relevance, and adaptability to future technological advancements.

Scope and Delimitations of the Study

This research and development project shall store the digital formats of a wide range of research papers, theses, dissertations, and scholarly works produced by faculty, researchers, and students affiliated with Naga College Foundation. The study shall cover designing user-friendly interfaces and advanced search functionalities to ensure proper accessibility for faculty, researchers, students, and other stakeholders. It shall also include development of features that shall facilitate alumni engagement, enabling former members of the academic community to access and contribute to the repository, subject for critical evaluation and approval of the research committee, prior to inclusion into the knowledge base pool. The scope includes an administrative dashboard with tools for efficient content curation, updates, and overall management, empowering librarians and administrative staff.

This research and development project shall not include non-academic materials, ensuring a focus on scholarly works and research outputs relevant to the academic mission of Naga College Foundation. It shall also exclude non-digital formats (e.g., scanned documents) as the search algorithm greatly depends on the text content of the document. And the repository's primary focus is on the digital copy of the research documents submitted by individuals affiliated with Naga College Foundation, and it will not encompass research from external institutions unless in collaboration with NCF. For the first development phase of the system in compliance with the Thesis 1/Capstone 1 subject, the scope would only be limited to the College of Computer Studies department of Naga College Foundation.

Significance of the Study

The NCF Archives: A Research Repository for Naga College Foundation will be beneficial to the following:

Faculty and Researchers. Faculty members and researchers at Naga College Foundation can benefit from a centralized repository by having easy access to a wealth of research materials. This facilitates their scholarly pursuits, curriculum development, and collaborative research initiatives.

Students. Students, both undergraduate and postgraduate, can access a comprehensive repository of research papers, theses, and other academic

resources. This aids in their academic endeavors, research projects, and enhances the learning experience.

Administrative Staff. Administrative staff involved in academic planning, accreditation processes, and institutional research can benefit from a well-organized repository that provides quick access to historical and current academic data.

Librarians. Librarians responsible for managing and curating academic resources will find the repository valuable for efficient cataloging, archiving, and retrieval of research materials.

Alumni. Alumni seeking to stay connected with the academic output of Naga College Foundation can benefit from a repository that preserves and showcases the institution's research legacy.

Future Researchers. The repository contributes to the legacy of the institution by preserving and making accessible a wealth of academic knowledge for future researchers.

Definition of Terms

The following terminologies were defined to provide clarity and understanding throughout the study, both conceptually and operationally.

Repository. A repository is a centralized location where data, information, or resources are stored, organized, and managed. It serves as a container or database that holds a collection of items, and it is designed to facilitate easy access, retrieval, and management of those items.

Configurability. Refers to the ability of a system or a software application to be easily customized or adapted to meet specific requirements or preferences without requiring significant changes to its underlying code or structure. A configurable system allows users or administrators to modify certain settings or parameters, tailoring the system's behavior to suit their needs.

Microservice. Microservices are an architectural and organizational approach to software development where software is composed of small independent services that communicate over well-defined APIs. These services are owned by small, self-contained teams.

API. API is an abbreviation for Application Programming Interface which is a collection of communication protocols and subroutines used by various programs to communicate between them. A programmer can make use of various API tools to make their program easier and simpler. Also, an API facilitates programmers with an efficient way to develop their software programs.

I. METHODOLOGY

This methodology provides a structured and iterative approach to the development, implementation, and maintenance of NCF Archives, ensuring that it aligns with the needs of the Naga College Foundation and its user community.

Methods of Research

The Research Methodology approach will utilize a mixed-method approach .Research Repository for Naga College Foundation depends on the specific goals, objectives, and research questions of the study. Research methodology that combines qualitative and quantitative approaches to provide a comprehensive understanding of the repository's effectiveness, usability, and impact.

Respondents of the Study

The study will be conducted at the Naga College Foundation, Inc. The respondents of the study will be:

1. **Students.** Both undergraduate and postgraduate students can provide perspectives on how the repository enhances their access to academic resources, aids in research projects, and contributes to their overall learning experience.
2. **Researchers.** Individuals actively involved in research activities at Naga College Foundation can offer valuable feedback on the repository's effectiveness in supporting their research endeavors.
3. **Faculty Members.** Uploading research papers, utilizing the repository for curriculum development, and engaging in collaborative research.

Data Gathering Procedure

The methods and instruments the researcher used to collect data are listed in this section.

Surveys. It is a process for gathering information from the person or people being interviewed by asking questions and providing prompts.

Research Tool

Survey Questionnaires: The researcher can obtain vital data and specifications required for system design and development with this type of guided-response questionnaire. Questionnaires regarding the proposed system and the current system will be provided.

Software Evaluation Tool. The researcher can assess user feedback on the system and satisfaction with this tool. We'll evaluate the system's functionality, usefulness, durability, and portability.

Statistical Tool

The following tools and techniques will be employed by the researcher to examine and clarify the information gathered from study participants.

Descriptive Statistics. It will give the researcher the measurements of variability and averages in addition to a summary of the data.

Inferential Statistics. It allows you to extrapolate predictions and inferences about a larger population from a sample of data collected from that community.

Systems Design

The NCF Archives aims to provide a centralized platform for storing and accessing the finished thesis and dissertations of the students. Hence, the proposed system is designed to handle large numbers of concurrent users while ensuring efficient search and retrieval of the documents, to achieve those aims, the system would comprise 3 components; interface layer, application layer, and data management layer.

The interface layer is designed to have a simple and user-friendly interface, ensuring an efficient and seamless user flow experience in searching, accessing and downloading documents. The search query could require varying choices such as document name, author, and other metadata.

The application layer is designed to achieve an efficient data retrieval and handling concurrency. This layer would house the APIs, search engine, and other microservices that would process the system core functionalities.

The data management layer would be responsible for the data storage and process any retrieval request sent from the other layers. This layer consists of the DBMS or the Database Management System that would store all the system information, such as user credentials, documents, and etc.

The system is also designed to offer several key features that are aimed to enhance efficiency and user experience, such as; a full-text typographic error-safe search engine, faceted or filtered search, download and access control, document upload and management, version control, and data analytics.

For scalability and performance, the system is designed to implement asynchronous processing, and caching added with horizontal scaling powered by its backend architecture.

Systems Architecture

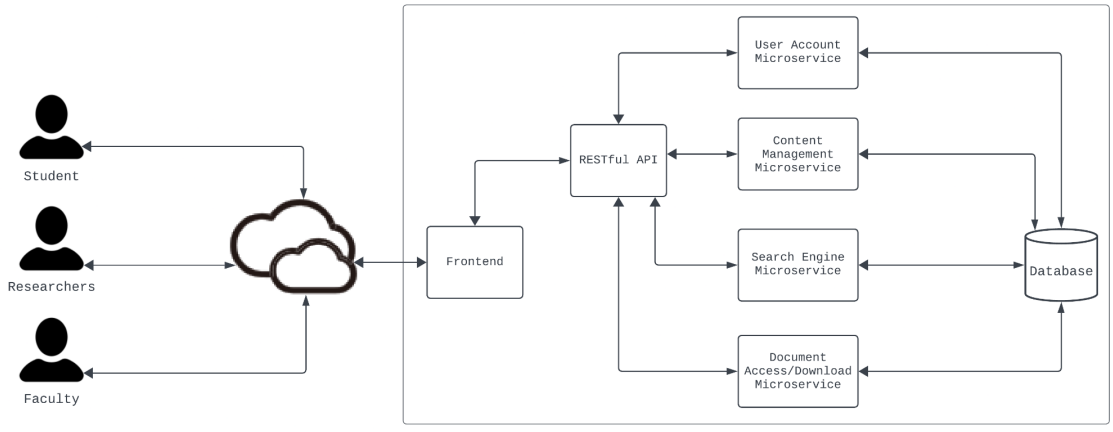

Figure 1.1. System Architecture

The proposed system for NCF Archives encompasses a hybrid approach by combining a Client-Server and Microservice Architecture wherein the core functionalities are subdivided into several microservices, such as User Account, Search Engine, and Document Access/Download Microservices.

User Account Microservice: This microservice houses the account creation, and management features of the system. It would be responsible for processing of user credentials to store and retrieve data from the database.

Content Management Microservice: This microservice would be responsible for document upload, store, update, archive, and report. The results processed by these microservice would be displayed back to the frontend for the user's view.

Search Engine Microservice: This microservice would be responsible for processing the search query of the user. It would be the one to implement the searching algorithm, and results filtering. The results processed by these microservice would be displayed back to the frontend for the user's view.

Document Access/Download Microservice: This microservice would be responsible for the content access control and download control that would limit content access and download via authorization, preventing unwanted user access on certain documents.

These microservices are designed to deliver NCF Repo-C-tory' core functionalities efficiently and effectively. Meanwhile, these microservices would deliver the data to the frontend using the RESTful API.

The proposed system is expected to be developed using Flutter for the frontend, Node JS for the backend, and PostgreSQL for the DBMS(Database Management System).

End Notes

1. A Web Based Digital Repository for Scholarly Publication. (2020). A Web Based Digital Repository for Scholarly Publication. 13(4), 67–75.
2. Okon, R., Eleberi, E.L. and Uka, K.K. (2020) A Web Based Digital Repository for Scholarly Publica-tion. Journal of Software Engineering and Applications, 13, 67-75
3. Development of a research project repository. (n.d.). Development of a research project repository. 18(1), 156–165.
4. Jiang, L., Zhang, Z., Li, Z., Gai, Z., Wang, P., Li, C., & Zhou, Y. (2023). The construction practice and prospects of global scientific data repository platform. *China Scientific Data*, *8*(1). https://doi.org/10.11922/11-6035.csd.2023.0027.zh
5. Bach, F., Soltau, K., Göller, S., & Minella, C. B. (2022). Current developments in the Research Data Repository RADAR. Research Ideas and Outcomes, 8. https://doi.org/10.3897/rio.8.e96005