COMET LAB

Project Documentation Standardization with GitHub

DevOps Project – Fall 2023

01/01/2024

Produced by:

Hasaranga Jayathilake

Supraja Pericherla

Anuja Sawant

Prakashini Reddy

Contents

Table of Figures and Tables

I. Introduction

Since the number of projects that are being managed by the COMET lab has been increased substantially. The importance of having a capability of tacking the progress of each project is much more important to ensure on sustaining the quality of each project output.

Therefore, identified that having a standardized project documentation may,

- Enrich the quality of the project output,
- Simplify the task that needed to be accomplished to complete the project,
- Smooth process of monitorization of project progress,
- Reduce the difficulty of transition time for handing over the responsibilities to the new team members and
- Enhance the understandability of the project progress for the internal and external project stakeholders.

A. Project Overview:

The purpose of this project is to create a standardized GitHub repository that can be utilized to retain all the documents of the projects that have been carried out by the COMET lab.

The scope of the project covers on how to create a GitHub profile for the COMET Lab, how to set up profiles for each student who works under the lab, how to organize project groups, and how to create a sample standardized GitHub repository with relevant sample files and an instructional webpage under the COMET lab website.

B. Problem Statement:

One of the main reasons for starting this project was the lack of a common repository for the lab to store and showcase its work.

Before this project, each project leader had their own personal repository to do the lab work and share it with other members of the project team. This was a major problem that needed to be solved.

C. Stakeholders and Key Participants:

The key participants and internal stakeholders of this project are all the members of the Lab. The project owners, who are the external parties that own the projects that the Lab students work on, are the main external stakeholders of the project.

Additionally, current and former Lab students, as well as potential employers of current and future Lab alumni, are considered as second-tier external stakeholders of this project.

D. Deliverables and Expected Outcomes:

Table 1: Projects' deliverables and their expected outcomes

Deliverables	Expected Outcomes
COMET Lab GitHub Profile	To achieve the all the labs' project data
Standardized Sample Repository	Expected structure of project repository with standardized sample files on reporting
Instructional Web Page	Guide on repository usage for the lab members (Guideline and training purpose)

E. Dataset Details:

Not Applicable

Initially, conducted face-to-face interviews with a few lab members to identify the project problem.

II. Methodology and Approach

To identify the issues that are associated with the lab members, As a team, did the following:

- Conducted a job profiling of the lab student member to understand the issues that are relevant to their workflow (*please go through the appendixes.*)
- Conducted focus group interviews and brainstorming sessions with some veteran students who had worked in the lab for at least 2 semesters prior to find out the most important issues that needed more priority at the beginning.
- Settled on the problem that has been taken for this project to be completed.

Table 2: Content of the Appendixes in this document

Appendix #	Content Description
Appendix 1	Life Cycle of a Graduate Research Assistant in COMET Lab.
Appendix 2	Graduate Research Assistance Timeline and Major Events
Appendix 3	Job Role Portfolio: Graduate Student Research Assistant - Data Science
Appendix 4	Job Role Portfolio: Project Manager - Data Science

III. Technical Details

Technology Stack:

Table 3: Projects' technological stack based on the tools and technologies used.

Deliverables	Tool	Technologies
COMET Lab GitHub Profile	GitHub Enterprise version	Git based repository architectural pattern
Standardized Sample Repository	Sample text-based documents – MS word	Word Processing
	Sample presentations – MS PowerPoint	Presentation Software
	Sample python code related files – Visual Studio Code	Integrated Development Environment (IDE) with python programming Language
	Sample CMD notes – Visual Studio Code	Integrated Development Environment (IDE) with markup language
Instructional Web Page	WordPress (Content Management System)	Classic theme with HTML and Java

B. Infrastructure Requirements:

Basic internet connection is required to login into GitHub to access the repository.

IV. Conclusion

A. Summary:

- Finalized all three deliverables on November 14, 2023, and made them available for lab members to use for their projects due at the end of the fall semester 2023.

Table 4: Project outcome summary

Deliverables	Outcomes
COMET Lab GitHub Profile	https://github.iu.edu/orgs/polis-comet-lab
Standardized Sample Repository	https://github.iu.edu/polis-comet-lab/template
Instructional Web Page	https://comet.luddy.iupui.edu/cmt23s/training/standardize-project-repositories-and-presentation/

- Presented, demonstrated, and helped on how to use the repository to complete the reports of the members' projects.

As at end of the semester 2023, as a polit project, under the new template 17 data science related project documentation has been completed.

Access to view all repositories: https://github.iu.edu/orgs/polis-comet-lab/repositories

Example of the completed project work based on the standardized repository: https://github.iu.edu/polis-comet-lab/NHANES

B. Challenges:

It has been identified that a main concern is the difficulty of getting on board with this mechanism for projects under the domain of user experience and user interaction, since not much exposure to GitHub is possessed by most of the members in that domain and having to do double work by updating their work on GitHub as well as on Figma is their main concern. Therefore, for using this mechanism for all the work of the COMET Lab inside one platform in the future, there is a concern that needs to be overcome by them.

C. Learnings:

A standardized operating procedure with a standardized repository for the Data Science project is beneficial for the students who work on data science projects. It helps them to have a good understanding of what they need to do in terms of completion of the project. It also helps them to improve their focus more on their project than finding how to present or document their work to be concluded.

D. Next Steps and Milestones (optional):

Key future steps that have been identified are as follows, which can be worked on in the spring semester of 2024:

- Creating instructions and advocating for the use of project tabs for the work assignments process among each other in each project, using GitHub functions like below.
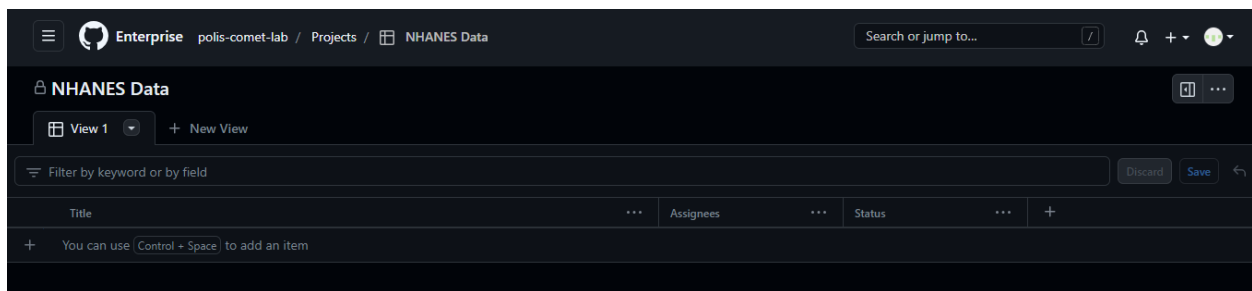

Figure 1: Sample window of work assignment among the team members.

- Finding a mechanism that can ensure that lab members who are doing projects on UI/UX domain use this standardized repository mechanism.

E. Closing Remarks:

This project aims to reduce or eliminate the unnecessary time spent by lab members on formatting the structures they need to create for their assigned projects. It also aims to enhance the collaboration and transparency of each member's work, and to facilitate the review and monitoring of the project from the supervisory perspective.

Most importantly, as a team, we want to acknowledge the support and guidance that we have received from the Polis Center - James Parker Dowling, Sagar Varma Samanthapudi and COMET Lab - Prof. Robert Skipworth Comer to successfully conduct this project.

V. Appendix

Appendix 1

Life Cycle of a Graduate Research Assistant in COMET Lab.

1. Recruitment and Onboarding:

- Graduate students are recruited based on their academic background, research interests, and relevant skills.

- Onboarding includes orientation to the research lab projects, introduction to ongoing projects, and familiarization with lab protocols.

2. Project Assignment:

- Students are assigned to specific research projects based on their expertise and interests.

- The project assignment includes defining roles, responsibilities, and expected contributions.

3. Training and Skill Development:

- Workshops and seminars are provided to support continuous learning and professional development.

- Graduate Research Assistants will have training to enhance their technical skills if they want to improve their knowledge on the GitHub repository or any other important tools that want to use during their project works, including programming languages, data analysis tools, and research methodologies.

4. Research Contribution:

- GAs actively contributes to ongoing research projects by conducting literature reviews, formulating hypotheses, collecting, and analyzing data, and developing algorithms.

- They collaborate with faculty, fellow researchers, and other stakeholders.

5. Publication and Presentation:

- GAs has opportunities to contribute to academic publications, conference papers, and presentations.

- They may present their research findings at conferences, seminars, and research group meetings.

6. Collaboration and Networking:

- Engaging in collaborative efforts within the research lab and with external partners fosters a rich research environment.

- Networking opportunities within the university and broader research community are encouraged.

7. Progress Evaluation:

- Regular evaluations, feedback sessions, and progress meetings help assess the student's contribution to the research lab.

- Adjustments to project assignments or additional support may be provided based on the evaluation.

8. Existing:

- Successful completion of project/s may help student get practical exposure and to include it in their portfolio.

9. Transition or Continuation:

- Graduates may transition to new roles within academia, industry, or research.

- Some may continue their involvement with the research lab as postdoctoral researchers, contributing to ongoing and new projects.

10. Alumni Engagement:

- Graduates remain part of the research lab's alumni network, fostering ongoing connections with the lab, university, and fellow researchers.

- Alumni engagement may involve mentorship, collaboration, or participation in events.

11. Legacy and Impact:

- The impact of Graduate Research Assistants is seen not only in their individual contributions but also in the collective legacy they leave within the research lab.

- Their work contributes to the lab's reputation, ongoing projects, and the advancement of knowledge in the field.

Appendix 2

Graduate Research Assistance Timeline and Major Events

Table 5: Description on key events of the lab member

Key Event	Small Description
Onboarding into the Project Assignment	Information Received from the project Supervisor. (Debriefing from the supervisor about the project requirements and the clients' expectations)
Operations	- The project team will write code for each task that the supervisor assigns. - Then they will explain the code and the process to the supervisor. - The supervisor will communicate with the clients and present the work done by the development team. (The development team does not have any direct contact with the client.) - If the client requests any changes or modifications to the current work, the supervisor will inform the development team. - The development team will make the necessary changes and send them back to the supervisor.

Appendix 3

Job Role Portfolio: Graduate Student Research Assistant - Data Science

COMET Lab – IU Luddy Indianapolis

Position Overview:

As a Graduate Student Research Assistant within the data science projects at COMET Lab, students will be able to utilize their knowledge into practice on real world impactful data intensive projects to gain hands-on project experience.

Key Responsibilities:

Research/Project Support:

- Assist in the design, implementation, and evaluation of data science projects. In-terms of identification of data schemas and analytical methods

- Collaborate with lab supervisors, peer research assistants and relevant external parties or stakeholders who needed to success of the project deliverables.

Data Analysis:

- Collect, clean, and preprocess data for analysis.

- Apply statistical and machine learning techniques to derive meaningful insights.

- Contribute to the development and implementation of data analysis pipelines.

Programming and Software Development:

- Utilize programming languages such as Python, R, or SQL for data manipulation and analysis.

- Contribute to the development of software tools or algorithms supporting project objectives.

- Collaborate with team members on code reviews and optimization.

Collaboration and Communication:

- Work closely with interdisciplinary teams to achieve project milestones.

- Participate in project related group meetings and discussions.

Professional Development:

- Attend relevant workshops, seminars, and conferences to enhance skills and knowledge.

- Engage in continuous learning and contribute to a positive research culture.

- Take advantage of networking opportunities within the university and broader research community via project base.

Qualifications:

- Current enrollment in a graduate program in Data Science.

- Strong analytical and problem-solving skills.

- Proficiency in programming languages such as Python or R.

- Familiarity with data analysis tools and libraries.

- Excellent written and verbal communication skills.

- Ability to work collaboratively in a project team.

Preferred Skills:

- Previous project experience in data science or a related field.

- Knowledge of machine learning algorithms and frameworks.

- Experience with big data technologies and tools.

- Strong organizational and time management skills.

Appendix 4

Job Role Portfolio: Project Manager - Data Science

COMET Lab – IU Luddy Indianapolis

Position Overview:

The Project Manager for the Data Science Research at COMET Lab will play a pivotal role in planning, executing, and overseeing data science projects, ensuring seamless collaboration among students who work under the project, faculty, and external stakeholders. This position requires a blend of project management and technical knowledge in data science and analytics with effective communication skills.

Key Responsibilities:

Project Planning and Execution:

- Develop comprehensive project plans, timelines, and budgets for data science research projects.

- Coordinate project activities, ensuring alignment with research objectives and timelines.

- Oversee the execution of projects, tracking progress and addressing any issues that may arise.

Stakeholder Collaboration:

- Foster collaboration among interdisciplinary teams, including students (Lab members), faculty, and external partners.

- Communicate project updates and milestones to stakeholders, ensuring transparency and engagement.

- Facilitate effective communication to bridge the gap between technical and non-technical team members.

Resource Management:

- Allocate resources efficiently, considering personnel and technology requirements.

- Collaborate with lab supervisors to ensure the availability of necessary resources for projects.

- Optimize resource utilization to enhance overall project efficiency.

Risk Management:

- Identify potential risks to project timelines and deliverables.

- Develop and implement risk mitigation strategies to address challenges proactively.

- Maintain a risk log and update project stakeholders on risk status regularly.

Quality Assurance:

- Implement and enforce quality assurance processes to maintain quality research standards.

- Collaborate with students to ensure adherence to ethical guidelines.

- Monitor and evaluate project outcomes to identify opportunities for improvement.

Qualifications:

- Proven experience in project management, preferably in a research or academic setting.

- Familiarity with data science concepts, methodologies, and tools.

- Proficiency in project management software and tools.

- Strong organizational and multitasking abilities.

- Excellent interpersonal and communication skills.

Preferred Skills:

- Knowledge on the project management in IT industry

- Experience with Agile or Scrum methodologies.

- Familiarity with data science programming languages (e.g., Python, R, SQL).

- Knowledge of research compliance and ethics in data science.