1. Introduction

CKAN is an open-source DMS (data management system) for powering data hubs and data portals. CKAN makes it easy to publish, share and use data. It powers hundreds of data portals worldwide. Our client, Link Digital, is a co-steward of CKAN. We will adopt CKAN as the primary tool for data management, performing an analysis similar to those conducted by civil society organizations engaged in data analysis research, social research, or policy analysis. Specifically, we will examine and document the determining factors of housing rental prices in the ACT region to provide insights. These research activities will serve as a valuable indicator of CKAN's usability, providing important feedback to the community. Moreover, we will also summarize and reflect on our journey of learning about CKAN and working with the client Link Digital. Through the results of these reports, we will be able to contribute to the value pursuit of our client.

2. Project

2.1 Objective

The project's ultimate purpose is to identify and address obstacles or limitations that hinder our team’s progress through a data analysis project where CKAN is employed as supporting data infrastructure. The project does not have a specific product development direction; instead, participants will explore the open source ecosystem, experience it as members, provide advice on CKAN and then add new value. Moreover, we will also provide our reflection on learning CKAN and working with Link Digital as expected by the client.

2.2 Problematization

2.2.1 Problem Statement

Although CKAN has achieved a significant market penetration and has a sizable installation base, there are several inherent limitations in achieving full utilization of open data and government transparency. The areas that require improvement include:

1. User Experience Improvement: The need for a more user-friendly interface or more functions.
2. Innovation: Fostering more innovative uses of the platform.
3. Public Engagement: Increasing community involvement and engagement.

4. Transparency: Improving the transparency and accountability of government through data.

Students participating in this project will explore these limitations and work on developing solutions.

2.2.2 Solution Approach

The exploration of solutions will be divided into two phases over two semesters. In the first semester, students will engage as members of the CKAN community, using CKAN to conduct a data analysis research project on a chosen topic. Through this process, they can gain deep insights into CKAN. They will generate CKAN analysis reports regarding the advantages and shortcomings of CKAN. In the second semester, based on the insights from their analysis, students will work on practical improvements, potentially developing new libraries or proposing enhancements to the CKAN system.

2.3 Bounds

In the first semester, this team will conduct the following:

○ **Engagement with CKAN Community:** Actively participate in the CKAN community to familiarize yourself with the ecosystem and its functionalities.

○ **Data Analysis Research Project:** Conduct a data analysis research project. In this process, CKAN will be used as the data management system for data management of this project.

○ **Evaluation of Platform Usability:** Assess the usability of CKAN during the data analysis process, identifying any issues related to user experience and accessibility.

○ **Identification of Limitations:** Perform an in-depth analysis of CKAN's existing limitations, such as user experience, data accessibility, and metadata quality.

○ **Documentation of Findings:** Compile and document the analysis and findings to submit as the final deliverable for the first semester assignment.

2.4 Success criteria

1) **Data Analysis Research Completion**: A comprehensive data analysis project is successfully conducted using CKAN as the data management system, and the findings are documented and submitted.

2) **Opportunity Identification**:Identify and analyze at least three key opportunities to improve CKAN during the data analysis process.

3) **Active Community Participation**: Students actively participate in the CKAN community, contributing to discussions and utilizing community resources.

4) **Insightful Report Submission**: A detailed report is submitted that documents the analysis process, the opportunities identified, and the proposed areas for improvement.

5) **Insightful Reflection Report Completed**: A detailed and comprehensive reflection that summarizes and reflects on our journey of learning about CKAN and working with the client Link Digital is completed.

3. Data Analysis Research Project Agenda

Topic: The determining factors of rental price in ACT .

In this process, we will be using CKAN to support our data management (including data collecting, data cleaning and data storage). Then we will generate 'CKAN analysis report' to reflect on our experience of using CKAN to support our research. We will comment on what are the benefits of using CKAN and what are the shortcomings that can be improved by fixing bugs or adding extensional function. Finally, we will reflect on this journey of learning CKAN and working with the client Link Digital. It is notable: although this research is now restricted to the range of ACT, it may be relocated into another geographical range due to reasons like legality

The Agenda of this research is as following:

1) **Data collection**: We will conduct web scraping on real estate websites such as Domain.com.au to collect rental data. In this process, we will also consider the possibility of automating this process by developing a program.

2) **Data management**: We will use CKAN to store and manage our data.

3) **Data cleaning**: We will use Python to clean our data. This process involves the use of CKAN to support.

4) **Generating CKAN analysis report**: Based on the above experience of using CKAN, we will generate an initial analysis report of CKAN. We will state our experience of using CKAN to manage our data, describe how CKAN supports our tasks of data collection, data cleaning, and then comment on the advantages of using CKAN. We will also provide advice on how this process and CKAN can be improved.

5) **Data analysis**: We will use machine learning models to analyze the dataset collected. With the capability of machine learning models to discover potential patterns, it is believed

that we would be able to discover the determining factors of rental price. Moreover, we will consider and try to discover the implication under these determining factors.

6) **Generating data analysis research report**: We will use the results obtained above to generate a research report.

7) **Publishing data analysis research report:** We will publish our data analysis result using website constructing technology.

8) **Generating CKAN analysis final report** : After completing the above stages, we will generate the final analysis report of CKAN. We will state our experience of using CKAN to manage our data, describe how CKAN supports our tasks and then comment on the advantages of using CKAN. We will also provide advice on how CKAN can be improved.

8) **Generating Reflection**: we will then summarize and reflect on our journey of learning about CKAN and working with the client Link Digital.

4. plan

Week	Task Description	Deliverables
W1	Project kick-off and team formation	Team Formation
W2	Initial client meeting (defining specific tasks, requirements, and deliverables) establishing team member roles and responsibilities understanding task background (Evidence Act and CKAN)	Statement of Work Charter

W3	Develop an understanding of CKAN, Data for Policy and Link Digital: 1. **Understand Link Digital's business value drivers:** Develop an understanding of how Link Digital supports CKAN and how it creates direct and indirect value for its clients, staff and community stakeholders. a.Kick off Session b.Watch: Initiative with Paul, Ian, Adria, Sergey c.Watch: Public Forum on Link's business model changes d.Read: Blog post on Link's site 2. **Introduction to CKAN and community activity**: Gain an overview of CKAN and consider its capabilities to support data-driven decision-making in the public sector. a.Review: CKAN Documentation, especially API Guide b.Watch: Table Designer YouTube clip c.Review: POSE Phase I program outcomes for CivicDataEcosystems 3. **Overview of Public Sector Data Challenges**: Develop an understanding of the challenges and opportunities in data management for policy-making in government settings. a.Watch: Keynote at Canadian Open Data Summit 2023 b.Watch: Adaptive Policy Management with Pia Andrews c.Review: Data for Policy topic and community of interest d.Read: USA Evidence Act 2018 and community of interest interview e.Watch: ONDC Clip for Australian intentions from one year ago	a.Your own view on how the team might use your strengths to best deliver a project related to CKAN, public sector data operations and the business value drivers of Link Digital. b.Your ideas for ways the team can build a CKAN portal that demonstrates value for the various stakeholders you have learnt about from the material covered.

W4	Requirements Engineering: 1. **Identify Key Workflow Value Drivers:** Identify and rank workflows in public sector policy-making that can benefit from improved data management and analysis via CKAN itself, or via periphery components that can interoperate with CKAN's primary value drivers. a.Pair up to have three pairs, each looking into different aspects of workflow or data pipelines that you consider relevant. Such as: **- inbound logistics** of the data pipeline (features to support harvesting, ELT/ETL, data enrichment before load, data cleanup or policy checks before load, , **- data operations** (data governance, tools to support data management and publishing process, data quality checks, data lifecycle auditing, automation or other controls) **- outbound logistics** (data viz, data stories, data referenceability, data subscriptions, etc - how data is delivered for value to end users or other machine users) b.Nominate a coordinator to bring the team's efforts together in an additive manner, to evolve the team's understanding of workflow components and generate a way of listing and describing these for points 2 and 3 below. **2. Stakeholder Analysis:** Identify and cross reference stakeholders (policy makers, analysts, data managers, civic tech contributors, civil society) and their specific needs for data management and analysis with key workflows. a.Same pairs, or mix it up, but add a dimension to your analysis for suggesting who would value various workflow or data pipeline capabilities. b.Continue with coordination to bring the analysis together	Analysis report about identifying key workflow value drivers and stakeholders.

W5	1. **Topic Selection**: The topic to be analyzed with the using of CKAN is selected:The determining factors of rental price in ACT; 2. **Research/literature review:** conduct review of selected articles regarding rental market. In this process, focus on: how to collect data using Python, how to clean data using Python and how to use machine learning models to analysis dataset ; 3. **Data collection**: start with Data collection. Decide the features of data to be collected. 4. **CKAN installation**: Configure the CKAN environment and then install it.	Summarizing documents on findings about the literature reviewed
W6	1. **Data collection**: continue on data collection 2. **Data management**: upload the collected dataset to CKAN at the end of this week in order to manage data using CKAN. 3. **Data cleaning:** start using the data cleaning function provided by CKAN to conduct the initial data cleaning. Then use python to conduct further data cleaning 4. **CKAN analysis:**Begin to identify potential limitations that CKAN platforms encounter during the general data management process including data uploading,data cleaning, such as data accessibility, processing efficiency, or user interface issues.	A in-the-prograss dataset collected for analysis with features selected A in-the-progress analysis report regarding CKAN
W7	1. **Data collection**:finish data collection 2. **Automation of data collection**: consider the possibility of automating this process by developing a program. Reflect on this consideration and summarize discussion regarding this. 3. **Data cleaning:** Finish the data cleaning. 4. **CKAN analysis:**Complete data analysis 5. **Data analysis:** apply machine learning methods on the cleaned dataset to decide what are the most determining factors of rental price in ACT 6. **CKAN analysis:** Draft a preliminary analysis report.	A dataset collected for analysis with features selected A report about automation of data collection with discussion of advantages, disadvantages, feasibility and potential issues.

		A cleaned dataset a preliminary analysis report.
W8	**CKAN analysis and gain feedback:** Continue to work on the report.Prepare a summary and share it with the CKAN community for feedback.Collect feedback on findings and suggestions for improvements.	Documentation for feedback and insights from the CKAN community.
W9	**Final report preparation and Presentation:** Drafting the final report, including data analysis,CKAN analysis report including: limitation identification, usability assessment, and community feedback.Ensure that the report clearly presents findings, challenges and proposed solutions or CKAN improvements. Moreover, start to prepare the presentation of our data analysis result using website constructing technology.	Draft version of the report.
W10	**Final report and presentation refining:** Refine and finalize reports. Review and refine reports with emphasis on clarity, completeness and consistency.Incorporate additional insights or suggestions for improvement based on further analysis or community feedback. Prepare reports for submission. Refine and improve the result presentation website.	Prepare final report for submission and presentation. Draft version of result presentation website.
W11	**CKAN analysis report submitting and Reflection Report starts:** Submit the final report. Submit the final report as a term assignment deliverable. Start	a final CKAN analysis report

W12	**Result discussing**: discuss the final report with client on the following aspect: client satisfaction, client expectation on the next semester	a discussion record

5. basis

5.1 Assumption

1) Availability of Resources: All necessary resources, including personnel, technology, and data, will be available throughout the project. This includes access to CKAN documentation, tools, and infrastructure for development and testing.

2) Technology Stability: The CKAN platform, along with Python and GitOps, will remain stable and supported throughout the project lifecycle.

3) Community Support: The CKAN community will continue to provide support through forums, documentation, and other resources, ensuring that students can access needed information and assistance.

4) Data Availability: External data sources required for the analysis will be accessible and reliable throughout the project period.

5.2 Dependency

1) **External Data Sources**: The project relies on the availability of external data sources, such as government datasets and reports. Any delays or issues in accessing these could impact the project timeline.

2) **CKAN Ecosystem**: Success depends on the existing CKAN ecosystem, including documentation, community support, and other CKAN resources.

3) **Participant Engagement**: The active participation and commitment of all project members are crucial for the successful completion of the project phases.

5.3 Constraints

1) **Time Constraints**: The project must complete all phases and deliverables within the specified timeline. This includes participating in the CKAN community, conducting data analysis, documenting findings, and delivering presentations, making effective time management essential.

2) **Data Accessibility**: There may be limitations in the availability and accessibility of external data sources (e.g., government reports, datasets). Delays or issues in acquiring the necessary data can impact the project schedule.

3) **Technical Limitations**: There may be functional limitations within CKAN and related technologies. Certain requirements might not be readily supported by CKAN's default functionalities and might require customization.

4) **Regulatory and Legal Constraints**: The project must comply with legal requirements, particularly concerning data privacy and security, which may require additional time and resources.

5) **Stakeholder Requirements**: Balancing the needs and expectations of various stakeholders may require adjustments and additional coordination.

6) **Technology Stack Stability**: The stability of CKAN, Python, GitOps, and related technologies is critical; any technical issues could impact the timeline and deliverables.

7) **Scope Creep**: The project must avoid scope creep, ensuring that only in-scope activities are pursued within the given timeframe and resources.

5.4 Risk and mitigation table

Potential Risk	Mitigation
Team management	Clearly define and assign the tasks to each member with clear required quality of work and due date. Regular meeting and discussion of the project.
Data Accessibility Issues	Establish alternative data sources early in the project and ensure regular communication with data providers to address potential delays.

Technical Challenges with CKAN	Leverage community support and CKAN documentation; allocate extra time for troubleshooting and learning.
Time Management	Develop a detailed project timeline with milestones and regularly review progress to stay on track.
Legal risk on data collection	Review the term of use and policy regarding data scraping before conduct data collection. Respect and follow these policy and relevant laws.

6. Acceptance

Diao Fu (Team Leader):

Signature: Diao Fu Date: 25/08/2024

Chuang Ma (Team Member):

Signature: Chuang Ma Date: 25/08/2024

Yuxin Mu (Team Member):

Signature: Yuxin Mu Date: 25/08/2024

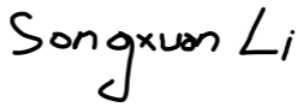
Anbo Wu (Team Member):

Signature: Anbo Wu Date: 25/08/2024

Qifeng Zheng (Team Member):

Signature: Qifeng Zheng Date: 25/08/2024

Songxuan Li (Team Member):

Signature:  Date: 25/08/2024

Doeun Han (Team Member):

Signature:  Date: 25/08/2024

Steven De Costa (Client):

Signature:  Date: 28/08/2024