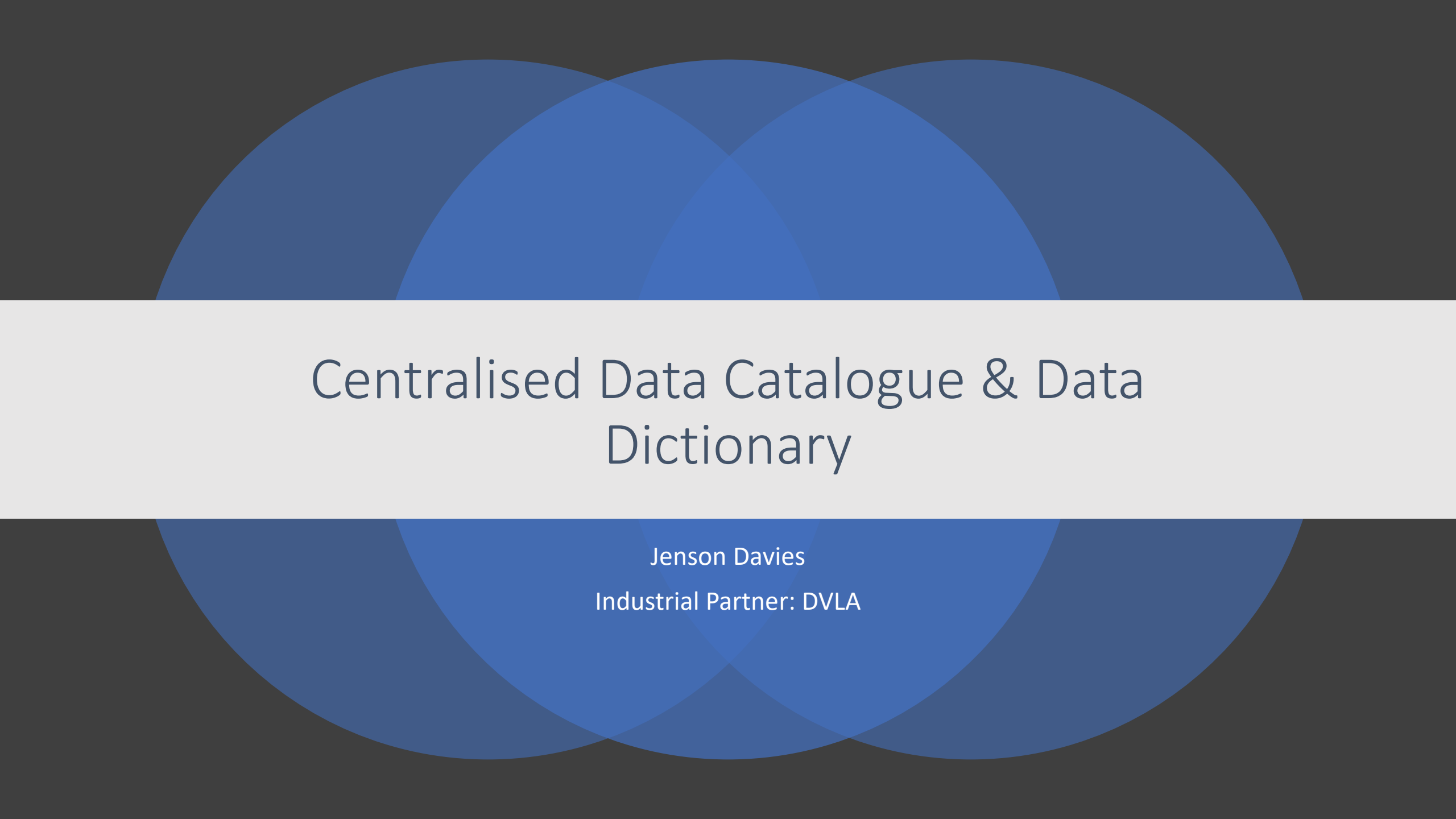




Centralised Data Catalogue & Data Dictionary

Jenson Davies
Industrial Partner: DVLA

Research Context

Current Situation:

- No single, up-to-date data catalogue or data dictionary
- Dataset and field information can be difficult to locate
- Engineers rely on existing documentation and colleagues' knowledge
- Additional time may be required to identify suitable data for business requests

Impact:

- Increased development time
- Difficulties with data discovery and governance
- Reliance on individual knowledge
- Greater risk of misunderstanding unfamiliar data

Proposed Project Solution:

A centralised web-based application that will provide authorised users with one location to:

- Discover available datasets
- Understand the purpose and ownership of specific datasets
- Access information regarding fields

Research Problem

Engineers rely on scattered existing documentation and individual knowledge to locate and understand data within the DVLA Data Engineering Team.

Data Discovery:

- Difficult to locate suitable datasets.
- Time spent searching for relevant data.
- Information is not available through a unified source.

Data understanding:

- Unfamiliar datasets can be difficult to interpret.
- Limited information regarding the purpose and meaning of data.
- Information regarding individual fields may be difficult to locate.

Data Governance:

- Difficult to establish who is responsible for data.
- Limited information regarding ownership and purpose.
- Makes consistent data understanding more difficult.

Research Aims & Objectives

AIM:

Design & develop a centralised web-based data catalogue and data dictionary application for the DVLA, to improve the discoverability and understanding of existing data.

Objectives:

1) Analyse:

Analyse the current processes used by the DVLA Data Engineering Team to locate and understand data.

2) Evaluate:

Evaluate academic resources and existing industrial data catalogues and data dictionaries.

3) Design & Develop:

Design and develop a functional application providing: a centralised data catalogue, data dictionary, search & filtering functionality and data navigation.

4) Evaluate Effectiveness:

Evaluate whether the application: improves data discoverability & understanding, supports data governance and reduces the time and difficulty with locating required data.

Literature Review

Data Catalogue:

- Provides a central means for discovering and comprehending available data assets.
- Gives an overview of distributed data assets.
- Helps users find suitable datasets.
- Modern catalogues can provide a range of features to support data discovery.

Data Dictionary:

- Provides detailed information regarding individual fields within datasets.
- Includes information such as:
 - Field names
 - Data types
 - Field descriptions
- Helps users determine whether the contents of a data set match their requirements.

Data Catalogue:

- Find the appropriate dataset.



Data Dictionary:

- Understand the contents of the dataset.

Literature Review

Development of Data Catalogues:

- Data catalogues have developed alongside the increasing distribution of organisational data.
- Modern catalogues provide a wide range of functionality.
- A review of 39 catalogue tools identified 24 unique features.

Data Discovery & Accessibility:

- Search, filtering and data inventory are important for discovering relevant datasets.
- Navigable interface can make data easier to locate.
- Information on the purpose and contents of a dataset help users to assess its suitability.

Data Governance & Understanding:

- Data catalogues can support governance by providing information on data ownership and purpose.
- Ownership information can help clarify responsibilities.
- Field descriptions can help improve user's understanding.

Relevance to Project:

Literature findings and DVLA user requirements will help to establish an informed design for the application.

Research/Development Methodologies

Research Philosophy

Interpretivism:

- Understand how the Data Engineering Team's locate & understand data.
- Identify difficulties experience during the processes.
- Establish requirements based on users' experiences.

Research Approach

Inductive:

- Investigate the existing process.
- Identify common issues and user requirements.
- Use the findings to determine the functionality of the application.

Research Strategy

Software Development + Case Study:

- Develop a practical software solution to the organisation problem.
- Investigate the current data discovery process.

Development Process:

Requirements Gathering → Application Design → Iterative Development → Testing & Refinement → Evaluation

Data Collection:

Team Discussion + Documentation + User Feedback

Resources/Materials

Technical

- Databricks
- Databricks Apps
- Unity Catalog
- Development Software

People

- Data Engineering Team
- Requirements
- Evaluation feedback

Documentation

- DVLA documentation
- Confluence
- Existing processes

Financial

As existing resources will be used no additional expenditure is anticipated.

Project Plan

Core System – Must Have:

- Centralised Data catalogue
- Data Dictionary
- Search & Filtering
- Dataset and field information

Additional Functionality – Should Have:

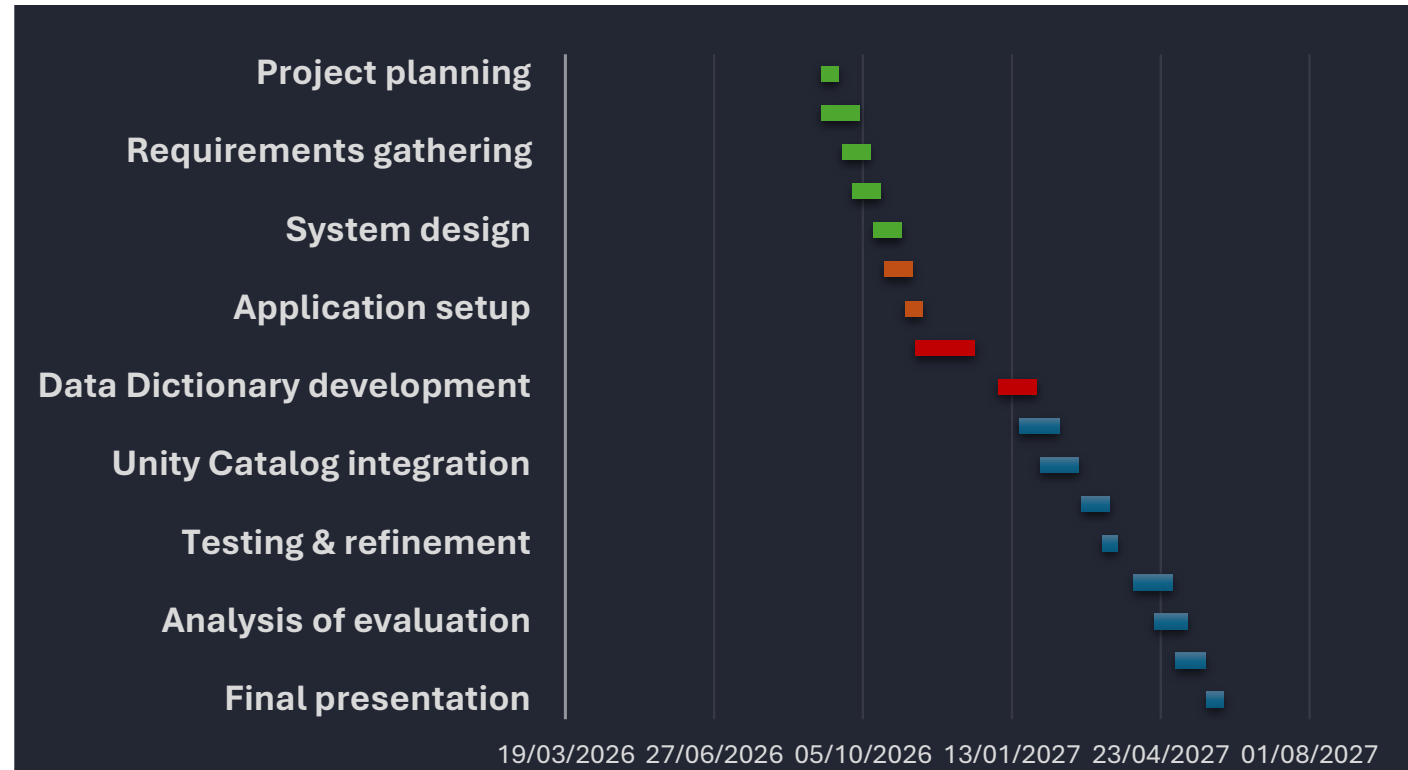
- Additional features where time allows
- Improvements base on user requirements

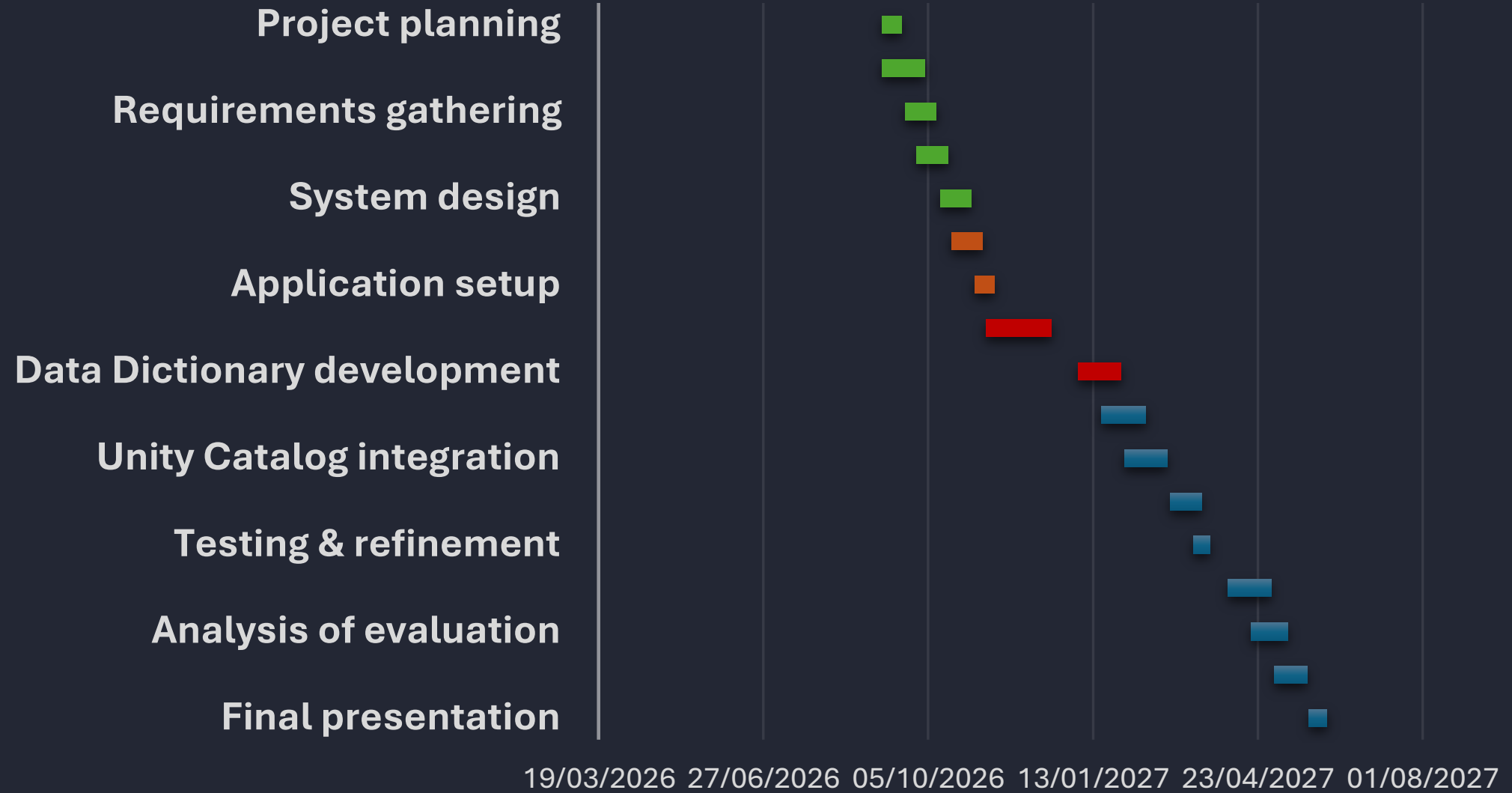
Future Functionality – Could Have:

- AI feature to assist with understanding data

Project Management:

- MoSCoW Prioritisation
- Timeboxing
- Regular Reviews





References

- [1] D. Petrik, A. Untermann, and H. Baars, “Functional Requirements for Enterprise Data Catalogs: A Systematic Literature Review,” *Lecture Notes in Business Information Processing*, pp. 3–18, 2024, doi: 10.1007/978-3-031-53227-6_1.
- [2] M. Tonnarelli, I. Kumara, S. Driessen, D. A. Tamburri, W.-J. van den Heuvel, and P. Oor, “Data catalog tools: A systematic multivocal literature review,” *Journal of Systems and Software*, vol. 230, p. 112584, Dec. 2025, doi: 10.1016/j.jss.2025.112584.
- [3] O. Olesen-Bagneux, “The Enterprise Data Catalog,” *Google Books*, 2026.
https://books.google.co.uk/books?id=9SWuEAAAQBAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false (accessed Sept. 04, 2026).