Project plan

Euclid – Delta project 2025

10/10/2025 – v1

By: Saamie Vincken

Version control

Version	Date	Changes	Author
0.1	29/09/2025	Setup of document	Saamie Vincken
0.2	01/10/2025	Background, context and stakeholder sections	Saamie Vincken
0.3	02/10/2025	Introduction, project group and methods sections	Saamie Vincken
0.4	03/10/2025	Name changes to sections	Saamie Vincken
0.5	04/10/2025	Initial setup of planning section	Saamie Vincken
0.6	09/10/2025	Research questions, deliverables, scope and risks sections	Saamie Vincken
1.0	10/10/2025	Finalizing document version 1.0	Saamie Vincken

Distribution

Version	Date	Distributed to	Status
1.0	10/10/2025	Canvas	None

Table of Contents

1. Introduction

This document describes the project plan for the Euclid – Delta Project 2025, carried out as part of the Delta excellence program at Fontys University of Applied Sciences. The project takes place in collaboration with the **Euclid Strong Lensing Working Group (SWG)**, which forms part of the wider **Euclid Consortium** coordinated by the **European Space Agency (ESA)**.

The purpose of this project is to develop a research-oriented software pipeline that supports ongoing scientific research into strong gravitational lensing using data from the Euclid Quick Data Release 1 (Q1). The developed pipeline will consist of several (beta-level) software features that address specific research questions formulated together with the SWG. These questions focus on improving data processing, machine learning analysis, and understanding systematic differences between simulated and real Euclid data.

The project combines technical implementation with scientific exploration. Its outcomes are intended to provide both practical tools and research insights for members of the Euclid Consortium, supporting future data releases and model development within the Strong Lensing Discovery Engine (SLDE).

This plan outlines the project's background, stakeholders, objectives, methodology, scope, risks, and expected deliverables, along with the evaluation approach and timeline that will guide execution during the semester. The document serves as both an internal coordination framework for the project group and a formal reference for coaches and stakeholders.

2. Background

The Euclid Space Telescope is a mission that maps one third of the observable universe to understand the properties of dark matter and dark energy. One of the main ways to study these is through strong gravitational lensing, which happens when the light from a distant galaxy is bent by the gravity of a massive object in front of it. By studying these effects, scientists can learn more about the distribution of dark matter and improve measurements of the Hubble Constant, which describes how fast the universe is expanding.

The Euclid mission releases its data in several stages, including the Quick Data Release 1 (Q1). This release was first available only to members of the Euclid Consortium but became public in March 2025. It includes the Strong Lensing Discovery Engine (SLDE) catalogue[1], which contains all strong lensing systems found in Q1.

3. Context and Stakeholders

The strong lensing discovery engine (SLDE) catalogue has already been used for many (scientific) goals such as training and retraining machine learning networks which will be applied to future data releases, however there are many research questions open that have not yet been addressed. These research questions are mainly formed by the scientists' part of the Strong Lensing working group (SWG) as part of the Euclid Consortium. These scientists are also responsible for the creation of the SLDE catalogue and will be the main users of the results coming out of the research questions addressed during this project.

Therefore, the scientists from the SWG are the main stakeholders for this project. The specific stakeholders forming the research questions and providing feedback during this project will be:

1. **Dr. Karina Rojas** (based in Brugg, Switzerland) – Stakeholder role within this project consists of giving feedback on data and software applications created.
2. **Dr. Prof. Tom Collett** (based in Portsmouth, England) – Co/Lead of the SWG, Stakeholder role within this project consists of giving feedback on scientific relevance and creation of the research questions.
3. **All other SWG members (50+ participants)** including other Co/Leads of the SWG **Dr. Prof. Fred Courbin**, **Dr. Prof. Raphael Gavazzi** and **Dr. Prof. Ben Metcalf**.

[1] https://zenodo.org/records/15025832

4. Project group

This project is carried out by six team members, each with their own focus but working together on a single integrated software pipeline. The group combines expertise in data analysis, software development, and machine learning.

- **Saamie Vincken (project lead):** responsible for project management, coordination, and communication. Focus on data analysis and ML testing.
- **Amal Khairunnisa Mardhiyyah Indarto:** focus on data analysis and ML applications within the software pipeline.
- **Vincent van Gent:** focus on data analysis and data conversion within the software pipeline.
- **Christian Măgureanu:** focus on ML applications and integration in the software pipeline.
- **Stan Rareş:** focus on ML applications and integration in the software pipeline.
- **Malwina Raczyńska:** focus on data analysis and data conversion within the software pipeline.

Contact between project members will take place in a Microsoft Teams group for informal discussion and communication about meetings etc. Formal documentation and planning will take place on GitHub (as discussed in the Methodology section).

5. Methodology

This project applies the DOT framework[2] to structure the overall research approach. The framework ensures that the research questions are explored through different complementary perspectives rather than a single method. The five parts of the DOT framework; Library, Field, Lab, Showroom, and Workshop are used as a basis for the research approach.

In this project, the main focus is expected to be on the Library, Field, and Lab perspectives. The Library research will be used for literature study and analysis of previous work related to Euclid data processing, lens classification, and image preparation. The Field perspective will be applied for data analysis and collection of relevant datasets to gain a better understanding of the data characteristics. The Lab perspective will focus on the technical implementation and validation of experiments, including data analytics, model validation, and model evaluation.

The method for project planning and version control or outcomes will all be within the GitHub environment. A shared repository will be used for all source code and documentation. Planning, backlogs and progress will be organized using a GitHub project board in combination with the SCRUM methodology. SCRUM provides an adaptive way of working, allowing the team to continuously evaluate results and adjust direction where needed based on stakeholder feedback or complexity of tasks. Two stand/up meetings per week will be held to monitor progress and discuss challenges, and three weekly demo and reflection moments will be used to present results and evaluate personal and technical progress within the group.

[2] https://ictresearchmethods.nl/dot-framework/

6. Research questions

A set of research questions has been defined in collaboration with the project members and the Science Working Group (SWG). Each project member will initially focus on one of these questions. The main research questions are:

1. *Which image compression format provides better performance for storing photometric color band data compared to the current JPEG format?*

2. *Which color processing method provides more accurate mapping of photometric color bands compared to the current RGB approach?*

3. *Can images of gravitational lenses from other missions (e.g., Hubble or James Webb) be transformed to match the visual characteristics of Euclid images, in order to expand the training dataset for machine learning applications?*

4. *What types of classifications can be performed on the current SLDE catalogue to provide more detailed insight into the distribution and characteristics of its subclasses?*

5. *Can unsupervised learning methods be applied for unbiased classification of one or more subclasses within the SLDE catalogue?*

6. *What is the impact of domain shifts between simulated data and the real data in the SLDE catalogue on model performance and generalization?*

Each of these main questions will be supported by several sub-questions, which will later be developed into specific, actionable tasks during the course of the project. Each project member will formulate their own sub-questions based on their assigned topic. These will be reviewed and discussed collectively within the group to ensure a consistent and coordinated workflow across all project members.

7. Expected deliverables

Each research question will result in a deliverable that consists of two parts:

1. **Documentation** describing the approach, findings, and evaluation results.
2. **An implemented feature** within the shared software pipeline that performs the corresponding processing task.

All deliverables will be combined into one integrated (beta) pipeline that can be used and further developed by members of the SWG. The main deliverables are:

- **Image compression comparison:**
  A pipeline feature that processes raw Euclid data and converts photometric color bands into an optimized compressed format.
- **Color processing method:**
  A pipeline feature that implements an improved color mapping method to create more accurate representations of the photometric bands captured by Euclid.
- **Cross-mission image transformation:**
  A pipeline feature that converts gravitational lens images from other missions into Euclid-like versions for use in machine learning training datasets.

4. **SLDE catalogue classification:**
   A pipeline feature that performs classification and visualization of subclasses within the SLDE catalogue to gain insight into their distribution and characteristics.
5. **Unsupervised learning application:**
   A proof of concept on applying unsupervised learning methods for grouping or clustering lens systems in the SLDE catalogue. Depending on the results, this may later be integrated into the pipeline as a feature.
6. **Domain shift analysis:**
   A research task to study the differences between simulated and real Euclid data and how this affect model performance. With a goal of gaining a better understanding on the impact of domain shifts and defining main issues in the current simulation pipelines of the SWG.

Each of these deliverables will form both a technical component and a research contribution. Together, they create a single pipeline with various features (or packages) that directly addresses the research questions and can be applied to future Euclid data releases.

8. Scope

This project focuses on the development of a software pipeline that supports research on strong gravitational lensing using Euclid Q1 data. The scope is defined by the research questions listed in Section 6 and their corresponding deliverables in Section 7.

The main scope boundaries are:

In scope	Out of scope
Analysis of the SLDE Q1 data	Analysis of any data not in the public Q1 set
Implementation of software features addressing the six defined research questions	Implementation of a complete production ready software pipeline
Analysis of data compression formats	Deployment or integration of new compression formats into official existing Euclid processing pipelines
Analysis of color mapping techniques	Validation of color processing on non-Euclid data sources
Analysis of ML learning techniques	Deployment of production ready machine learning models
Integration of implemented software features into one functional beta pipeline	Continuous maintenance post project completion
Documentation of research outcomes and software functionality	Formal publication or peer review within the semester timeframe

The project’s scope therefore focuses on research, prototyping, and software design to address specific scientific questions within the Strong Lensing Working Group, rather than large-scale operational deployment.

8.1. Risks

The overall project risk is considered moderate with most risks can be mitigated through early planning, efficient communication within the team, and continuous coordination with the Euclid SWG.

Risk	Description	Mitigation
Data access	Access to data outside of Euclid SLDE may be limited.	Communicate with SWG if there are ways to receive additional data. Verify all required data at the start of the project.
Limited computational resources	Machine learning experiments may not be possible on large scale with available hardware.	Use smaller test samples, coordinate with project lead on use of shared cluster.
Students are beginner programmers	Data and project complexity can overwhelm students not familiar with advanced software, ML and data science techniques.	Provide structured onboarding workshops including git workflow, relevant (Euclid) tools and internal documentation. Divide complex tasks into smaller components.
Students are not astrophysicists	Project group consists of ICT students, not astrophysicists, topics within the research questions may be more complex for this group.	Keep close collaboration with SWG for scientific clarification, define importance of the groups' understanding about astrophysics-cosmology topics (no need to know and understand everything immediately!)
Time constraints	The semester timeframe (~5 months) may be too short to fully explore all six research questions.	Prioritize deliverables with highest scientific and stakeholder relevance.
Stakeholder feedback delays	SWG group may have limited availability to provide consistent feedback.	Schedule early review sessions, share updates trough project lead who has more consistent meetings with SWG.
Unclear or inconclusive research results	Some complex analyses may not form clear conclusions.	Document all findings and highlight methodology, testing and discussion.

9. Evaluation and validation

The evaluation and validation process ensures that both the technical and research outcomes of the project meet the intended objectives defined in Sections 6 and 7. It focuses on verifying if the contents are correct (and non-contradicting), functionality and scientific relevance of the developed software features.

Each deliverable will be evaluated according to the following general criteria:

- **Functionality**: The feature performs its intended processing task as defined in the related research question.
- **Reliability**: The feature executes without technical errors and produces consistent results across multiple runs.
- **Performance**: The feature executes within acceptable (depending on the feature) execution times and memory given the available hardware resources.
- **Scientific validity**: The feature outputs are scientifically meaningful aligned with expectations or feedback from the SWG.
- **Reproducibility**: The feature results can be reproduced using the shared GitHub repository, code and documentation.
- **Documentation**: The feature includes clear user and developer documentation.

9.1. Validation methods

To ensure the above criteria are met, validation will be performed through a combination of code testing, performance measurement, and stakeholder review. All software modules will be validated by at least one other team member trough peer review, ensuring functionality and consistent structure.

Machine learning and image analysis components will be validated using metrics (such as accuracy, precision, F1-score and ROC-AUC) to measure performance. Where applicable, results will be compared to existing Euclid benchmarks and applied to official test sets.

For features producing visual or comparative outputs, such as color mapping, compression or cross-mission image transformation, validation will be performed trough visual inspection and expert feedback from the SWG.

The SWG stakeholders, including Karina Rojas and Tom Collett, will provide scientific validation by reviewing the relevance and interpretation of results. This feedback will guide refinement during the semester and ensure the outcomes align with the goals of the SWG. All validation activities and outcomes will be documented in the GitHub repository, including evaluation notebooks, validation logs, and descriptions of applied methods. The final project documentation will include a summary of validation results, quantitative metrics, and qualitative assessments, supported by demonstrations of the functional beta pipeline during the final presentation

10. Planning and milestones

The project timeline follows the sprint timeline consistent with the general dates within the semester, as displayed below (from: Canvas OL course overview):

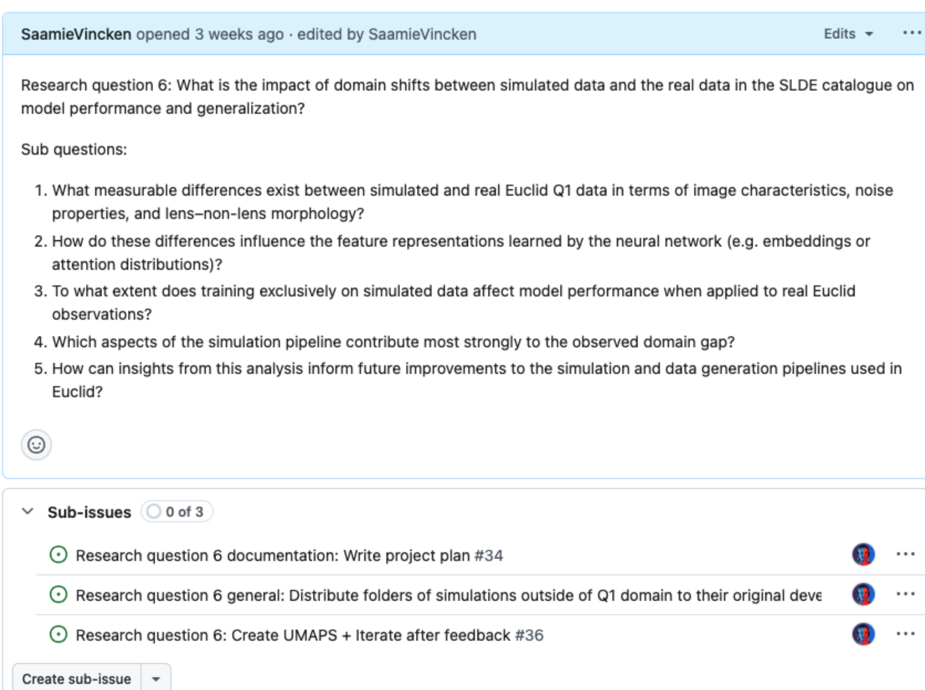
Week	1	2	3	4	5	6		7	8	9	10	11	12	13	14	15		16	17	18	19	20
Date	01/09	08/09	15/09	22/09	29/09	06/10	13/10	20/10	27/10	03/11	10/11	17/11	24/11	01/12	08/12	15/12	22/12	05/01	12/01	19/01	26/01	02/02
Sprint	Start (0)			1				2			3			4				5			End talks	End talks
Days off							Autumn break										Christmas break					
Events		Delta project day		SP**		Demo Retro		SP**		Demo Retro	SP**		Demo Retro	SP**		Demo Retro		SP**		Demo/ II*	Retro	

* II: Innovations Insight
** SP: Sprint planning

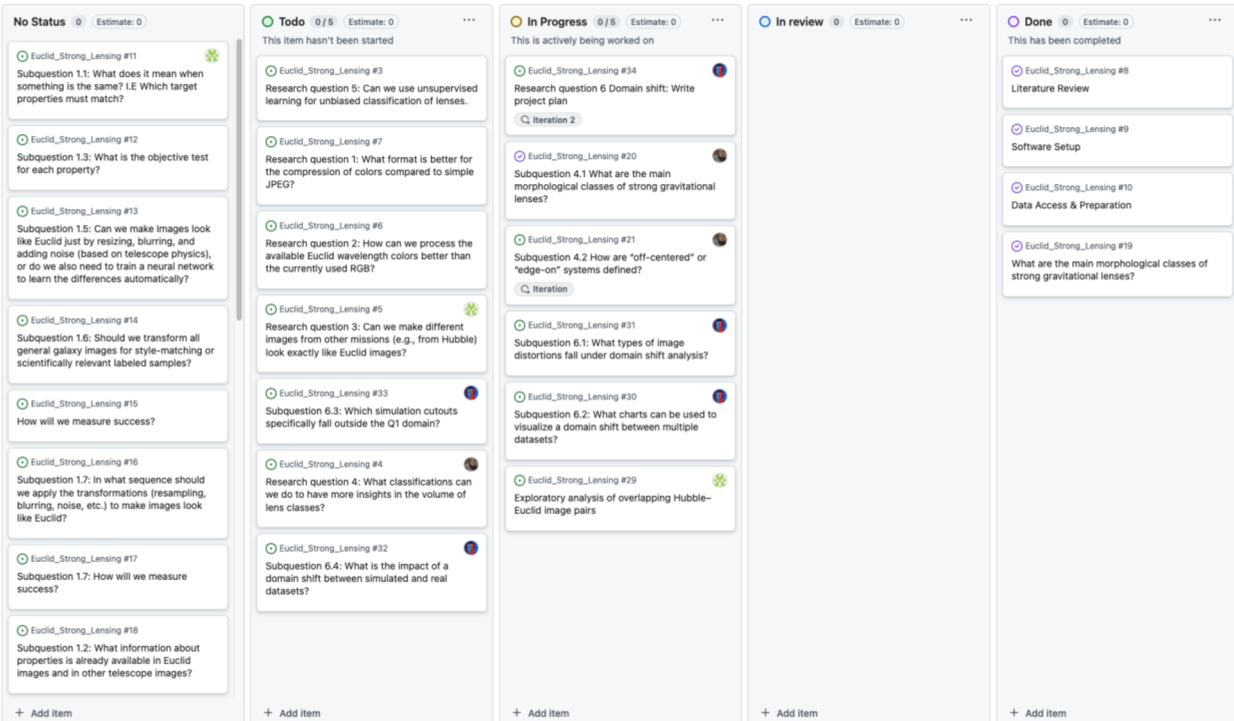
The project will have a total of 5 (0-4) sprints each consisting of 3 weeks – with the exception of the holidays and study days. As mentioned in throughout this document, the semester goal is to:

Develop a beta level software pipeline that includes features which directly addresses, and solve, the research questions formed by the stakeholders.

With this goal in mind each main research question is divided into smaller sub questions by the person responsible of that specific research question. Each main and sub question is documented in the project plan and in a GitHub issue part of the repositories' project board, where the questions can then be further split up into actionable tasks. Below is an example of one of the issues:



For this project, each main research question and deliverable (as described in Section 7) directly correspond to the main milestones. The connecting sub-questions correspond to the sprint milestones. All of the actionable tasks are automatically loaded into the projects' backlog and scrum board with the categories **No Status** (backlog), **Todo** (current sprint), **In progress**, **In review**, and **Done**, as displayed below:



Before a task is moved from '**In review**' to '**Done**' it must be reviewed by at least one other group member. The definition of '**Done**' for each task is defined during the sprint planning and documented inside the sub-tasks' description in GitHub, but will always follow the main evaluation criteria described in Section 9.

11. Resources and Links

Euclid mission overview:
https://www.esa.int/Science_Exploration/Space_Science/Euclid

Stakeholder overview (function = Strong Lensing Working Group):
https://www.euclid-ec.org/consortium/roles/

Strong Lensing Discovery Engine (SLDE) dataset:
https://zenodo.org/records/15025832

GitHub repository (request access with owner):
https://github.com/SaamieVincken/Euclid_Strong_Lensing

GitHub project board (request access with owner):
https://github.com/users/SaamieVincken/projects/1/views/1

DOT framework research approach:
https://ictresearchmethods.nl/dot-framework/