Research plan document

By: Mohammed Bouali, Calvin Kwan, Andrei Pipeline, Björn Wildeman

For: Group project SUE data science platform Advanced software

# Problem/Opportunity Statement

Currently, the SUE platform offers a CLI tool for data scientists that facilitates essential tasks such as creating notebooks and persistent volume claims (PVCs). However, the CLI can only be run locally, limiting its integration into the broader infrastructure and preventing the platform from offering these services as an integrated, scalable solution. The opportunity lies in creating an API that has the same functionalities as the CLI but that can be accessed via gRPC and integrated into SUE’s infrastructure. This would allow SUE to offer these functionalities as a service to its clients, improving scalability, accessibility, and maintainability. The primary challenge of this project is to ensure the API mirrors the CLI’s capabilities while being suitable for infrastructure-wide deployment and external use.

# Main Research Question

**How do we implement the functionalities that are in the current CLI into an API that can be integrated into SUE’s infrastructure so that they can offer it as a service to their clients?**

This central question drives the research and development of transforming the CLI into a robust, scalable, and secure API.

**Sub-question 1: How will we implement the CLI application’s functionalities into the API?**

**ICT DOT Framework Research Strategies:**

1. **Literature Study:**
   * A literature study will focus on understanding existing API design patterns in GO. Researching best practices, design patterns, and common challenges. This strategy will help in shaping the overall structure and functionality mapping between the CLI and API.
2. **Prototyping:**
   * Prototyping the API in iterative stages is essential to explore and validate the feasibility of translating CLI commands to API functions. By creating small, functional API modules that replicate specific CLI functionalities, we can test their integration and identify potential issues early in the process.
3. **IT Architecture Sketching:**
   * IT architecture sketching will help define how the API will be integrated into SUE’s infrastructure. It will provide a blueprint of how the backend will interact with the API and how existing data flows need to be modified to support API interactions. This step will also clarify the necessary changes to SUE’s current infrastructure.

**Sub-question 2: How will the API be deployed?**

**ICT DOT Framework Research Strategies:**

1. **Design Pattern Research:**
   * Researching various design patterns will help determine the most efficient, secure, and scalable method for deploying the API. For example, exploring microservices architecture, containerization, or serverless deployment patterns will help optimize the deployment based on SUE’s infrastructure.
2. **IT Architecture Sketching:**
   * Sketching the deployment architecture is crucial to understand how the API will be integrated with other components of SUE's infrastructure, such as authentication services, load balancers, and databases. This ensures the deployment process is aligned with the broader technical framework.

**Sub-question 3: How can we ensure that our software solution is GDPR compliant?**

**ICT DOT Framework Research Strategies:**

1. **Security Check:**
   * A security check will ensure that all data handled by the API is encrypted and secured in line with GDPR’s data protection requirements. It will also focus on ensuring that user data is handled, processed, and stored following GDPR principles, such as data minimization and consent management.
2. **Ethical Check:**
   * Ethical reviews will assess the API's overall data handling procedures to ensure it respects users' privacy rights. This strategy ensures that the API aligns with ethical standards and doesn't inadvertently violate GDPR through questionable data practices.

**Sub-question 4: How can we make our API able to communicate with SUE’s infrastructure?**

**ICT DOT Framework Research Strategies:**

1. **Explore User Requirements:**
   * Understanding the specific user and system requirements is crucial for ensuring seamless communication between the API and SUE’s infrastructure. This step will focus on gathering input from stakeholders and analyzing how the API must interact with SUE’s existing services, databases, and authentication systems.
2. **Literature Study:**
   * A literature study will help identify the best approaches for API-infrastructure integration, specifically focusing on gRPC-based communication. This research will provide insights into protocols, data formats, and existing technologies that can be used to facilitate smooth integration.
3. **System Tests:**
   * Performing system tests will be essential to validate that the API communicates effectively with the SUE infrastructure. Testing will focus on various aspects like latency, error handling, and load management to ensure that the API operates efficiently under real-world conditions.

# Intended Deliverables and Timeline

The research aims to deliver a fully documented and prototyped API that replicates the CLI’s functionalities and integrates seamlessly with SUE’s infrastructure. The API will be deployable via gRPC, ensuring it can be offered as a service to SUE’s clients while remaining secure, scalable, and GDPR-compliant.

**Estimated Timeline:**

* **Literature Study**: 2 weeks
* **Prototyping and Testing**: 4-6 weeks
* **Design Pattern Research and IT Architecture Sketching**: 3 weeks
* **Security and Ethical Check**: 1-2 weeks
* **Final System Testing**: 2 weeks
* **Development**: 15 weeks

**Total Duration**: Approximately 17-18 weeks