Software Requirements Specification for Software Engineering: subtitle describing software

Team #18, Gouda Engineers
Aidan Goodyer
Jeremy Orr
Leo Vugert
Nathan Perry
Tim Pokanai

September 30, 2025

Contents

Revision History

Date	Version	Notes
Date 1	1.0	Notes
Date 2	1.1	Notes

1 Purpose of the Project

1.1 User Business

The business of the user is to use the system to aid in the creation and analysis of experiments relating to a large but fragmented data set of rat trials that investigate compulsive behaviour of rats and how they relate to various factors such as drug injections or brain legions. This data set is comprised of videos of rats moving on a flat surface, files of their (x,y) coordinates and plots that show the trajectories of these coordinates. Users may want to extract a concise set of pre-existing data that can be applicable to their hypothesis to streamline their workflow and avoid the necessity of carrying out trials. Users may also want to use the system to provide analysis on the data set as a means of further supplementing ther academic work. This can include drawing behavioural metrics of the rats based on various parameters (i.e. rate of compulsion based on injection type) or data visualizations based similarily on compulsive behaviours as they relate to attributes of the rats.

Insert your content here.

1.2 Goals of the Project

The high-level goal of the project is to provide users in the field of behavioural sciences, a unified and non-technical way of drawing from the vast amount of available data on OCD rat trials to aid in their academic work and experiments. This will be accomplished via the following sub-goals:

- Create a DBMS that provides query access to the entirety of the data set of rat trials as well as all metadata associated with the data.

- Provide a UI that makes querying this data approachable to non-technical users by incorporating familiar and intuitive filtering and searching techniques (Attribute filters, Natural Language search bar). Further the UI must provide users with pre-generated query options that are likely to be useful for users that may not know where to start. This could include selecting all rat trials of a particular drug injection along with all saline control data.

- Provide an algorithm that can identify compulsive behaviour by looking at each rat trial and provide options for behavioural metrics and data

visualizations based on various attributes that the user can extract for their purposes.

Insert your content here.

2 Stakeholders

2.1 Client

The clients for this project are Dr. Henry Szechtman and Dr. Anna Dvorkin-Gheva, two professors at McMaster University who are experts in the fields of Psychiatry and Behavioural Neurosciences and Bioinformatics, respectively. They have worked extensively with these OCD rat trials and were responsible for generating and depositing this data in the FRDR repository where it currently is stored. Dr. Szechtman and Dr. Dvorkin-Gheva want to drastically improve the availability of the vast data set they have created so that the trials they conducted can become useful to other students and academics in related fields. *Insert your content here.*

2.2 Customer

Since our application is intended to be open-source and accessible to the public for free, our customers will simply be a group of end-users of our application. Our end users will primarily involve behavioural neuroscience researchers like Dr. Henry Szechtman and Dr. Anna Dvorkin-Gheva, as well as graduate students and lab members, who will all benefit from the user-friendly and accessible functionality of the platform. Additionally we will have data scientists as customers, who will benefit from our application architecture anf offered functionality for their purposes. Lastly, collaborating institutions and organizations from the open research community will be users of our application, since the application is intended to be free-use and accessible to all. *Insert your content here.*

2.3 Other Stakeholders

Other stakeholders for our project would be defined as any person or group of interest in the project outside the client and the customer. The first one would be our group, Gouda Engineers, as our task is to turn the project

from the client/customer's idea to a real product. We focus on creating documentation, building code, testing and deploying the software. Another stakeholder is our TA, Tiago de Moraes Machado, who is there to help our group along the way with our project by keeping us on track and providing feedback where it is needed. Lastly, our professor, Spencer Smith and our class, Capstone 4G06, would be the final stakeholders as they provide the guide and structure for building the project, knowledge on the topics for each step, and feedback where it is needed. *Insert your content here.*

2.4 Hands-On Users of the Project

Insert your content here.

2.5 Personas

Insert your content here.

2.6 Priorities Assigned to Users

Insert your content here.

2.7 User Participation

Insert your content here.

2.8 Maintenance Users and Service Technicians

Insert your content here.

3 Mandated Constraints

3.1 Solution Constraints

Insert your content here.

3.2 Implementation Environment of the Current System

Insert your content here.

3.3 Partner or Collaborative Applications

Insert your content here.

Collaborative System	System Overview
FRDR Repository	The Federated Research Data Repository is a 'bilingual bilingual publishing platform for sharing and preserving Canadian research data. It is a curated, general-purpose repository, custom built for large datasets.' This is where the data set of the rat trials is physically located and is dispersed across 29 independent datasets. Our system will provide a unified database schema but will pull data from this repository.
ratbat.mcmaster.ca	While not a directly collaborative system, this is a system made by a previous years' capstone team to address the same problem that our system seeks to. It will be a collaborative system in the sense that ir provides a reference of potential ways to approach our solution, ideas that work well and can be carried forward and providing visibility to shortcomings of the system will help us to avoid repeating mistakes.

3.4 Off-the-Shelf Software

- The system shall make use of the FRDR API in order to fetch the relevant data from the data set as we will not be hosting a seperate database to physically store the data for our system

- In the case of implementing natural language processing, an Off-the-Shelf LLM will be used to accomplish this (In other words we will not be writing our own LLM model for this purpose). Note that this model may need to be trained to provide results that are acceptable for our purposes.

3.5 Anticipated Workplace Environment

Insert your content here.

3.6 Schedule Constraints

Due to the nature of the capstone course, the scheduling constraints map directly to dates in which major deliverable related to the project are due in the capstone course. They are laid out below:

Project Milestone	Scheduling Constraint
Software Requirements Specification	Oct 6 2025
Verification and Validation Plan	Oct 27 2025
Design Document Revision -1	Nov 10 2025
Proof of Concept Demonstration	Nov 17-28 2025
Design Documentation Revision 0	Jan 19 2025
Revision 0 Design Demonstration	Feb 2-13 2025
Verification and Validation Report	Mar 9 2025
Extras (Performance Report + User Manual)	Mar 9 2025
Final System Demonstration	Mar 23-29 2025
Final Documentation	April 6 2025

3.7 Budget Constraints

There is very limited budget available for this project. The department of Computing and Software at McMaster University will provide $125 CAD for approved expenses. Outside of that funding, we are asked not to exceed spending of $500 CAD culmulatively as a team. Thus, the total budget constraints for this project are $500 CAD in total and $375 of our team's personal funding.

3.8 Enterprise Constraints

Insert your content here.

4 Naming Conventions and Terminology

4.1 Glossary of All Terms, Including Acronyms, Used by Stakeholders involved in the Project

Insert your content here.

5 Relevant Facts And Assumptions

5.1 Relevant Facts

- The entirety of the data set that is being worked with is available in the FRDR repository but is fragmented into 29 independent dats sets. 29 being the number of studies written that required relevant rat trials to be performed.

- The culmination of all the relevant data for this project is 11 terabytes and contains nearly 60,000 data objects with 20,000 hours of video files alone.

- Every data object has many metadata annotations (or attributes) associated within them. In the current data hosting implementation on the FRDR, these are all contained in a README file that is stored in the root directory of each dataset. This README file has a row for each data object in the dataset and columns for each attribute.

5.2 Business Rules

This system does not exist within a traditional business environment and is rather in the realm of academia. As a result, there are no specific business rules that shape the requirements, the system acts as a tool in academic pursuits. Additionally, any business rules related to protection of information do not apply as all data being used is publicly available to the general public already.

5.3   Assumptions

- The availability and accessibility of the data in the FRDR will be upheld in definitely or at least for the lifecycle of this project.

- The system will not at any point, need to physically store any of the relevant data and can use the FRDR as a datastore.

- The API provided by the FRDR repository will be made available for our use.

- The query performance of the FRDR API will be reasonably efficient for our purposes.

- An out of the box machine learning model(s) will be available and capable of being trained in a way that it can perform natural language queries and categorization of rat behaviours.

- API requests to the repository are independent of the user's operating system and device state. i.e. Users will not require any setup on their local machine to use the system.

6   The Scope of the Work

6.1   The Current Situation

Insert your content here.

6.2   The Context of the Work

Insert your content here.

6.3   Work Partitioning

Insert your content here.

6.4   Specifying a Business Use Case (BUC)

Insert your content here.

7 Business Data Model and Data Dictionary

7.1 Business Data Model

Insert your content here.

7.2 Data Dictionary

Insert your content here.

8 The Scope of the Product

8.1 Product Boundary

The product is a web-based data analysis platform designed specifically for the Dr. Szechtman Lab's OCD behavioral neuroscience dataset. Its scope includes database design, backend services, frontend user interfaces, and data analysis pipelines to enable global researchers to access, search, analyze, and visualize rat behavioral data.

In-Scope Features:

- **Database and Backend**: PostgreSQL schema, FastAPI endpoints, metadata management, video file management.

- **Frontend**: Up to date interface with advanced search/filtering, trajectory visualization, and intuitive design for non-technical researchers.

- **Data Processing**: Python based analysis tools for behavioral metrics, pattern detection algorithms, automated workflows, and an extensible architecture for future video-based analysis.

- **Deployment and Accessibility**: A fully deployed platform accessible to international researchers, with documentation, tutorials, and open-source analysis tools.

Out-of-Scope Features:

- Direct integration with external datasets outside the OCD rat data repository.

- Real-time data collection from new experiments (focus is on analysis of existing dataset).

- Advanced machine learning video recognition (future extensibility only, not initial deliverable).

This boundary ensures the platform delivers a usable, scalable, and research-ready tool while avoiding scope creep into areas requiring separate infrastructure or regulatory compliance.

8.2   Product Use Case Table

Use Case	Actors	Description	Preconditions	Outcome
UC1: Search & Filter Data	Researcher	Query dataset	User logged in; dataset available	Filtered dataset returned
UC2: Download Data	Researcher	Download trials	Dataset query completed	Files downloaded
UC4: Deployment & Access	DevOps Engineer	Deploy platform online or locally	Backend/frontend configured	Platform accessible

Table 1: Product Use Case Table

8.3   Individual Product Use Cases (PUC's)

PUC1: Search and Filter Data The researcher searches the dataset using filters such as trial number, injection count, or behavioral pattern. The system returns relevant subsets of data.

PUC2: Download Data The researcher downloads selected subsets of trials, metadata, or video files for offline analysis.

PUC3: System Deployment and Access The DevOps engineer deploys the platform to ensure that researchers can access it globally, with necessary backend and frontend services running reliably.

9 Functional Requirements

9.1 Functional Requirements

Insert your content here.

10 Look and Feel Requirements

10.1 Appearance Requirements

- The system shall resemble a webstore or online boutique in which possibly useful queries are packaged and presented for selection to guide a non-technical user audience.

- The system shall not overwhelm the user with too many or overly complex visible features as to not intimidate a non-technical user audience.

- The system shall present the filtering and searching capabilities of the system in a way that is familiar and intuitive, such as the filtering provided in an online shopping storefront.

10.2 Style Requirements

Insert your content here.

11 Usability and Humanity Requirements

11.1 Ease of Use Requirements

- The system shall have a very low ease of use floor. This floor being scrolling through pre-packaged query options and clicking on the one most relevant to their needs.

- The system shall have a reasonably low ease of use ceiling with intuitive an intuitive filtering interface as well as a simple natural language search interface and provided options for generating metrics and visualizations.

11.2 Personalization and Internationalization Requirements

Insert your content here.

11.3 Learning Requirements

- Assuming that the user is familiar with nature of the data itself, the system should have a learning curve of less than 30 minutes. Users shall easily be able to find and use searching mechanisms and querying pre-packaged and custom searches shall be very intuitive.

- The system shall provide options to generate available metrics or visualizations based on the queried data so the user does not need to learn to create their own.

11.4 Understandability and Politeness Requirements

Insert your content here.

11.5 Accessibility Requirements

Insert your content here.

12 Performance Requirements

12.1 Speed and Latency Requirements

- The system shall return all query results within 2 seconds for result sets of up to 5,000 records.

- All network requests shall have latency below 100 ms under normal operating conditions.

12.2 Safety-Critical Requirements

- There are no safety-critical operations for this system, as it handles only user interface and data processing with no direct impact on human safety.

12.3 Precision or Accuracy Requirements

- All query results shall be accurate and filtered according to user specifications.

- There shall be no inconsistencies between the source database and query results.

12.4 Robustness or Fault-Tolerance Requirements

- The system shall log and report all input errors without crashing.

- The backend shall handle errors gracefully, storing detailed logs for debugging and monitoring purposes.

12.5 Capacity Requirements

- The system shall support up to 250 concurrent users and handle 5,000 transactions per hour.

- The database and associated controllers shall be able to manage up to 11 TB of data.

12.6 Scalability or Extensibility Requirements

- The system shall allow additional modules to be integrated without requiring major changes to existing features.

- Under heavy user loads, the system shall maintain at least 80% of its performance efficiency.

12.7 Longevity Requirements

- The system shall be designed to operate reliably for at least 5 years, with multiple teams and developers able to maintain and extend it.

- The system shall be compatible with any operating system when running offline.

13 Operational and Environmental Requirements

13.1 Expected Physical Environment

- The system shall be able to run on a standard desktop with atleast 12 GB RAM and average processor.

- The software shall function correctly on Windows, macOS, and Linux operating systems.

- The software should be able to run in normal office environment conditions which includes temperatures between 15°C and 30°C.

13.2 Wider Environment Requirements

- The system shall be able to run the top three popular browers.

13.3 Requirements for Interfacing with Adjacent Systems

- Data exchanged with adjacent systems shall use JSON format and be transmitted over HTTPS.

13.4 Productization Requirements

- The system will use Docker and Kubernetes as it's envoirment for assebility to run on all systems.

- The system will provide all code changes and software updates on the GitHub Repository.

13.5 Release Requirements

- The system releases shall follow the versioning defined by MAJOR#.MINOR#.PATCH#.

- The system will provide all code changes and software updates on the GitHub Repository. Code then can be pulled and tagged from the appraite GitHub release.

14 Maintainability and Support Requirements

14.1 Maintenance Requirements

Insert your content here.

14.2 Supportability Requirements

Insert your content here.

14.3 Adaptability Requirements

Insert your content here.

15 Security Requirements

15.1 Access Requirements

There are no access requirements for our system, since it will be publicly accessible.

15.2 Integrity Requirements

The product shall prevent itself, the databases, and other files within it from intentional abuse.

15.3 Privacy Requirements

The system does not collect or use any user-related data that could be considered private and/or confidential.

The system does not store any personal or sensitive personal data, it only stores and presents the data which is already publicly available through FRDR.

There are no privacy requirements for the system.

15.4 Audit Requirements

There are no audit requirements for the system. There are no data changes or transactions involved on the user's end and they only view already publicly available data, therefore no auditing is required.

15.5 Immunity Requirements

Insert your content here.

16 Cultural Requirements

16.1 Cultural Requirements

There are no cultural requirements for the system.

17 Compliance Requirements

17.1 Legal Requirements

There are no legal requirements to be considered for the development of this project. The project does not involve the storage or use of user data, it only offers the searching and presentation of an existing public data set.

We have consent from our supervisors, Dr. Henry Szechtman and Dr. Anna Dvorkin-Gheva to use their data set in the development of the project. Additionally, we are allowed to reference the work of the capstone group, ratbat.mcmaster.ca, which they supervised during the previous year.

17.2 Standards Compliance Requirements

Insert your content here.

18 Open Issues

Insert your content here.

19 Off-the-Shelf Solutions

19.1 Ready-Made Products

Several open source or commercially available products provide out of the box functionality relevant to the project. These include:

- Databases

  Products like PostgreSQL, MongoDB, and AWS RDS are database offerings that provide robust data storage and indexing capabilities. These products meet the needs for data management at scale, but may require customization to fit the project's use.

- Artificial Intelligence

  Google's Vertex AI and the OpenAI API offer a suite of products pertaining to natural language processing and image analysis. These may be effective tools for the natural language and data processing tools central to the project, but require integration with a third party API external to the system. Specialized tools like DeepLabCut offer opensource tools for animal behavioral analysis.

19.2 Reusable Components

Dr. Henry Szechtman and Dr. Anna Dvorkin-Gheva possess several preexisting software modules that can be incorporated into the system with minimal modification. This suite of tools exists as a system called RatBat: a previous effort to implement an interactive data platform for the OCD rodent dataset. These modules include components for interfacing with the existing dataset, as well as tools for querying the dataset. The existing database includes an extensive set of trials, outcomes, and metadata comprising a core reusable project component.

19.3 Products That Can Be Copied

The client has provided several resources to serve as references for the implementation of the system. One similar interface is the GDC Data Portal which provides a querying interface that may be replicated by the system.

Furthermore, the client expressed a goal of incorporating a user-friendly, e-commerce-like catalogue interface, citing retailers such as Amazon as a reference for user interface design.

20 New Problems

20.1 Effects on the Current Environment

The new system will be developed from the ground up, there is no current implementation environment to affect. The only existing system that will be interacted with is the FRDR repository which provides a stable API for accessing the relevant data sets that will be used and will thus be unaffected by any developments made.

20.2 Effects on the Installed Systems

The 'installed system' that concerns this project is the FRDR repository as it will be used as a part of the system to provide the required functionality. The interactions between the new system and the FRDR repository will all be through a REST API that the FRDR provides. As a result no effects or conflicts are anticipated, although this conclusion operates under the assumption that the provided API is sound.

20.3 Potential User Problems

An important note to make about the project is that it is not a mandatory replacement for the existing data set and infrastructure available through FRDR. Users besides the commissioners of our project, our supervisors, can simply keep using the existing system if that better suits their needs.

The comissioners of our project, and external users of the project due to its public availability, could have a potential issue with the Natural Language Processing (NLP) used in querying raw data files. Since natural language will be converted to a query statement using an AI model, it cannot be guaranteed the model will be accurate 100% of the time when creating these queries from natural language.

20.4 Limitations in the Anticipated Implementation Environment That May Inhibit the New Product

Insert your content here.

20.5 Follow-Up Problems

Insert your content here.

21 Tasks

21.1 Project Planning

Insert your content here.

21.2 Planning of the Development Phases

Insert your content here.

22 Migration to the New Product

22.1 Requirements for Migration to the New Product

There are no requirements for migration to the new product other than that the user must familiarize themselves with the new domain and UI if they are used to using the FRDR repository.

22.2 Data That Has to be Modified or Translated for the New System

The new system will take the raw data directly from the FRDR repository and present it to users. There are use cases in which the data will be processed for the purposes of producing metrics or visualizations which is relevant to the functional requirements rather than this section. Thus, no data needs to be modified or translated for this new system.

23 Costs

There are no costs associated with the development of this project.

24 User Documentation and Training

24.1 User Documentation Requirements

- The system shall be released alongside a user manual guiding users through the possible inputs they have access to, how to properly execute them as well as the corresponding outputs that will be produced as a result. The inputs that will be over: querying pre-packaged result sets, generating custom filters, natural language searching, producing behaviour metrics, producing data visualizations.

24.2 Training Requirements

There are no explicit training requirements for this system. The interface should be sufficiently intuitive and the user manual must be interpretable by a non-technical user if they need to learn specific functions. No explicit training will be needed

25 Waiting Room

Insert your content here.

26 Ideas for Solution

Insert your content here.

Appendix — Reflection

The purpose of reflection questions is to give you a chance to assess your own learning and that of your group as a whole, and to find ways to improve in the future. Reflection is an important part of the learning process. Reflection is also an essential component of a successful software development process.

Reflections are most interesting and useful when they're honest, even if the stories they tell are imperfect. You will be marked based on your depth of thought and analysis, and not based on the content of the reflections themselves. Thus, for full marks we encourage you to answer openly and honestly and to avoid simply writing "what you think the evaluator wants to hear."

Please answer the following questions. Some questions can be answered on the team level, but where appropriate, each team member should write their own response:

1. What went well while writing this deliverable?

2. What pain points did you experience during this deliverable, and how did you resolve them?

3. How many of your requirements were inspired by speaking to your client(s) or their proxies (e.g. your peers, stakeholders, potential users)?

4. Which of the courses you have taken, or are currently taking, will help your team to be successful with your capstone project.

5. What knowledge and skills will the team collectively need to acquire to successfully complete this capstone project? Examples of possible knowledge to acquire include domain specific knowledge from the domain of your application, or software engineering knowledge, mechatronics knowledge or computer science knowledge. Skills may be related to technology, or writing, or presentation, or team management, etc. You should look to identify at least one item for each team member.

6. For each of the knowledge areas and skills identified in the previous question, what are at least two approaches to acquiring the knowledge or mastering the skill? Of the identified approaches, which will each team member pursue, and why did they make this choice?