# Abstract

With the internet userbase constantly growing, privacy becomes more and more of an issue as user activities become tightly linked to virtual identities which many users pay little to no attention to when browsing online.

This project provides a solution to this issue by focusing on the idea of compartmentalising by separating these identities by different uses while still maintaining the core aspects of a password manager by unifying some of their features.

The development of this project has undergone multiple phases where a methodology was chosen and followed in order to deliver a final product, this includes:

* Background research
* Design
* Implementation
* Testing and evaluation

Research was conducted on existing password managers to obtain their shortcomings in order to form a specification of requirements as a way of improving them, and with these requirements, designs were created to display a clear structure of the application as well as how it should appear and operate where interfaces were designed and algorithms such as health score calculation and encryption/decryption were created. These designs were then later implemented to create the system along with any amended changes to the requirements where later on this implementation was tested using unit testing to finalise and ensure that the system work as expected.

Once the testing was completed, the system was evaluated using the results from the following:

* Unit tests – All passed
* Code quality metrics – Optimal values
* User feedback – Positive views
* Competitor comparisons – Unified and improved features

All of these evaluation methods provided positive results, thus affirming that the system is successful, and its aim was met.

This report discusses and documents all of these phases in detail where both the process and reflection of developing this system is shown from start to end.

# Introduction

The internet userbase is constantly growing, and according to a report (Kemp, 2020), for the first time, more than half of the world’s total population is now using social media, which is 3.96 billion people, which is 51% of the global population. This shows the cruciality of internet in our day-to-day life as it has become an ecosystem that is unavoidable, you can store and access all kinds of data, especially what you’d call sensitive, and so data has become the digital form of gold, this, however, brings attention to security problems/cyber threats due to the value of data.

­­­­­­­

While your data can secured behind a password, there’s very little that can be done if the database that the data is contained is breached, which then it is the company’s responsibility in hope that the data is encrypted and secured well, in many cases, it isn’t; a perfect example of this is where millions of Facebook records were exposed on a public Amazon server, and, in another incident, an unsecured database exposed the information of 80 million US households (Webster, 2020). Furthermore, a survey conducted by LastPass for 3,250 respondents (LastPass, 2020) reveal that 66% of them would use the same password or a variation of it for accounts, and 48% would only change these passwords if required, the same survey also reveals that an average user has around 38 online accounts, this would mean that they would use the same password in 38 different accounts which if one of those accounts are compromised, the others can be considered compromised too; It all narrows down to the difficulty of having to remember different passwords for different accounts, especially if the passwords are too complex (Rawlings, 2020).

The current most popular solution is to use a password manager, which does these rigorous tasks with a few clicks, the only issue is that one must again trust the company to store all these credentials securely, especially if they’re stored online for synchronisation and while they endorse complete safety and security for your data, yet there are instances in which they get hacked through a vulnerability, such as the LastPass hack in 2015 (LastPass, 2015), and so another solution is a local password manager like KeePass which only allows offline access. However, disregarding all of the above, it is pointless to have a strong and unique password for every account if every other data attribute of those accounts remains the same across different accounts; when an account shares the similar or same attributes as another account in two separate websites, in which both of the sites have their database hacked, not only are these accounts compromised but they can also be linked together regardless of how strong their passwords were, which causes the user to be more vulnerable to further threats such as identity theft.

This is the problem these password managers do not tackle as they narrow the user into using strong passwords behind a paywall but not do not teach them anything about bad browsing habits nor how to recover from a data breach or to minimise damage from them.

To truly secure your privacy online, one must build a privacy model and compartmentalise different accounts into different personas, which means that for an example a person who works should not reuse the same username, email, password, and other relevant data in other areas that aren’t related to their work as a simple search of a username could reveal a lot about a user’s online habits; a common saying is to never post something on social media that could otherwise harm your chance in obtaining a job, but it may not necessarily be relevant to work and thus they should’ve never mixed in the first place. Therefore it’s essential to create virtual identities to solve these issues and this also solves the issue in the case of a data breach as it acts as one of the layers that’s breached rather than the person’s real information.

## Aims and Objectives

The project’s aim is to create a system that unifies some of the essential security and privacy features needed to provide the user some assurance against future browsing issues and threats such as compromised accounts and publicised personal information through minimising the amount of the user’s data breached by not only allowing the user to save passwords but also build a privacy model creator and compartmentalise the credentials based on the identities created by it. It is aimed to be very straight forward to use and offer more than the average password manager with these features.

The objectives to achieve this aim are the following:

* Gather enough knowledge to ensure that the system is fit for its purpose, which includes examining existing software such as password managers that provides a partial solution to these problems to make the system better by outlining their benefits and drawbacks and what could be done to compensate for these drawbacks, which may require researching alternative libraries and APIs to use to solve these issues.
* Using the conducted research, form the specification of the system by defining the user requirements to display what the purpose and the capability of the final product should be, this will be split into two types of requirements: functional and non-functional, where the non-functional requirements may be shifted towards the usability of the system.
* Based on the approximate length as well as the difficulty of completing these tasks, choose and follow a suitable methodology so that the project can be carried out in an structured manner and have it delivered before the deadline so that there is also enough time for improvising.
* Design several interfaces before an actual working program is built, a key aim of the system is to make it as easy and straight forward to use, and so it will be crucial to put just as much effort into the interface through usability engineering. This follows up from the requirements as these prototypes will ensure that these requirements are/can be met, concerning this, several diagrams such as flowcharts will be made to show the control flow of the program, which will give an idea of how it will operate.
* Once the designing phase is completed, using the created designs, implement the system starting from the logic as the focus with just a functional but incomplete interface, meaning that the front end is expected to be completed last due to the anticipated complex algorithms in the back end. This will make use of the gathered libraries and APIs considered from the prior research where for an example, an alternative method of encryption may be used to store the credentials and another library could be used to tailor these credentials to their respective identities made by the user.
* Test the system with several different methods to ensure that the system will be fully functional and meets all the specified requirements as well as the aim of the project, this is a continuous process and should be documented in some way to ensure the found errors are corrected and prevented from being repeated. This should also include test cases on the encryption as well as the efficiency, which will involve automated testing so that on each build, the integrity and performance remains or improves, the test data for the credentials should be completely fake and therefore no personal information will be used.
* Evaluate the completed system in different perspectives through the test cases and use cases as well as user feedback and compare it to other solutions to check whether it provides as a solution to the previously mentioned privacy/security issues to ensure that it is a success, afterward, document what was done well and what could’ve been improved or done differently for any similar future projects. This is mainly in relation to what may separate this password manager from its competitors where the system will be evaluated to see if it assists in compartmentalising online identities and enables the user to recover from setbacks such as data breaches.

## Project Approach

To begin, secondary research will be conducted to ensure that the system will be built for its purpose to tackle the existing security/privacy issues that similar systems may not be able to, for this, these systems will be examined and evaluated to find their benefits and drawbacks. Moreover, considering it will involve the protection of data, there will be further research to ensure that it is also GDPR compliant.

Once this research is completed, a specification concluding the functional and the non-functional requirements of the system will be formed, this acts as a set of tasks to complete to accomplish the aim of the system. Afterwards, the design phase will begin where designs will be made for the interface tailored to these requirements, which then follows onto the design of the logic of the system where various diagrams will be created to show the flow of the system using flowcharts and other similar notations, all of which are planned to be produced using tools such as “draw.io”.

Ideally, an agile approach is most suitable for this project, and so the currently considered software development process is Scrum due to the expected influx of changes needed to be made throughout the project and thus requiring flexibility, and therefore choosing a fixed approach such as the waterfall model wouldn’t be suitable.

This approach will be carried out over the course of multiple sprints where a backlog for each of those sprints is produced on the sprint before it, these sprints will be produced weekly or bi-weekly to keep the project in constant development. To make the best out of this approach, an application known as “Trello” will be used to store and track the progress where all the GitHub repository commits linked to the tasks will be attached and displayed.

The virtual identities simply serve the purpose of compartmentalising the user’s credentials which can either be imported or generated based on the identity attributes, and so it will be designed to show the vulnerabilities in those identities (through weak or breached credentials, duplicates, etc.) and how to mitigate them, these will all be presented clearly and visualised using graphs and statistics.

It is also planned to include other features such as a web crawler to search for existing public information given the identities or the credentials as well as data breach checkers to check to see if those same credentials are leaked online.

The considered method of encrypting these credentials is “AES” given its credibility from its usage in government agencies for protecting top secret data, and so it’s intended to use the master password as the key to encrypt these credentials into a local file, however, a double encryption is also considered where the master password itself is encrypted and instead its salted hash will be used for the encryption/decryption key instead of the actual master password for a second layer of security; for this, the “bcrypt” method is ideal due to its work factor which causes it to hash the credentials slower, hence increasing the time it takes to crack compared to other popular methods. To accomplish implementing these features, several libraries and APIs will be used, all of which will be gathered in the conducted research earlier, the implementation is planned to be carried out using C# due to the vast range of compatible libraries, however, Rust, is in consideration to be used due to its focus on performance and memory safety.

Lastly, there will be extensive testing done not only towards the end of completion of the system but also throughout the process of the development, this will make use of development techniques such as unit testing. The main objective of this is to ensure that the system is fully functional and the possibility of a vulnerability that can exploit the system is minimised as much as possible, and so platforms such as “CircleCI” will be considered so that much of this is automated (including performance/efficiency) and clearly presented. The test data for creating identities and credentials and testing their encryption is intended to be filled with fake data in CSV or JSON files generated using sites such as “fakenamegenerator” and “Mockaroo”, this is to eliminate time consumption of manually creating data or gathering and using personal information, these data may have to be pre-processed so that they are compatible with the system. These test data will be used by importing them into the system and running test cases on them to ensure that they are encrypted well and to show the outcome of when some of these credentials are compromised, which is merely imitated for the purpose of the test case, which will result in changes to the displayed statistics to show that the system was successful in these tests.

## Specialism

The reason this project and its approach was chosen is due to its strong relation to the software engineering specialism as this system will be built from scratch following a development cycle with extensive research and planning which will require time management and analytical skills along with designing and implementing. Due to this, completing this project will give an insight to how major software are built in the industry through the knowledge gained and the tools and techniques used throughout this project.

## Research ethics

The user feedback for the evaluation of the system at the end will involve human participants, and therefore this project requires ethical approval, this was obtained before any feedback is collected as shown in Figure 8‑1 where a letter was received displaying the approval for research within the given period to complete the project alongside some conditions. The data collection will not gather any personal information as it will be completely anonymised and stored securely on an encrypted drive.

## Dissertation Outline

As the dissertation is expected to contain a lot of information, it’s tedious to try and briefly describe the structure of it, and so a flowchart was created displaying the flow and the inputs/outputs of each chapter/section as shown below:

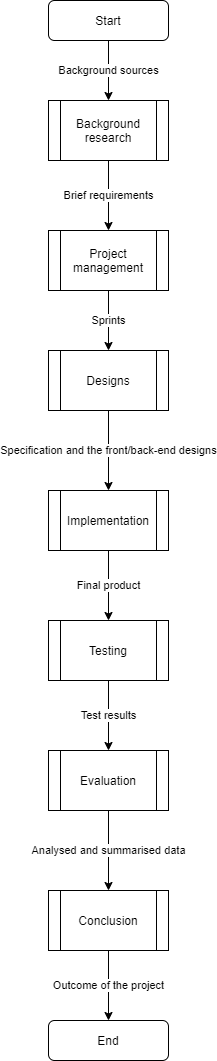


Figure 1‑1 Flow of the dissertation

One thing to note is that the testing and the evaluation will both be in the same chapter although they take different inputs and thus, they were shown separately. The order of the dissertation will be the same as shown above so that all information is coherent and easy to follow.