**C868 – Software Capstone Project Summary**

**Task 2 – Section A**



|  |  |
| --- | --- |
| **Capstone Proposal Project Name:** | Dossier Management System |
| **Student Name:** | Phuc Thach |

**Table of Contents**

Contents

[**Table of Contents** 2](#_Toc88002685)

[**Business Problem** 3](#_Toc88002686)

[The Customer 3](#_Toc88002687)

[Business Case 3](#_Toc88002688)

[Fulfillment 4](#_Toc88002689)

[**Existing Gaps** 5](#_Toc88002690)

[Traditional Communication 6](#_Toc88002691)

[Computer-Mediated Communication 6](#_Toc88002692)

[**SDLC Methodology** 6](#_Toc88002693)

[**Milestones** 7](#_Toc88002694)

[Milestone Alpha / Backend System 7](#_Toc88002695)

[Milestone Beta / Mobile Application 7](#_Toc88002696)

[Milestone Gamma / Web Application 7](#_Toc88002697)

[**Implementation** 8](#_Toc88002698)

[**Validation and Verification** 9](#_Toc88002699)

[**Environments and Costs** 9](#_Toc88002700)

[Programming Environment 9](#_Toc88002701)

[Environment Costs 10](#_Toc88002702)

[Human Resource Requirements 10](#_Toc88002703)

[**Project Timeline** 11](#_Toc88002704)

[**References** 12](#_Toc88002705)

# **Business Problem**

**The Customer**

Corruption is a complex social problem causing immense damage to individuals and societies; freedom and the rule of law, downgraded individual trust in their government, environmental impact such as deforestation, and economic cost for victims to build and grow wealth. According to the World Economic Forum, the global cost of corruption is at least 5% of the World’s Gross Domestic Product (United Nations, 2018). In Asia, nearly 1 in 5 who used a public service paid a bride in the previous 12 months, equivalent to 836 million people in the region (Transparency International, 2020). Transparency International, an organization specializing in fighting corruption, proposes that one truth element to fighting corruption is for societies to embrace transparency. Societies must be made aware of formal and informal rules, plans, processes, and actions involving everyone and everywhere, politicians, government officials, public service, businesspeople, or members of the public. Corruptions are staggering high, and it is unthinkable how perpetrators can get away with so many evil deeds.

Corruptions are rampages and are the crown jewel of top Cambodian leadership. In January 2012, the Cambodian National Assembly passed the country’s first law on public procurement to fight endemic corruption in the public sector. In 2016, Global Witness published a report in which the Prime Minister of Cambodia and his family members control almost every aspect of Cambodia’s economy through their registered private domestic companies. The Sen family’s business enterprises link to major international brands, including Apple, Nokia, Visa, Procter & Gamble, Nestlé, and Honda. The report exposes the Sen family to be worth $500M - $1B, despite the prime minister’s claim to only earn $1,140 per month (Global Witness, 2016). Each day there are hundreds of thousands of cases of corruption behaviors going unnoticed; public servants taking money or favors in exchanges for services, politicians pocketing public funds, and corporations bribing officials to get lucrative deals. Corruptions occurred at the highest level of the country’s judicial and political dynamic.

The proposed Dossier Management System must achieve these two objectives; a.) allow activists to report their struggles and findings (dossiers) to Transparency International or similar non-government organizations. b.) Enable analysts of NGOs to pull and review dossiers and publish their results for public viewership. The long-term goal of the system is to give the public better information and thus contribute to the rationalization of society, “resulting in more democratic and more affluent societies (Meijer, 2009) “.

## **Business Case**

Most victims turned activists initially do not report corruption cases at all. Victims are often trapped in a harsh reality; if they report corruption crimes, they may get caught, activists pay heavy consequences because of their heroic actions. Or they do nothing, hear nothing, and continue to be victims. Dr. Kem Ley, a Cambodian activist, physician, and political commentator, known for his political commentary, including trenchant criticism of the Sen family, was assassinated on July 10, 2016, at a petrol station in Phnom Penh. He was the third notable activist to be killed after union leader Chea Vichea in 2004 and environmental activist Chut Wutty in 2012 (Wikipedia contributors, 2021). The Dossier Management System must benefit anti-corruption activists by allowing them to report their findings safely. There must exist a kind of anonymity mechanism protecting activists from potential backlash from perpetrators. Safety for activists is a top priority.

The Dossier Management System is a centralized system consisting of four subsystems: the mobile application, VPN connection, back-end system, and the dossier management web application.

1. The mobile application is the activist’s swift knife. It enables activists to send dossiers and report corruption-related crimes to non-governmental organizations (NGOs). It is shipped with a virtual private network connection (VPN) profile enabled by default. The VPN profile protects activists from being snooped by untrustworthy internet connections, gains anonymity by hiding the activists’ actual location, and prevents ISP from logging the activists’ activities, especially while submitting dossiers.
2. The VPN system enables secure data transfer between the mobile application and the non-governmental organization network.
3. The back-end system consists of the application interface (API). The mobile front-end application and the case management web interface communicate with this system to send and retrieve dossiers. The back-end system stored its data in an open-source document database, MongoDB.
4. The dossier management web application enables analysts to download and view dossiers coming from activists.

Essential features of the Dossier Management System are unidirectional communication flow and anonymity; both activists and analysts are protected. Users log into the system with mnemonical keywords and PINs. It does not store personal data points such as names, phone numbers, email addresses, or postal addresses. Activists and analysts do not engage in direct communication as in no correspondence. When analysts or activists turn rogue, their associates’ identities remain safe and protected in the worst case.

## **Fulfillment**

The Dossier Management System consists of four subsystems: the mobile application, VPN system, back-end system, and the dossier management web application.

1. The mobile application is built with the React Native Framework. It also acts as a VPN client and establishes a secure connection between the activist and the organization’s virtual private network. Activists enter the secured network connection by clicking on the Connect button when the application is first launched. Once the connection is established, the activist may proceed to submit dossiers via the Report Case Form. Once dossiers have been submitted, the application will disconnect and take the user back to the login screen. The application will always return success messages, regardless of server errors. The confusing return behavior is a safety feature to protect the activists and a preventive deterrence related to dictionary attacks against the back-end system.
2. The VPN system utilized open-source VPN technology called WireGuard. According to Donenfeld (2017), WireGuard is a modern VPN that uses state-of-the-art cryptography. It aims to be faster, simpler, leaner, and more practical than IPsec while avoiding the massive headache. It intends to be considerably more performant than OpenVPN. “WireGuard is designed as a general-purpose VPN for running on embedded interfaces and super computers alike, fitting many different circumstances” (Donenfeld, J. A, 2017). Administrators of NGOs will be able to set up the VPN Server simply by running a few Linux commands. NGOs could also contract trusted server hosting facilities specializing in WireGuard to assist with the deployment.
3. The back-end system utilized the open-source Parse Platform, formally a product of Facebook, Inc. The Parse application stack includes storage, user authentication, push notification modules out of the box. These modules allow software engineers to develop features faster because they do not need to be concerned with low-level designs, specifically processing and storing data. Instead, developers can be guaranteed that the data are processed structurally and persist in MongoDB, the database system of choice. Deployment of the Parse application stack is as simple as running a few command lines. The back-end system does not expose itself in the public network; instead, both users (activists and analysts) must first log into the organization network via VPN Protocol to consume the back-end system API endpoints.
4. The dossier management web application is built using the React-Admin Framework. It is a front-end framework for building data-driven applications running in the browser, on top of REST/GraphQL APIs, using React Framework and Material Design. It is open-sourced and maintained by the Marmelab consulting firm located in France. The dossier management web application has the following features:  
   * Search functionality allows multiple row results and displays, allowing analysts to search dossiers by keywords.
   * Reviews functionality allows analysts to add, modify, and delete reviews associated with a dossier including audit logging.
   * Report functionality allows analysts to generate columns, multiple rows, date-time stamps, and title reports.

# **Existing Gaps**

There are many general purpose communication systems for activists to submit dossiers to their intended organizations. Those existing systems could be categorized into traditional communication, which is direct/indirect personal contact between activists and analysts, and computer-mediated communication, which is using computerized technology as the mediator for that communication. The Dossier Management System does not intend to replace those methodologies; instead, it serves as a complementary tool to both methods with the added benefits of additional data transfer security and personnel anonymity.

## **Traditional Communication**

Traditional communication is an information exchange that takes outside the cooperative collection of inter-networks (Doerfel et al., 2003). The technique includes these communication channels; direct person-2-person contact, where both participants meet at a secret location to exchange dossier documents; indirect person-2-person contact, where participants agreed to a private place for dossiers to be dropped off and picked up; analog telecommunication, where both participants communicate through land-line telephone communication; snail mail, where both participants exchange information through the postal service system. This mode of communication functions at a rudimentary level; both participants risk exposing their identities. For example, activists could be working for preparators to harm the analyst researching the topic matter.

The Dossier Management System could be a robust complementary tool in facilitating exchanging dossiers between activists and analysts in a secure manner. Activists could upload their dossiers from wherever and whenever while reducing their probabilities of being exposed. Analysts could retrieve those dossiers quickly thru the dossier management web application. Both participants achieve dossiers exchange without having to put their identities at risk.

## **Computer-Mediated Communication**

Computer-mediated communication is an “information exchange that takes place on the global, cooperative collection of networks…. The resulting information exchange can involve a wide range of symbols people use for communication” (Doerfel et all., 2003). Computer-mediated communication is software programs that utilizes the internet connection. Some examples are online message boards, chat rooms, messaging platforms, email platforms, customer relationship platforms, and social media platforms. While these channels provide avenues for information exchange, they generally are not built for transferring and managing of sensitive information like how one would not want to eat soup with a fork.

The Dossier Management System (DMS) is a form of computer-mediated communication; however, it is built to exchange dossiers and other sensitive data. DMS includes features that other forms of computer-mediated communication may not support: anonymous features, available to search and filter dossiers, built-in VPN connection. DMS is built with the interests of securing the personal identities of activists and analysts from harm.

# **SDLC Methodology**

The SDLC methodology utilized in this project is Agile. The Dossier Management System is a specific form of computer-mediated communication for a particular set of users, activists, and analysts. It needs to take full advantage of fundamental Agile techniques to achieve its intended goal: frequent customer communication, incremental development, and focus on quality.

* Customer Communication – A primary contact from the development team will communicate directly with the two types of customers, activists and analysts, enabling everyone else free to perform their other duties. The development team will receive immediate feedback on newly released features, allowing better, more helpful products.
* Incremental Development – The project is fast, iterative, and incremental. The iterations are relatively short, with a duration of a couple of weeks. Each iteration incorporates every development step: requirement gathering, high-level design, low-level design, development, testing, deployment, and maintenance.
* Focus on Quality – Due to the fast iteration cycle, developers do not have time to spend chasing down bugs; therefore, extra efforts are put to write good, solid code rather than rushing through the programming.

# **Milestones**

The Agile Methodology does not have the ideal of deliverables due to its nature of having fast iterations. It applies continuous integration and continuous deployment to its development cycle, allowing customers to see and test newly built features in a short number of times. Milestone is a term used in agile methodology to denote a specific point used in the project that marks a significant stage of the development. The team has determined these three deployment milestones: back-end system, mobile application, and web application.

## **Milestone Alpha / Backend System**

The project has reached this milestone by completing and deploying the following tasks associated with back-end services.

* Set up and configured Parse Application Stack with MongoDB database system. Developers must be able to make GraphQL API calls to the back-end system.
* Set up and configured VPN Server using the open-sourced VPN technology, WireGuard. Users must be able to log in to the private network using the WireGuard Client application.

## **Milestone Beta / Mobile Application**

The project has reached this milestone if it has completed and deployed the following tasks associated with the mobile application, where activists submit dossiers.

* Users must connect securely to the organization’s virtual private network.
* Users must report classified dossiers to the organization by completing the Submit Dossiers Form and pressing the Submit button.

## **Milestone Gamma / Web Application**

The project has reached this milestone by completing and deploying the following tasks associated with the web application, where analysts pull and read dossiers.

* Users must be search, read, and view assets associated with each dossier.
* Users must be able to add, modify, and delete notes with audit logs.
* Users must generate reports for building metrics and other actions such as providing oversights to the public.

# **Implementation**

The implementation of this application is rudimentary. It is a new system; therefore, no outages are necessary. Deployment to production can be staged before the customer communicates with the user base. If significant, unexpected problems occur, there is not a need to roll back any changes. The team will try redeploying again later.

Implementation coordinate with project management and involves several key members of the development team, including a member in site reliability organization. The implementation team consists of the following members:

* Delivery Manager – overseer the entire deployment process
* Technical Lead & Software Engineers – on standby in case of technical problems during deployment
* Site Reliability Engineer – execute deployment process
* Acceptance Testers – run acceptance tests to confirm all features function and meet the customers’ requirements.

Implementation steps are as follow:

1. In the production server, the Site Reliability Engineer pulls the latest codes from the project git repository. The Site Reliability Engineer will need permissions to the project repository.

git clone git@github.com:thachp/dossa.git

1. Site Reliability Engineer navigate into the root directory of the project and run the command:

docker compose up

The project directory includes the DockerFile used by Docker Compose to define and run multi-container Docker applications. With Compose, Site Reliability Engineer uses a YAML file to configure the application’s services. Then, the engineer creates and starts all system services from the configuration file with a single command.

Once Docker Compose complete executing the command, all services should be running on the following ports:

* + - [udp://0.0.0.0:51820](http://0.0.0.0:51820) – WireGuard VPN Server
    - [mongodb://0.0.0.0:27017](http://0.0.0.0:27017) – MongoDB Database Server
    - <http://0.0.0.0:1337> – Backend API Service running on Parse Application Stack
    - <http://0.0.0.0:3000> – Dossier Management Web Application

1. Acceptance Testers proceed with their tests with automation testing or manual testing.

# **Validation and Verification**

Ideally, the team sits down, writes codes that perfectly satisfy the requirement, and is done with the project. Unfortunately, it is not the case in the software development industry, but a façade and an unrealistic realization. Bugs are bound to happen.

Testing will be a comprehensive full lifecycle test to ensure that the application has met the requirements as designed. The project applies the following testing techniques: unit testing, automated testing, system testing, and acceptance testing.

* Unit testing – Developers verify the correctness of a specific piece of code. It is the first chance to catch bugs. Developers must commit to writing unit tests for all methods and classes they contribute to the project.
* Automated Testing – Continuous integration is set up against the project repository. Automatic testing runs unit tests against the front-end and back codebase each time developer commits new changes to the trunk branch. Load testing is a part of automated testing, where a testing tool simulates many users, all running simultaneously to measure performance.
* System Testing – SDET (Software Development Engineer in Test), where end-to-end testing is done against the entire system. SDET may use a testing framework such as Selenium to automate UI testing.
* Acceptance Testing – Customer Representatives sit down with the application and run through all the user cases identified during the requirements gathering phase to ensure everything works.

In addition, software developers also incorporate software development principles such as the SOLID principles, Command and Query Responsibility Segregation (CQRS) pattern, and code review processes between code submissions. The team hopes to limit bugs in the production environment by following software development principles and applying various testing techniques.

# **Environments and Costs**

## **Programming Environment**

The Dossier Management System is developed with open-source technologies and libraries. Open-source libraries are projects in which the copyright holder grants users the rights to use, study, change and distribute the software and its source code to anyone and for any purpose.

Besides a decent desktop/laptop and Visual Studio Code, the programming IDE of choice for the development team, completing the project requires utilizing the following open-source projects.

* Visual Studio Code is a source-code editor made by Microsoft for Windows, Linux, and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git. (https://github.com/microsoft/vscode)
* Git tracks changes in any set of files, usually used for coordinating work among programmers collaboratively developing source code during software development. Its goals include speed, data integrity, and support for distributed, non-linear workflows. (https://github.com/git)
* Docker Platform - a set of platforms as a service product that uses OS-level virtualization to deliver software in packages called containers. Containers are isolated and bundle their software, libraries, and configuration files; they can communicate with each other through well-defined channels. (https://github.com/docker)
* NodeJS - an open-source, cross-platform, back-end JavaScript runtime environment that runs on the V8 engine and executes JavaScript code outside a web browser. (https://github.com/nodejs)
* WireGuard is a communication protocol and free and open-source software that implemented encrypted virtual private networks and was designed to ease use, high-speed performance, and low attack surface. (https://github.com/wireguard)
* Parse server - an open-source back-end that can be deployed to any infrastructure that can run Node.js. (https://github.com/parse-community)
* MongoDB - a source-available cross-platform document-oriented database program. Classified as a NoSQL database program, MongoDB uses JSON-like documents with optional schemas. (https://github.com/mongodb)
* Expo - an open-source platform for making universal native apps for Android, iOS, and the web with JavaScript and React. (https://github.com/expo)
* React Framework - a free and open-source front-end JavaScript library for building user interfaces or UI components. It is maintained by Meta and a community of individual developers and companies. React can be used as a base in the development of single-page or mobile applications. (https://github.com/facebook/react/)
* React Native Framework - an open-source UI software framework created by Meta Platforms, Inc. It is used to develop Android, Android TV, iOS, macOS, tvOS, Web, Windows, and UWP by enabling developers to use the React framework and native platform capabilities. (https://github.com/facebook/react-native)
* React Admin – a front-end Framework for building B2B applications running in the browser on top of REST/GraphQL APIs, using ES6, React, and Material Design. (https://github.com/marmelab/react-admin)
* GraphQL is an open-source data query and manipulation language for APIs and a runtime for fulfilling queries with existing data. GraphQL was developed internally by Facebook in 2012 before being publicly released in 2015. (https://github.com/graphql)

## **Environment Costs**

Dossier Management System used open technologies for most of its internal implementation, and therefore with the exceptions of servers, there are no other associated costs. The nominal fee for WireGuard Server is expected to be $165 a month. Hosting servers for storage and database server are expected to be $300 a month. The total environment costs are expected to be $465/month. However, the monthly fees are expected to increase or decrease depending on the user traffics.

## **Human Resource Requirements**

The project is projected to take six months to complete and needs seven human resources working full-time for the project’s duration. The larger share of human resources is on the engineering side of the project, followed by project management and user interface design. Engineering consumes approximately 70% of the associated hours and dollars, while project management and UI/UX consume 30%. Based on the industry salary for each resource type, the total development cost is estimated to be $600K. There may be unknown factors that may prevent the project from progressing; therefore, we include a 20% margin of error on top of the $600K for a total of $720K.

|  |  |  |  |
| --- | --- | --- | --- |
| Resources | Responsibilities | Capacity | Est. Costs |
| Product Manager | Project Management, User stories | Full Time | $88K |
| UI/UX Developer | Mock UI / UX | Full Time | $76K |
| Technical Lead | Architect, High-Level Design | Full Time | $100K |
| Back-end Developer | Low-Level Design, Back-end | Full Time | $90K |
| Front-end Developer | Low-Level Design, Front-end | Full Time | $85K |
| Software Developer in Tests | E2E Testing, Automated Testing | Full Time | $85K |
| Site Reliability Engineer | Deployment, Back-end Services | Full Time | $76K |
| Development Costs | | | $600K |
| 20% Margin of Error | | | $120K |
| Total | | | $720K |

# **Project Timeline**

The project is projected to take six months to complete. The large share of the projected timeline is on the project’s engineering side, followed by the pre-development phase requiring only two months. Back-end services development and deployment would require three months. Mobile and Web application development would require five months. Important to note that project development may work in parallel. For example, while project management gathers actual user requirements, engineering could work on their development environment, perform spike tests, and code core UI components. Throughout the development cycle, the team also engages with customers for immediate feedback.

|  |  |  |  |
| --- | --- | --- | --- |
| Milestone | Deliverable | Description | Dates |
| Pre-development | Requirements | Meet with activists and analysts, and perform procedure review | 6/1/2022 – 8/1/2022 |
| Milestone Alpha  Back-end Services | VPN Server, Parse Application Stack | Must allow users to VPN into the organization’s private network.  Must allow developers to create classes and objects. | 7/1/2022 –10/1/2022 |
| Milestone Beta  Mobile Application | Android Mobile  IOS Mobile | Mobile applications must work on both Android and IOS.  Must allow activists to submit dossiers. | 6/1/2022 – 11/15/2022 |
| Milestone Gamma  Web Application | Web Application | Must allow analysts to read and download dossiers. | 6/1/2022 – 11/15/2022 |

# **References**

Doerfel, M., Taylor, M., & Perry, D. (2003). Internet-Based Communication in Crisis Management. *Management Communication Quarterly*, *17*, 206–232. https://doi.org/10.1177/0893318903256227

Donenfeld, J.A. (2017, January). *WireGuard: Next Generation Kernel Network Tunnel*. Network and Distributed System Security Symposium. https://doi.org/10.14722/ndss.2017.23160

Donenfeld, J. A. (n.d.). *WireGuard: fast, modern, secure VPN tunnel*. WireGuard. Retrieved November 16, 2021, from https://www.wireguard.com/

Global Witness. (2016, July 7). *Hostile Takeover*. Retrieved November 15, 2021, from https://www.globalwitness.org/en/reports/hostile-takeover/

Human Rights Watch. (2019, October 21). *30 Years of Hun Sen*. Retrieved November 15, 2021, from https://www.hrw.org/report/2015/01/12/30-years-hun-sen/violence-repression-and-corruption-cambodia

Kaspersky. (2021, July 5). *What is VPN? How It Works, Types of VPN*. Www.Kaspersky.Com. Retrieved November 15, 2021, from https://www.kaspersky.com/resource-center/definitions/what-is-a-vpn

Meijer, A. (2009). Understanding modern transparency. *International Review of Administrative Sciences*, *75*(2), 255–269. https://doi.org/10.1177/0020852309104175

Transparency International. (2020, August 10). *What is corruption?* Transparency.Org. Retrieved November 15, 2021, from https://www.transparency.org/en/what-is-corruption

United Nations. (2018, September 10). *Global Cost of Corruption at Least 5 Percent of World Gross Domestic Product, Secretary-General Tells Security Council, Citing World Economic Forum Data | Meetings Coverage and Press Releases*. https://www.un.org/press/en/2018/sc13493.doc.htm

Wikipedia contributors. (2021, November 9). *Kem Ley*. Wikipedia. Retrieved November 15, 2021, from https://en.wikipedia.org/wiki/Kem\_Ley

Youtube. (2020, January 15). *WireGuard: Next Generation Secure Network Tunnel*. Retrieved November 15, 2021, from https://www.youtube.com/watch?v=88GyLoZbDNw