One Shared Story Capstone Report 1

Rachel Grace   
*MSDS Candidate*  
*University of Virginia*Charlottesville, VA  
rg5xm@virginia.edu  
  
Ami Kano  
*MSDS Candidate*  
*University of Virginia*Charlottesville, VA  
ak7ra@virginia.edu  
  
Chunru Zheng   
*MSDS, Economics PhD Candidate*  
*University of Virginia*Charlottesville, VA  
cz8yb@virginia.edu

*Abstract*—This progress report summarizes the topic of the One Shared Story capstone project, the milestones achieved, and intended future progress. While much of what has been accomplished is communication, the capstone team anticipates concrete achievements in the next weeks of the semester.

Keywords—Communication, Database, Dashboard, Data Model

# Introduction

This document describes the progress made on the One Shared Story capstone project. It defines the methodology and goal of the project, as well as a proposed schedule for the rest of the semester. The document provides an overview of the difficulties encountered and the solutions employed up until this point. Finally, it discusses client satisfaction.

# Goal

The concrete goal of this project is to create a functional, scalable, and searchable database for Robin Patton and One Shared Story (OSS), the nonprofit that she works for. The database must have a functional backend and be hosted reliably. Further, it must have a public dashboard application stored on the OSS website for searching the database to locate records, and a method to bulk upload spreadsheets of data.

## Sub-Goals

Identify applicable funding agency here. If none, delete this text box.

### Database

The data in the database has been restructured to meet the client’s goals. She asked that the data be restructured as a part of the project, not to add functionality to the database but so that future collaborations might be possible. This meant restructuring the data to honor the data model of Ancestry.com and On These Grounds (OTG). This also entailed changing the names of the data features so that search queries could be more easily written and the data model preserved.

Specifics of the data query have been established. The client requires the data to be searched by name, location, date, and record type. Additionally, the query must be flexible and capable of searching new datasets that are similar or identical in their features. This has been partially accomplished but is still being adjusted.

A final requirement of the database is detailed documentation. The client would like to be able to write new queries herself. This entails careful documentation of the query writing so that she can reproduce the work with some changes. This has not yet been done.

### Searching

The database must be searchable through a dashboard built with a free tool. Currently the team is exploring Starlette as a free dashboard option. The dashboard must have search fields like name, location, and event, and ideally would have capabilities to narrow or broaden the search. The output of the dashboard must be neat, ideally displaying results like the visualization on Ancestry.com. The query returns must provide a link to the hosted document image when available as a URL.

### Data Entry

The client must be able to upload data to the database herself. The team may create a dashboard so that she can do this, but another option is for her to directly upload the data through the MongoDB Atlas tool. The team has already demonstrated this capability for her. Either way, clear documentation must be written so that she can add data later. This has not yet been done.

### Hosting

The client and team must choose where to host the database. The client has stated that she prefers to host the database on her own website, but has recently indicated that she is open to keeping the database on the MongoDB AWS platform. The team has advocated for this solution. This decision still needs to be made.

The database also needs to be tested appropriately, and clear documentation must be created so that the client can maintain the database herself with no prior knowledge. This has not yet been done.

## Schedule

* 2/22/23 - [Database: Querying] Queries that return names in association with events, and are general enough to accommodate new datasets
* 4/1/23 - [Data Entry] A dashboard or direct upload pipeline
* 4/1/23 - [Searching] A dashboard made with a free tool, most likely Starlette
* 4/1/23 - [Searching] Appropriate search terms on search dashboard for returning genealogy records by name, event, and/or location
* 4/1/23 - [Searching] A neat and readable search result display
* 4/1/23 - [Data Entry] Working data entry code for adding an entire collection to the MongoDB database, or documentation for uploading a spreadsheet to MongoDB
* 5/3/23 - [Hosting] Appropriately tested and, when needed, fixed database
* 5/3/23 - [Hosting] Working database hosted appropriately
* 5/3/23 - [Searching] Edited and documented query function as needed to accommodate search dashboard
* 5/3/23 - [Database: Querying] A well-documented query function with instructions for replicating and editing queries in the future
* 5/3/23 - [Hosting] Detailed documentation for maintaining the database
* 5/3/23 - Complete project

# Related Work

This capstone project has been attempted before by a previous UVA MSDS capstone team. That team created a MongoDB database and dashboards with Flask API. However, the project was ultimately not reproducible. While the previous capstone group’s work informs the work of this team, it is not a precedent for this capstone as the client’s goals have changed.

# Methodology

## Communication

The team uses Microsoft Teams as a main communication tool, and also uses it to communicate with Dr. Judy Fox and Ian Liu, advisors to the project. The main communication tool used with the client, Robin Patton, is email, as well as in-person meetings. Additionally, the team prepares PowerPoint presentations each week for the capstone class. For project management, the team uses a GitHub project board with defined tasks that are assigned to group members and given due dates.

## Roles

The roles of the team are as follows: Rachel Grace is the communication lead, responsible for creating the PowerPoint update presentations, drafting all written communication, and mediating conversations with the client. Ami Kano is the coding lead, responsible for writing search queries and facilitating the backend of the database. Chunru Zheng is the dashboard and visualization lead, responsible for implementing the dashboard tool and creating visualizations for search output.

## Authority Breakdown

In the event of disagreement between the capstone class’s requirements and the client’s requirements, Ms. Patton has final authority over project priorities. Dr. Fox has final authority over what is realistic and appropriate given our time and expertise constraints. At all times, the capstone team has authority over their own availability on weekends and after work hours on weekdays.

## Workflow

A typical week entails a Monday morning capstone class with updates and advice from Dr. Fox and Ian Liu. Around midday on Monday, the team meets with the client to update her on the past week’s progress and ask needed questions. On Tuesday afternoon, the team meets to further discuss deliverables for the weeks. Finally, auxiliary meetings may be scheduled between the client and individual team members to clarify any outstanding issues or questions.

# Difficulties and Solutions

## Communication

Various difficulties have arisen from communication. A main point of conflict was a misunderstanding between the client’s goals and the previous capstone group’s progress. When the current team used the data from the previous capstone group, the client was not pleased and expressed a desire to change that work. In response, the current team scrapped the current work and revised the project scope, adhering to the client’s goals. Another point of frustration has been the client’s desire to understand every detail of the team’s work, and her resistance to anything she does not understand. This comes from her hope to administer and maintain the database herself. In response, the team has agreed to closely document all work, and has worked hard to explain every aspect of the project to the client. The team also drafted a comprehensive scope document that the client has agreed to, which has helped eliminate scope creep and frustration on all sides.

## Cost

The client’s organization, OSS, does not wish to allocate any funds toward this capstone project. This requires all the tools used for the project to be free, and limits the team from using paid tools like Tableau, even with a budget from UVA. For hosting, the free tier of MongoDB is adequate for the capstone needs. However, the dashboard creation will require much more involved work without a paid tool. The dashboard has not been created yet, but this consideration will impact the team’s work as the dashboard is developed.

# Client Satisfaction

At this point in the project, the client has expressed satisfaction with the work and communication done so far. She is pleased with the new data structure and the database created. As the team makes more progress, her satisfaction with the dashboard and its capabilities will be a significant marker of future accomplishments.