Matthew Edwards – me4bb

Colin Warner – ynq9ya

**Progress Report #2**

**Executive Summary:**

One Shared Story’s mission is: “to expand what we know about the past so we can better understand the present and work together to create compassionate, inclusive communities for the future.” They do this by encouraging research to reveal hidden histories of individuals and communities, provide support to amplify underrepresented stories, and foster dialogue within communities, and the general public to expand our common narrative. They have been around for a few years and have many spreadsheets of data consisting of transcribed historical records. Robin Patton is a part of One Shared Story and she contacted UVA in need of some help with their data. Through multiple meetings we have come up with three objectives: Standardize the way volunteers transcribe data, store transcribed data in a database, create an interactive search tool.

The data is the most important part of this, or any other, project. One Shared Story has many excel sheets with data transcribed from historical documents from the local area. These have differing schemas depending on what the document is: wedding records, personal property tax info, slave laborer records, etc. One Shared Story has shared the following four variables that will have some info in every dataset: Event, Source, and People. There will be more data depending on what was being recorded, but these will be our starting point as we work with the data.

One issue that One Shared Story has run into is the fact that they have many different types of datasets they are transcribing. They search out historical documents, gain access to them, and then transcribe the data from them. Each document tracks something different and therefore will have different data types from other documents. The volunteers transcribing this data currently do so on excel sheets. The first thing we need to do is create a web portal that standardizes the way volunteers gather data. This web portal will also lead to a uniform storage location for the data.

One question we need to answer as we get more of the data is what type of database is going to be best. With the differing types of data, and the fact that each new dataset could potentially be new and different as well, we are going to approach this problem with a NoSQL type database like MongoDB. This will allow for more flexibility as data is stored, and will give volunteers a chance to transcribe all sorts of records in the future.

One Shared Story often visits churches on the weekend to allow individuals to learn about family ancestry. This is where the third objective comes in for our group. They need an interactive search tool to help this process go smoothly for new users. The goal for this tool would be to allow these families to search for their ancestors and view information as it relates to the sources, events, people, and roles that are defined in our data model.

**Overview:**

To accomplish what Robin has asked, we started with a first pass at a database and transcribing tool. This was built off the schema of two datasets sent by One Shared Story. We presented this tool to One Shared Story and they were very impressed and pleased with the initial step. We felt that the database actually would be better if it was in the form of a MongoDB database and then created a newer tool that has the capability of keeping the transcribing tool active, while moving the database into a MongoDB format. This will allow for flexibility of storage for future projects. Robin was a little wary of this setup in a meeting we had, but we were able to ease many of her concerns. It seems that at times she still doesn’t know exactly what she wants but is supportive of the direction we are going with the database tool.

Because she is not interested in any real analysis, the lit reviews we have done have been in setting up databases. And the challenges for us so far have been in our knowledge of how to get the database set up. Once we have it constructed and functional, the next step is going to be how to store it (on a local computer, AWS, etc.) and will do some research and planning with Robin to decide what is going to be best going forward for them.

**Data:**

We were finally able to get “all” of the data from Robin. I use quotes because I feel like there is more data that she hasn’t given to us yet. Some of the data is in excel files, including files with multiple tabs for the data. She also sent us a few access db files of data. That has been trucky because none of us has Microsoft access and it has been problematic opening them another way. We are working on finding a solution to make sure we get everything One Shared Story has into the database.

The data that we can open consists of 5 excel sheets that contain the following information: Louisa County Birth records, Fluvanna County property taxes from multiple years, and Louisa County personal property taxes from multiple years. These are able to fit fairly nicely into the preferred classification method of using an event, source, and people mentioned above. The other information will then fit into those overarching categories and will be nicely stored.

The data currently is just being stored on Robin’s individual laptop. That is one of the main reasons for the implementation of a database. She is hopeful to be able to use the database to collaborate with other organizations.