

Carolyn Holmes, Gabrielle Curcio, & Xuanyi Zhao

February 10th, 2021

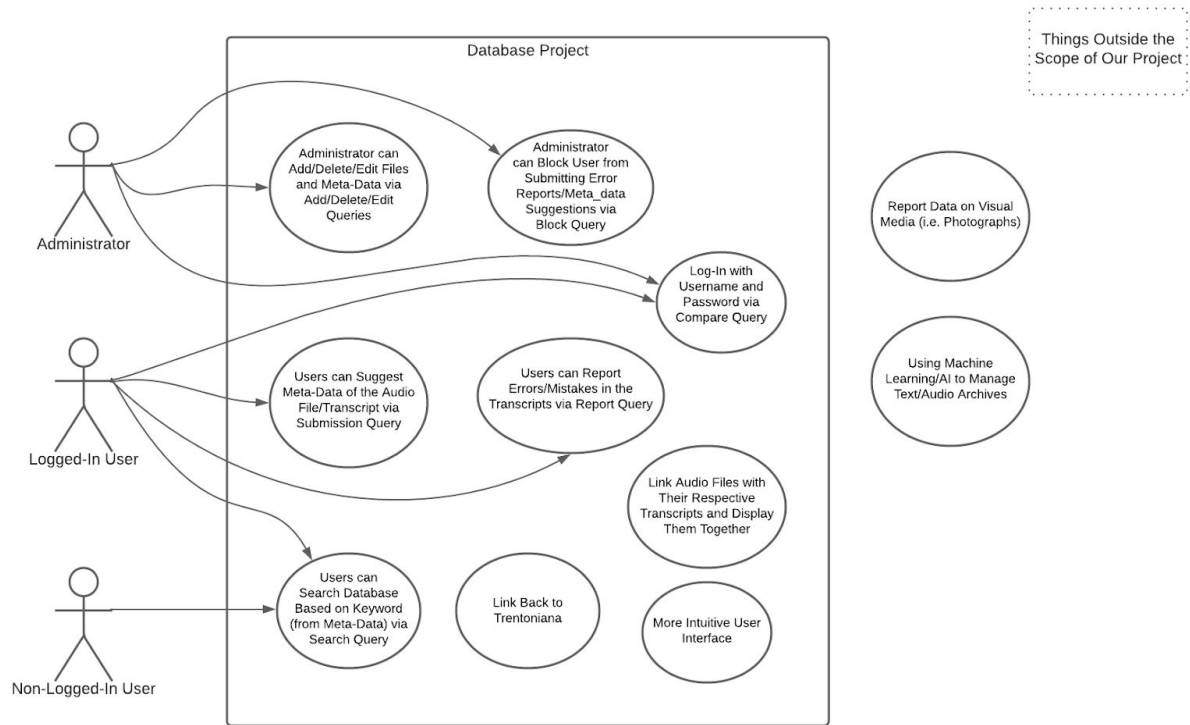
CSC 315-01

Project Proposal:

- **Problem statement:**
 - In the Trentonian Library online, audio files are stored inefficiently, making it difficult for users to search and sort through the files to find those that are best suited for their needs. Additionally, the user interface is overly complicated and confusing for a casual user. These problems with the library must be addressed to improve the user's experience.
- **Objective of the module:**
 - Creating a clearer user interface that allows for better access to the data (easier to use search and sort functions).
 - Creating a database system that handles audio-based data.
 - Link audio files to their corresponding transcripts.
 - Allow upper-level users to submit possible meta-data for files (such as topics addressed in the file) or issues/errors found with the file.
 - Allow casual-level users (such as people not logged on) to view transcripts and listen to audio files (presented on the same page).
 - Allow an administrator to edit meta-data or add/delete/edit files.
- **Description of the desired end product, and the part you will develop for this class:**
 - A database system, including a user interface, that allows better access and management of the Trentonian Library archive.
 - This database will contain different levels of permissions/operations to be done based upon who the user is.
 - Certain users will be able to submit issues or errors they found with the transcript and they can submit meta-data about the file (that could make the file more easily searchable) and the administrator can review the submission to see if they should change the file.
- **Description of the importance and need for the module, and how it addresses the problem.**
 - Currently, the audio and visual content on the Trentonian Library provides limited metadata and no transcriptions on certain content.
 - We would like to expand the versatility of the Trentonian Library and make it a better tool for research.
 - Allowing for users to submit error fixes or meta-data would reduce the amount of work a single person (like a librarian) would have to do to maintain the database.
 - We will make the user interface more intuitive and allow for easier searching of files with the inclusion of more meta-data, making it simpler for someone researching to find what they need (for example- they want an audio file that talk about moving costs, that could be a keyword they could search by in our database, which the Trentonian Library doesn't have).
- **Plan for how you will research the problem domain and obtain the data needed.**

- We will look at other audio storage systems (like other online libraries) to decide how we can best set up our user interface.
- Look at ways how transcript files are typically stored.
- Our data will come from the Trentonian Library and Dr. Steele's transcripts.
- Some data, as mentioned, will come from users sending in edits/meta-data suggestions and we could try out this functionality during the testing phase.
- **Other similar systems/ approaches that exist, and how your module is different or will add to the existing system.**
 - [Garland Encyclopedia of World Music Online](#): When a user finds a sound recording or a video, they can play audio while reading some metadata about the recording. An interesting feature about this library is it displays accompanying transcripts to videos.
 - [Schomburg Center for Research in Black Culture](#): When a user clicks on a sound recording entry, they will be taken to a page with the original audio and some metadata (title, date, keywords, etc). However, the page does not display the transcript for the interview.
 - The Library of Congress stores audio files with their metadata like interviewer name, date, and file topic but it doesn't show the transcripts for the files. It also allows a user to search files like our system does.
 - The Southern Oral History Project (<https://dc.lib.unc.edu/cdm/compoundobject/collection/sohp/id/12031/rec/1>) has a good search feature and allows users to sort records by certain fields. The audio files are stored with metadata (visible to the user) like the description of the file, interviewer name/date of birth, interviewer ethnicity, and citation for the record (which is a good idea to add to ours). They also show the transcript with the audio file.
 - While the libraries mentioned above have good user interfaces, ours improves upon them with the user suggestion function, thereby providing help to the librarian with file management.
- **Possible other applications of the system (how it could be modified and reused.)**
 - The system is a system that allows indexing of any data with machine-readable content, and can be potentially used for managing other text/audio archives.
 - The system could be updated to allow for visual media such as the photographs in the Trentonian Library (allowing for people to more easily make a report of who they believe the person in the picture is or submitting descriptions of the context of the photo/video).
 - Implement AI/Machine Learning to analyze the transcript to extract meta-data quickly and without a lot of work on the part of the librarians/database managers.
- **Performance –specify how and to what extent you will address this.**
 - Controlled redundancy: using repeated fields in a database. This helps improve query performance and increase database access speed
 - Denormalization: To avoid searching through multiple files to collect data

- Transcripts are large files that must be stored in an efficient manner, so we will need to store it and retrieve them in such a way that can handle the large amounts of data.
- **Security –specify how and to what extent you will provide security features.**
 - Only allow the administrator to have permission to change/add/delete the actual files.
 - Have encoded passwords so that people cannot get easy access to the administrator's account.
 - If time allows, have a function to allow the administrator to suspend a user if they submit too many error/meta-data suggestions in order to avoid spamming.
 - Assign stricter roles (like people have to be signed in to make error/meta-data suggestions).
 - Use parameterized queries and use constraints to prevent SQL injections.
- **Backup and recovery –specify how and to what extent you will implement this.**
 - Transaction log backup: save all transaction log entries and transaction events that occurred in the database
 - Differential backup: Saves all data since last backup; Perform a daily data backup routine
 - If part of a transaction is not completed, revert the database back to the state before the incomplete transaction began.
- **Technologies and database concepts the team will need to learn, and a plan for learning these.**
 - Build upon current knowledge of Postgres- learn by reading Chapter 2 of Seven Databases in Seven Weeks book and online tutorials
 - Programming language that interfaces with Postgres: learn by using online tutorials- possibly .Net, C, C++, JavaScript, Node.js
- **A diagrammatic representation of the system boundary that specifies what data you will model and which queries you will implement.**



- 1-page quad chart; see: Quad_instructions_template.ppt in the Canvas files section

Trentonian Audio/Transcription Database

Carolyn Holmes, Gabrielle Curcio, Xuani Zhao

Need

- Need a user-interface that is easy to navigate, easy to search the database, and quick to retrieve search results
- Need to be able to search by keywords that describe the content
- Need to be able to see what each file is about without having to read/listen to the entire file
- Need security so that files can't be changed without permission

Approach

- Rely on user comments to classify metadata about the files
- Rely on users to report errors in the transcripts (that admin reviews)
- List metadata for file in the search results
- Show transcript together with audio file
- Grant administrator access to all database features
- Only allow admin to add/change/edit files

Benefit

- An easier-to-navigate database will be more likely to be used by people doing research and they will receive more internet traffic from it
- The librarian will not have as large a workload with reading/listening to files in order to find errors and meta-data
- A simpler, more efficient database will be easier for them to manage

Competition

- Our user reporting error and metadata submission system is unique to our project and helps relieve workload on admin- not present in other systems
- Improves upon pre-existing Trentonian Library by being less confusing to use and having a better search function
- Attempt to improve upon efficiency of other databases