**Project Brief: Excel Data Mapping Automation Script**

|  |  |
| --- | --- |
| Project Name | Excel Data Mapping Automation Script |
| Date | In Progress |
| Project Client | Abbie Norris-Davidson |
| Project Overview | The project involves developing a Python script to automate the mapping of information from one Excel spreadsheet to another based on specified fields.  The script will handle additional requirements as specified during the project scope definition. |
| Goals  & Objectives | Goals:   * To automate the process of transferring data between two Excel spreadsheets based on predefined mappings, thereby reducing manual effort and potential errors.   Objectives:   * Develop a Python script capable of reading data from source Excel files. * Implement logic to map fields from the source file to corresponding fields in the target file. * Handle exceptions and errors gracefully to ensure robust performance. * Allow for flexibility in mapping configurations to accommodate different data structures. * Provide documentation and user instructions for seamless adoption and maintenance. |
| Constraints  & Assumptions | Constraints:   1. User Proficiency: Users are assumed to have proficiency in using Python scripting and Excel for basic operations. 2. Manual Input Requirement: The script requires manual input of the collection name to specify data processing. 3. File Location Dependency: Input Excel files are expected to reside in the same directory as the script for processing. 4. File Naming Convention: The naming conventions for input and output files follow the specifications provided by Abbie Norris-Davidson. 5. Cross-Platform Compatibility: The script is designed to operate seamlessly on any operating system with a Python environment installed. 6. Library Dependencies: Python libraries utilized within the script must be installed in the Python environment where the script will execute.   Assumptions:   1. Field Mapping Specification: The mapping of input and output field names adheres to the structure defined in the eGrove\_metadata\_crosswalk\_layout.xlsx. 2. No need for UI: The project does not include development of a graphical user interface (GUI). Interaction with the script is limited to command-line or script execution environments. |
| Project Scope | * Mapping Metadata from eGrove Excel Sheet to Preservica Excel Sheet: Develop a Python script to automate the mapping of metadata fields from an eGrove Excel spreadsheet to a Preservica Excel spreadsheet based on specified mappings. * Combining Fields for Specific Preservica Requirements: Implement logic within the script to combine multiple fields from the eGrove spreadsheet into specified fields in the Preservica spreadsheet to meet project requirements. * Handling Multiple Authors: Include functionality to identify and annotate rows in the output where there are multiple authors in the eGrove metadata. * Corporate Author Handling: Ensure the script identifies corporate authors in the eGrove metadata and writes the corporate name in the appropriate Preservica field as per project specifications. * Generating Final Output According to Naming Convention: Configure the script to produce the final output Excel spreadsheet formatted according to a predefined naming convention specified by the project guidelines. |
| Technology Used | Used the following python libraries:   1. **pandas**: Python library for data manipulation and analysis, particularly useful for handling structured data like Excel spreadsheets. 2. **os**: Python module providing functions to interact with the operating system, crucial for file management operations in the script. 3. **json**: Python module for parsing and creating JSON (JavaScript Object Notation) data, utilized here for configuring and storing mappings and other settings. |
| Command to run | 1. Open Command Prompt/Terminal on your system. 2. Navigate to the folder with the script. 3. Run python automate.py |
| Target Audience | Abbie Norris-Davidson |
| Success Criteria |  |
| Timeline | No specific timeline mentioned for this project. |