

# Merit Badge Counselor Lists Specification - Updated

## Goals

1. Identify Registered Merit Badge Counselors who are registered with Troop 12 or Troop 32 Acton MA
2. Identify Troop 12 and Troop 32 Unit Leaders who are not registered as Merit Badge Counselors
3. Identify Troop 12 and Troop 32 Merit Badge Counselors for all Merit Badge.
4. Generate separate webpages for the lists identified in the above three goals that can be incorporated into an existing website. a. Generate WordPress-compatible pages and HTML pages for each of the above three goals b. Enable the downloading of the list as CSV, PDF, and Excel files and the printing of the list on each of the generated webpages. c. Identify the generation date and time on each of the generated webpages.
5. Create a tool that will accept the inputs identified further below and re-generate the above webpages on demand. a. The inputs listed below that are links can have the provided links as default inputs. b. Generate clear prompts/instructions for supplying each required input. c. Generate clear and useful console logs to identify progress and errors, if any.

## Definitions

1. Scouting America - formerly Boy Scouts of America.
2. Council - A Scouting America organization that services members in a geographic area.
3. Unit - A local organization of Scouting America members sponsored by a Chartering Organization. Unit types are Packs, Troops, Crews, Boats.
4. Troop <number> (abbreviated as T<number>) - The identifier for a Troop unit type.
5. T12 - Acton MA girls-only troop. AKA T7012 where '70' prefix denotes a girls-only troop.
6. T32 - Acton MA boys-only troop.
7. Merit Badge (abbreviated as MB) - An earned badge recognizing skills in a specific subject matter
8. Rank - A stepped advancement level in Scouting America
9. Eagle Scout - The highest rank in Scouting America
10. Eagle-required Merit Badge - one of a specific set of MBs required to earn the Eagle Scout Rank
11. Youth Member - An under aged 18 person registered with a unit
12. Registered Adult - A person aged 18+ who is registered with Scouting America.
13. YPT - Youth Protection Training. Required to remain current, that is, not expired, for all Registered Adults
14. Merit Badge Counselor - A Registered Adult who has been authorized by the local Council to instruct Scouts on one or more MBs
15. Unit Leader (aka Scouter, Leader) - An Registered Adult who is registered with a unit. Note that a Registered Adult can be registered with multiple units and multiple positions within Scouting America
16. ScoutBook - The data management system used by members of Scouting America.
17. Heart of New England - The name of the local Council that serves T12 and T32

## **Architectural Framework**

1. Programming language: a. Create the tool using the latest released version of Python.
2. Enable the launching of the tool as a GUI executable that includes: a. Simple file selection (drag & drop) b. Progress indicators c. Built-in help/instructions d. Automatic output folder opening e. Error messages in plain English
3. Additionally provide a command-line interface (CLI) for: a. Batch processing multiple troop sets b. Automation and scripting capabilities c. Integration with other tools d. Advanced users who prefer terminal workflow
4. Distribution: a. Primary: Standalone executable (.exe/.app) requiring no Python installation b. Secondary: Python package installable via pip for developers c. Include sample data files for testing
5. File Organization: a. Create timestamped output folder (e.g., "MBC\_Reports\_2025-06-02\_14-30") b. Organize by format: /html/, /csv/, /pdf/, /excel/, /wordpress/ c. Include a summary report with statistics and validation results d. Auto-open output folder upon completion
6. Error Handling: a. Graceful handling of corrupted/incomplete HTML files b. CSV encoding detection (UTF-8, Windows-1252, etc.) c. Network timeout handling for scouting.org data d. Clear error messages with suggested solutions e. Partial processing capability (continue with available data)
7. Debug Logging: a. Print extracted data for each processed row in roster: i. First Name - column heading is firstname ii. Last Name - column heading is lastname iii. Unit Position - column heading is positionname iv. Email - column heading is primaryemail v. Phone - column heading is primaryphone b. Print extracted data for each registered merit badge counselor: i. First name ii. Alternate first name iii. Last name iv. Email addresses (multiple if present) v. Phone numbers (multiple if present, with home, work, mobile labels) vi. List of parsed merit badges c. Print compiled list of adult leaders and indicate whether they are also a registered merit badge counselor d. Write debug logging to the <tool name>.log file.
8. Configuration: a. Settings file for default URLs and preferences b. Output format preferences
9. Input Validation: a. CSV format verification (required columns: firstname, lastname, positionname, primaryemail, primaryphone) b. HTML content verification (contains merit badge counselor data in expected structure) c. Web connectivity check for scouting.org access d. File size limits and warnings e. Duplicate detection and handling
10. Progress Tracking: a. Overall progress bar (0-100%) b. Current step indicators ("Processing HTML files... 3/5") c. Time estimates for large datasets d. Cancel operation capability e. Background processing option

## Inputs

## 1. List of all merit badges

- Source: <https://www.scouting.org/skills/merit-badges/all/>
- They begin after "Merit Badges A-Z" heading
- Ignore the "Requirements Update" section at top of this page
- **Updated Cached List:** Must include "American Business", "American Cultures", "American Heritage", "American Labor", "Motorboating", "Multisport", "Citizenship in Society"
- **Remove:** "Eagle Scout" (this is a rank, not a merit badge)

## 2. List of Eagle-Required merit badges

- Source: <https://www.scouting.org/skills/merit-badges/eagle-required/>
- **Updated Eagle-Required List:** Must include "Cycling" and "Sustainability" in addition to standard badges

## 3. T12 Roster

- A CSV containing a list of T12 members including Registered Adults and Youth Members
- The column headers for the relevant rows are on the line beginning with "..memberid"; ignore the previous rows
- Extract the values for the following column headers: i. First Name - column heading is firstname ii. Last Name - column heading is lastname iii. Unit Position - column heading is positionname iv. Email - column heading is primaryemail v. Phone - column heading is primaryphone
- Youth Members are identified by positionname column value containing "Youth Member"
- Registered Adults are those not identified as a Youth Member

## 4. T32 Roster

- A CSV containing a list of T32 members including Registered Adults and Youth Members
- The format of the T32 Roster matches the format of the T12 Roster

## 5. List of Merit Badge Counselors - UPDATED FOR HTML INPUT

- **Input Format:** HTML files saved from ScoutBook Merit Badge Counselor search results
- **Source:** Search results from ScoutBook for Registered Merit Badge Counselors within 5 miles of zip code 01720 (Acton MA)
- **Structure Requirements:**
  - Each counselor entry must be in a `<div>` with `style="margin-left: 65px"`
  - Contact information in `<div class="address">`
  - Merit badges in `<div class="mbContainer">` containing `<div class="mb">` elements
- **Data to Extract:**
  - First Name (from main text node)
  - Alternate First Name (within parentheses if present)
  - Last Name (remainder after first name)
  - Phone numbers (Home and Mobile patterns)
  - Email (from mailto links)
  - List of merit badges (from mb container divs)

## Parsing Requirements - **CRITICAL UPDATES**

### Name Parsing

- **Handle complex last names:** McMillen, MacQuilken, deBlonk, Duffy-Reichert
- **Detect parenthetical alternate names:** "Christopher (Chris) White"
- **Prevent truncation:** Ensure full last names are captured
- **Validation:** Reject obvious non-names containing council-related terms

### Merit Badge Parsing

- **Council Phrase Filtering:** Remove the following before badge detection:
  - "Approved by the Heart of New England Council"
  - "Counselor for Heart of New England Council"
  - "Heart of New England Council"
  - "Heart of New England"
  - "New England Council"
- **Prevent False Matches:** Ensure "Art" is not detected from "Heart" and "New England" is not treated as a merit badge
- **List Independence:** Each counselor must have completely independent merit badge lists (use copy.deepcopy)

## Contact Information Cleaning

- **Phone Format:** Normalize to XXX-XXX-XXXX format
- **Handle 11-digit numbers:** Remove leading "1" country code
- **Email Validation:** Extract valid emails and handle malformed concatenations
- **Deduplication:** Remove duplicates while preserving all valid contact methods

## Outputs

### 1. T12/T32 Merit Badge Counselors

- Cross reference the compiled List of Merit Badge Counselors with the Registered Adults in the T12 Roster and T32 Roster to identify the T12/T32 Merit Badge Counselors
- **Matching Logic:**
  - Match the First Name in a roster to the First Name or Alternate First Name in the compiled List of Merit Badge Counselors
  - Match the Last Name in a roster to the Last Name in the compiled List of Merit Badge Counselors
  - Disregard Youth Members
- **Output Fields:**
  - First and Last Names
  - Unit Position
  - Registered troop(s)
  - All email addresses
  - All phone numbers
  - Complete list of merit badges
- **Phone Format:** 123-456-7890
- **Multi-troop handling:** If a merit badge counselor is registered with multiple troops, output one entry with troop number as "T12, T32"

## 2. T12/T32 Leaders not Merit Badge Counselors

- List the first and last names of Registered Adults in T12 and T32 who are currently not Merit Badge Counselors
- **Sort:** By Last Name, First Name
- **Multi-troop handling:** If a Registered Adult is registered with multiple troops, output one entry with troop number as "T12, T32"

## 3. T12/T32 Merit Badge Counselor Coverage

- List all merit badges and their T12/T32 merit badge counselors in the following order:
  - Eagle-required merit badges with corresponding T12/T32 merit badge counselors (sort alphabetically)
  - Eagle-required merit badges without corresponding T12/T32 merit badge counselors (sort alphabetically)
  - Non-Eagle-required merit badges with corresponding T12/T32 merit badge counselors (sort alphabetically)
  - Non-Eagle-required merit badges without corresponding T12/T32 merit badge counselors (sort alphabetically)

## 4. Output Formats

Support generation of the above three lists in the following output formats:

- **HTML:** Self-contained with embedded CSS, working download buttons
- **PDF:** Professional formatted reports with letterhead option
- **Excel:** Multi-sheet workbook with data validation and formatting
- **CSV:** UTF-8 encoded, properly escaped, Excel-compatible
- **WordPress:** Shortcode-compatible with embedded download functionality

## Testing Requirements

### Data Validation

1. Verify the total number of merit badges in "T12/T32 Merit Badge Counselor Coverage" matches the total number of merit badges in <https://www.scouting.org/skills/merit-badges/all/>
2. Verify the total number of Eagle-required merit badges matches <https://www.scouting.org/skills/merit-badges/eagle-required/>
3. Verify the total number of Merit Badge Counselors in "T12/T32 Merit Badge Counselors" plus the total number of adults in "T12/T32 Leaders not Merit Badge Counselors" match the total number of Registered Adults in the T12 Roster plus T32 Roster
4. Verify the "Download CSV File" and "Print" buttons work for each generated webpage
5. Verify the generated webpages are without system and console errors

### Performance & Usability



6. Performance: Process 1000+ counselors within 30 seconds
7. Usability: Non-technical user can complete workflow in under 5 minutes
8. Cross-platform: Test on Windows 10/11, macOS, Linux
9. Data integrity: Verify no data loss during processing
10. Output validation: All generated files open correctly in standard applications

## **Name Parsing Validation**

11. **Complex Names:** Verify correct parsing of McMillen, MacQuilken, hyphenated names (e.g., Duffy-Reichert), and names with lowercase prefixes (e.g., deBlonk)
12. **Alternate Names:** Verify proper handling of "Christopher (Chris) White" format
13. **Council Filtering:** Verify "New England" and "England Council" are not treated as counselor names






## **Merit Badge Validation**

14. **Council Phrase Filtering:** Verify "Art" is not incorrectly assigned from "Heart of New England Council"
15. **Complete Badge Lists:** Verify all counselors have their complete merit badge lists (e.g., Herb Philpott should have all 20 badges, not just 5)
16. **Missing Badges:** Verify "American Cultures", "American Business", "American Heritage", "American Labor", "Citizenship in Society" are properly detected when present

## **Developmental Order**

1. Implement necessary framework to support CLI and GUI and all output file types features a. Initially, only the essential functionality of CLI input and HTML output will be implemented in order to test data processing
2. Essential functionality a. Implement CLI input only where input files can be specified b. Implement HTML output file type only
3. Cache the lists of all merit badges and Eagle-required merit badges from scouting.org a. Repeated runs of the code will result with scouting.org denying access to these lists b. Upon first successful retrieval of these lists, save them in data structures in code as default values to use should there be any subsequent failure to retrieve them from scouting.org
4. Testing a. Test with dummy input data - verify output is correct b. Test with real input data - manually verify the output is correct
5. Implement GUI input after confirming CLI/HTML testing passes a. Repeat above Testing step
6. Iteratively implement and repeat testing with the following output file types: a. CSV b. PDF c. Excel d. WordPress

## Fixed Defects - RESOLVED

1. ~~The parsed list of merit badges associated with each merit badge counselor was incomplete~~  **FIXED:** Switched from PDF to HTML parsing
2. ~~Names were being truncated (McMillen → Mc, MacQuilken → Mac)~~  **FIXED:** Improved name parsing with duplicate detection
3. ~~False merit badge matches from council names ("Art" from "Heart", "New England")~~  **FIXED:** Added council phrase filtering
4. ~~Shared merit badge list objects between counselors~~  **FIXED:** Implemented copy.deepcopy for list independence
5. ~~Missing merit badges in cached lists~~  **FIXED:** Updated cached lists with all required badges

## Notes

1. Refer to [scouting.org](https://www.scouting.org) for more information about Scouting America as needed
2. The merit badge lists should be derived from only [scouting.org](https://www.scouting.org) as they are always current
3. The specifications within this document supersede prior chats
4. **Input Method:** Use HTML files from ScoutBook search results instead of PDF files for more reliable parsing
5. Request clarification where needed in this chat
6. Future enhancements:
  - a. The retrieval of rosters for specified units directly from ScoutBook
  - b. The retrieval of Merit Badge Counselors directly from ScoutBook API
  - c. Batch processing for multiple units
  - d. Automatic updating of unit websites
  - e. Generate test data and expected results from actual data and reports, once the reports are manually verified as correct