**Open Source Security Checklist**

**Software Title and Contact Information**

|  |  |
| --- | --- |
| 1. **Title of Software:** | Termset Generator |
| 1. **Name of Submitter (Office):** | <name> (Office) |
| 1. **Date Submitted:** | [Insert the date the Open Source Security Checklist was completed. *DD/MM/YYYY* format.] |

**Background**

The release of government-developed software as open source is now a priority as indicated by the draft policy titled “Leveraging American Ingenuity through Reusable and Open Source Software” released by the White House on 10 March 2016. This draft policy <https://www.whitehouse.gov/blog/2016/03/09/leveraging-american-ingenuity-through-reusable-and-open-source-software> cites the savings in taxpayer dollars realized by avoiding duplicative custom software purchases and by promoting innovation and collaboration across Federal Agencies. The FDA can benefit from having the brightest minds inside and outside the government reviewing the code to ensure that it is secure, reliable, and effective in furthering our public health objectives. While government-developed open source software can benefit both the government and stakeholders, there are two key points that must also be addressed: software quality and the potential exposure of FDA’s IT security infrastructure and confidential data.

The issue of software quality is beyond the scope of this checklist. It is important that developed software follow best programming practices. A comprehensive list of best programming practices can be found with an internet search. An example of a few of these practices are: knowing what the code block must perform, describing variables, keeping code simple, maintaining consistent naming conventions, validating input, architecting for security, using default deny practices, adhering to the principle of least privilege, sanitizing data sent to other systems, and practicing defense in depth. These principles must be applied in the development of the code from the beginning, and not at the stage of preparation of the code for open source. Government groups that want to release their software to open source should ask themselves if the software is of sufficient quality to merit release. If they feel the software is of sufficient quality, then they should complete this checklist.

Specific checklist items address the FDA Office of Information Management Technology (OIMT) security concerns for the possible release of secure or confidential information. Please check the candidate software for these issues and document how they were checked. Your description for checking should be clear and complete enough to be replicated by the OIMT ISSO’s if they want to check your work.

Provide the location of candidate open source code and SHA-256 hash string. If the hash string changes, then the checklist must be revisited before release. Only the source code that matches the approved hash string may be released.

**Summary of Security Checks Performed**

**Name of individual checking the software:** *Doug Billings*

**Date software check was performed:** *7/28/2021*

Please provide detailed descriptions of the checks performed on the proposed open source software. At a minimum, the following checks should be performed (or an explanation of why the check is not applicable to this software must be provided).

1. Validate that all account credentials, database connection strings with usernames/passwords, security keys, tokens, API keys, accounts/passwords for connections to third parties, computer or server names, IP addresses, and unique build parameters have been removed from all code, configuration files, and code libraries. (Look for keys which are usually hexadecimal strings of arbitrary length, server names are usually preceded by “\\”, and IP addresses in the form of “XXX.XXX.XXX.XXX” or “xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx”.)

Detailed description of how this was checked:

*Visual inspection of code by multiple reviewers.*

1. The same applies to any/all third-party software configuration files or code libraries that are included within the software’s repository. (Repeat above, but only if the code libraries have been edited to include FDA resources.)

Detailed description of how this was checked:

*N/A. No third party code is included.*

1. What are the types of licenses associated with the libraries used in this open source software (if any)? For example, MIT license, Creative Commons license, etc. To determine this, go to the websites of the libraries used.

List of libraries and open source license associated with it:

|  |  |
| --- | --- |
| **Library:** | **Open Source License:** |
| pandas | BSD |
| scispacy | Apache |
| streamlit | Apache2 |
| ujson | BSD |
| UMLS (data not code) (automatically downloaded by scispacy) | UMLS |

1. Validate that all security functions (i.e. user authentication and session management) that rely on cryptographic mechanisms and have corresponding keys, are properly handled and removed. (Search for strings that include “MD5” or “SHA”.)

Detailed description of how this was checked:

*Visual inspection of code by multiple reviewers. No security functions are included.*

1. Validate that all sensitive comments in the code are removed. Comments are a part of best programming practices; however, comments that point out weaknesses of the code or are not well tempered (i.e. expletives) must be removed. (Search for comment strings and use a dictionary list of specific words.)

Detailed description of how this was checked:

*Visual inspection of code by multiple reviewers.*

1. Insert SHA-256 hash string of software submitted. (use Windows 10 PowerShell to generate using the cmdlet “Get-FileHash”)

File: termset\_generator.zip

Date: 8/31/2021

0d9adbc72a217dd662ddbf424737b141fb2d50c8d2d5ac5bd7872ddfacc2919f

**Document Revision History**

Please track revisions to this document in the table below.

| **Checklist Version** | **Created/Updated By** | **Date**  **(MM/DD/YYYY)** | **Reason for Change** |
| --- | --- | --- | --- |
| 1.01 | Mark Walderhaug | 11/16/2016 | Initial Template |
| 1.02 | Vali Tschirgi | 12/2/2017 | Template modified to add header table and revision history table. Added date checks were performed. |
| 1.03 | Mark Walderhaug | 5/20/2019 | Added check for open source library licenses and place to insert hash string |