

- Real world softwares

- Widely used (software reputation, download times, integrated with other software/libraries)
- Continuously maintained

While considering the data source software, our team considered the following aspects:

1. Widely used
 - a) software reputation
 - b) integrated with other software/libraries
2. Continuously maintained

◆ FFmpeg



FFmpeg is a free open-source software project consisting of a suite of libraries and programs for handling video, audio, and other multimedia files, and streams.

Over 1,160 contributors are contributing to this project

Maintenance: The latest commit was 7 Oct 2021

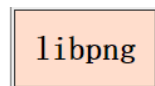
GitHub stars: over 26.4k

Forked: Over than 8.8k times

Integrated: 'MPlayer', 'HandBrake', 'Google Chrome' and .etc.

Link: <https://github.com/FFmpeg/FFmpeg>

◆ LibPNG



Libpng is the official PNG reference library. It supports almost all PNG features, is extensible, and has been extensively tested for over 23 years. You can easily read and write every pixel from a PNG file by using Libpng.

Over 45 contributors are contributing to this project

Maintenance: The latest commit was 13 Mar 2021

GitHub stars: over 751

Forked: Over than 412 times

Link: <https://github.com/glennrp/libpng>

◆ LibTIFF



Libtiff is a library for reading and writing Tagged Image File Format (abbreviated TIFF) files.

Over 66 contributors are contributing to this project

Maintenance: The latest commit was 04 Oct, 2021

Link: <https://gitlab.com/libtiff/libtiff/-/tree/master>

- NVD/CVE labelled
 - We need the labels from the data sets to train/test our models.
 - The vulnerability labels among these projects data are trustable. Because NVD is funded by National Institute of Standards and Technology (NIST) in 2005 which is a recognized authority in Vulnerability detection field.