FOSSology: Feature Overview

What is FOSSology?

A Web server application for license and copyright compliance of software components.

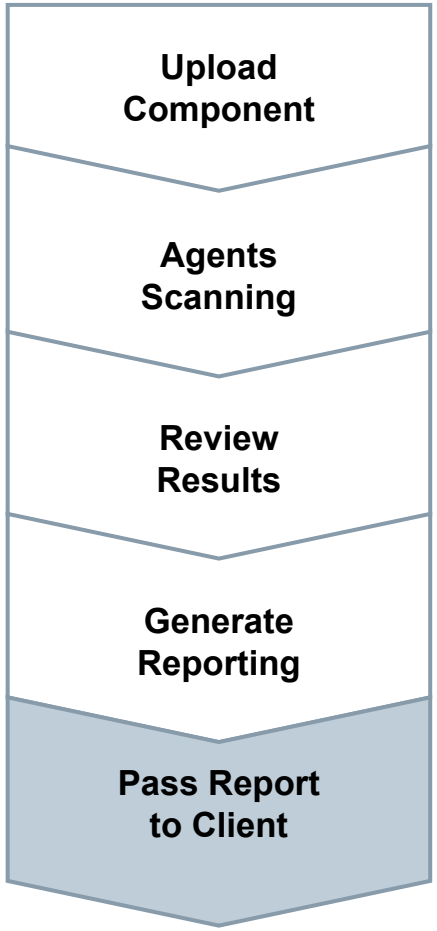
FOSSology Project
https://www.fossology.org/

- Published first in 2008, GPL-2.0
- 2015: Linux Foundation collaboration project
- Web server based and command line interfaces
- Scanning agents searching for license and copyright relevant hits (and more …)
- A multi-user / multi-tenant Web UI for review organizing clearing job

FOSSology Development
https://www.github.com/fossology/fossology

- Standard Web application stack:
  - Linux, Apache 2, PostgreSQL, PHP,
- Web-based UI in PHP, but scanners written in C / C++
- Besides Web UI:
  - Command line utilities

How does FOSSology work?



- Uploading source code archive (*.zip, *.tar.gz, etc)

- Agents scan for license relevant text
- Copyrights, ECC, your keywords to look for etc.

- Review scanner results for wrong license classification
- Review other scanner findings (copyrights, ECC)

- Result of the “clearing”
  - SPDX reporting
  - Generated notice or readme file
  - debian-copyright

fossology

Hands-On: Basic End-to-End Workflow

Functionality

1. **Using FOSSology End-to-End**
   - From uploading …
   - … to generating report: SPDX
2. **Uploading - offers a variety of selections**
3. **Review the uploaded file in the license browser**
4. **Review the found licenses in the aggregated view**
5. **Do the clearing work**
6. **Review the copyrights**
7. **Review the Export Control and Customs (ECC)**
8. **Generate desired report output**

Example

Upload ionicons-3.0.0

- Go to license browser by clicking upload name or selecting clearing from the action menu
- Select “go to all …. with licenses”
- Review licenses and apply decisions

Select the copyright, e-mail, url section from the yellow menu bar area

- Review copyright statements and correct in case

Select ECC from the yellow area and review

Go to Browse main view

- Select the pop-up menu of the icons upload

Features: Two License Scanner: Nomos and Monk

Nomos Keywords	Nomos Reg. Expressions	Bulk Phrase Matches	Monk Full text Matches
• Finds all kind of license relevant texts • Finds unknown Licenses	• Finds most license relevant texts • Identifies also derivatives of licenses	• Good for finding actual licenses • Identifies also derivatives of licenses	• Certainty that known license text is actually found and wording is exactly reproduced

Flexibility

• Very imprecise • Does not identify license • High number of false positives	• Only limited precision for identifying actual licenses • False positives	• Limited to known phrases only • Does not provide certainty about original or derivative	• Works only on known license texts • Actual occurrences are minority

Precision

fossology

Features: Bulk Scan, Reviewing Many Files at Once

The FOSSology software cannot know all possible license relevant text phrases

Use Case

- Finding standard and known license texts is straightforward
- ***Do I need to create a clearing decision for every file?***
- ***If the scanner yields wrong results, do I need to go into every file and correct then?***

Solution

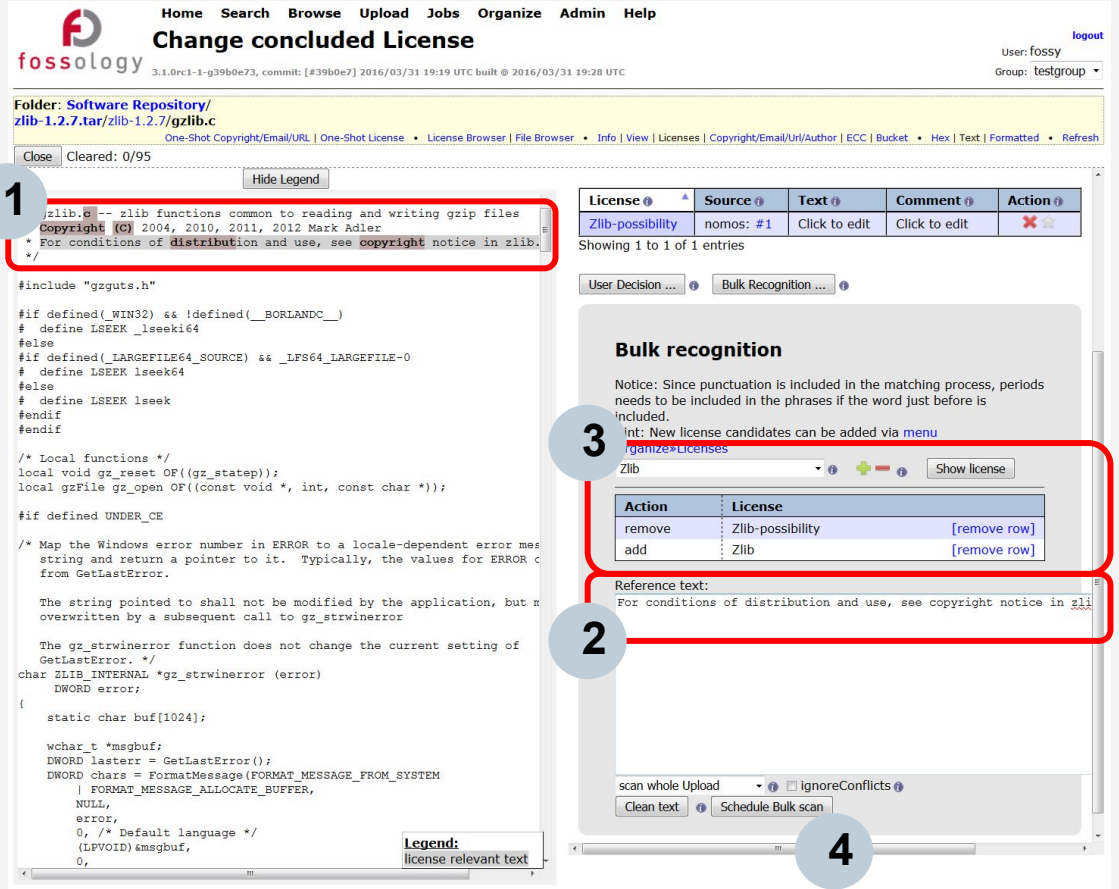
- FOSSology allows the user to define text phases …
  - **For confirmation …**
  - **… or correction**
- And assign license confirmations or corrections to it
  - Every time a file is found with this text phrase (100%) …
  - … the confirmation or correction is applied to it.
- No need to go through every file again.
- Does not apply with scan result conflicts

fossology

Features: Using Bulk Scan

User Interface

- Copy a characteristic text phrase from file view left (1)
- Paste into the bulk scan text field (2)
  - The application will search for file with this text
- Define scanner license findings (3)
  - For correction / removal
  - For confirmation, creating a clearing decision
- Bulk scan will run over all files of the package and apply clearing decisions or scanner corrections where matching (4)



Key Features: Bulk Scanning – A few Things to Consider

Scoping Bulk Scan	• Bulk scan can be applied on folders but also an entire uploads
Larger bulk scan sections	• Prefer larger text portion instead of smaller ones, this increases the precision • However, do not consider copyright statements for example, as they usually vary within an upload
Comment characters	• Comment characters such as #, //, * , /* are filtered out anyway, copying them in does not matter • But other characters are not filtered out („%“). They will make a difference
No characteristic text phrase	• Consider correcting license scan results directly using the edit feature • With the edit feature available in the license browser view, no text phrase is defined

Features: Edit License Conclusions in License Browser

Edit is about a short cut at license review

Use Case

- Within a software component, I know files / folders already.
- With a software component, I know that files are irrelevant.
  - Unused architecture
  - Test files, code examples
  - Build infrastructure
- **Do I need to look at the files?**

Solution

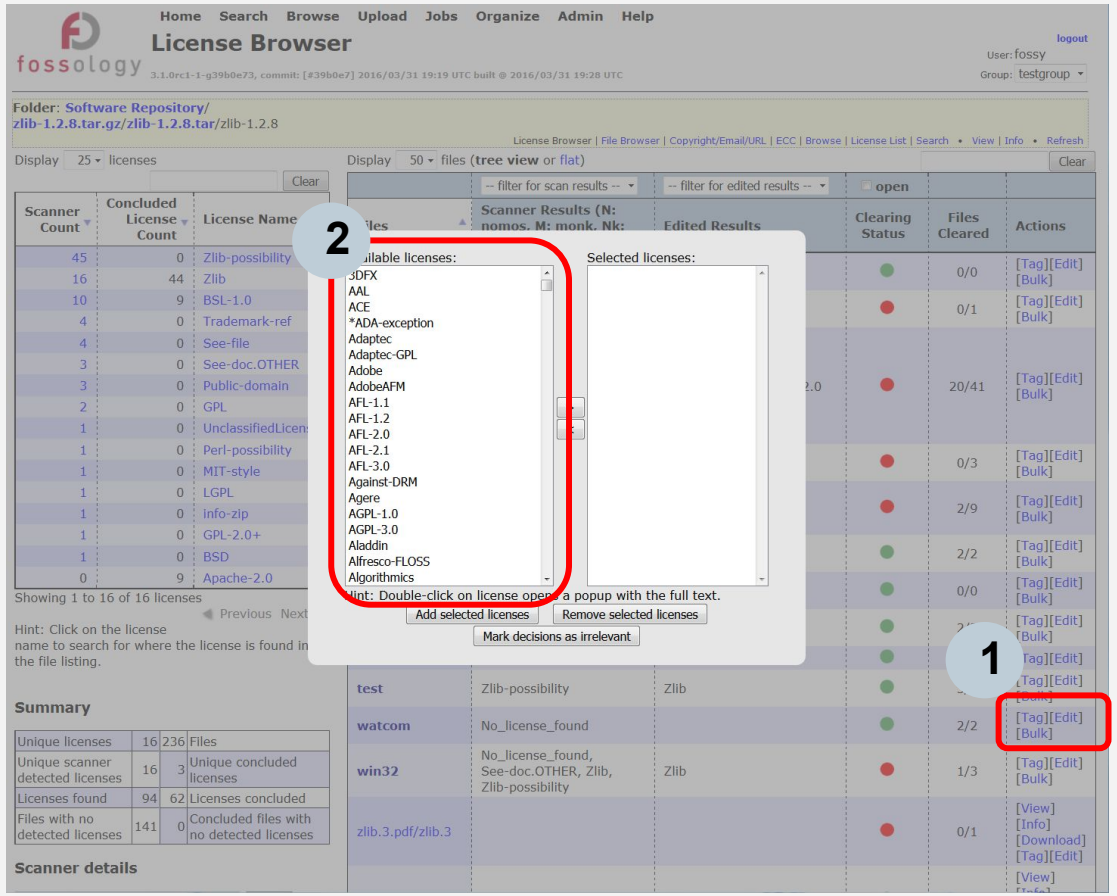
- FOSSology allow to set / unset license on folder level
- In license browser, the user can set the licenses on folders
  - Set or confirm licenses
  - Remove scan results from the clearing decision
  - Mark files as irrelevant
- What happens actually?
  - The “Edit” creates clearing decisions
  - The scan results are not touched (and preserved)
  - No clearing decision final on scan result conflicts

fossology

Features: Using License Conclusions in License Browser

User Interface

- At aggregated license browser view select link [Edit]
- Select either licenses for decision …
- … or mark file trees irrelevant (for distribution)



Features: Reuse of License Corrections

Reuse is about reusing already done work.

Use Case

- Upload and clearing of a component:
  - Going file-based
  - Maybe using bulk phrase
  - Reviewing copyrights
- ***Uploading a newer version***
  - Do I need to go through all files again?

Solution

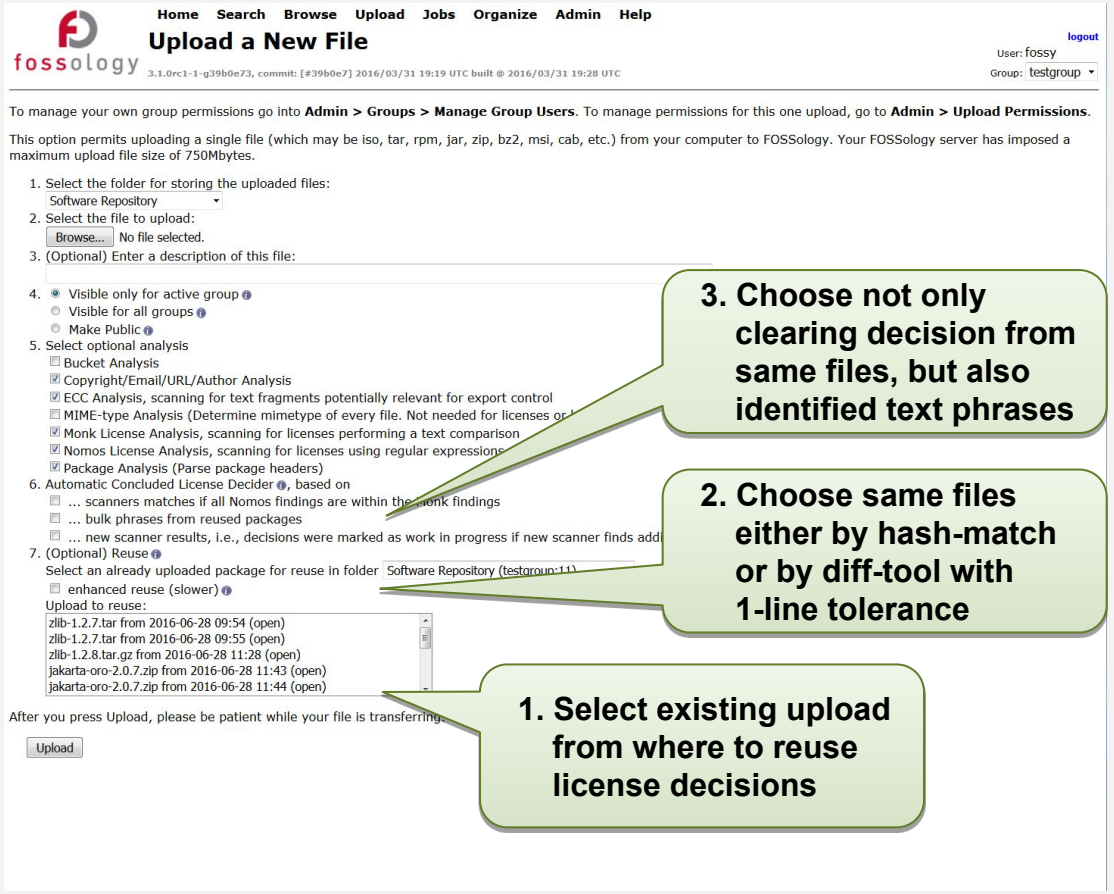
- FOSSology maintains hash values of every file
  - It already reuses scans (rescan possible for new agent)
  - **Users want to reuse clearing decisions = what the user has decided**
- Three different reuse functions
  - Reuse clearing based on same file hashes
  - Reuse bulk scan operation = individual text phrases
  - Reuse where file differs by one line (using diff-tool = slow!)

fossology

Features: Using Reuse of License Corrections

User Interface

- At upload you can select another existing package on server for reusing license review data
- Three main reuse options given:
  - Reuse license review data at same hash value calculated for files (1)
  - Reuse license review data at one-line tolerance using the diff tool (2)
  - Reuse bulk scan tasks entered for selected existing package also for new package (3)



Features: Copyright Statements

Edit is about a short cut at license review

Use Case

- Many license obligations ask for listing the copyrights at distribution.
- Example form BSD: *“Redistributions in binary form must reproduce the above copyright notice,…”*
- **How do I find out all relevant copyright statements of an upload?**

Solution

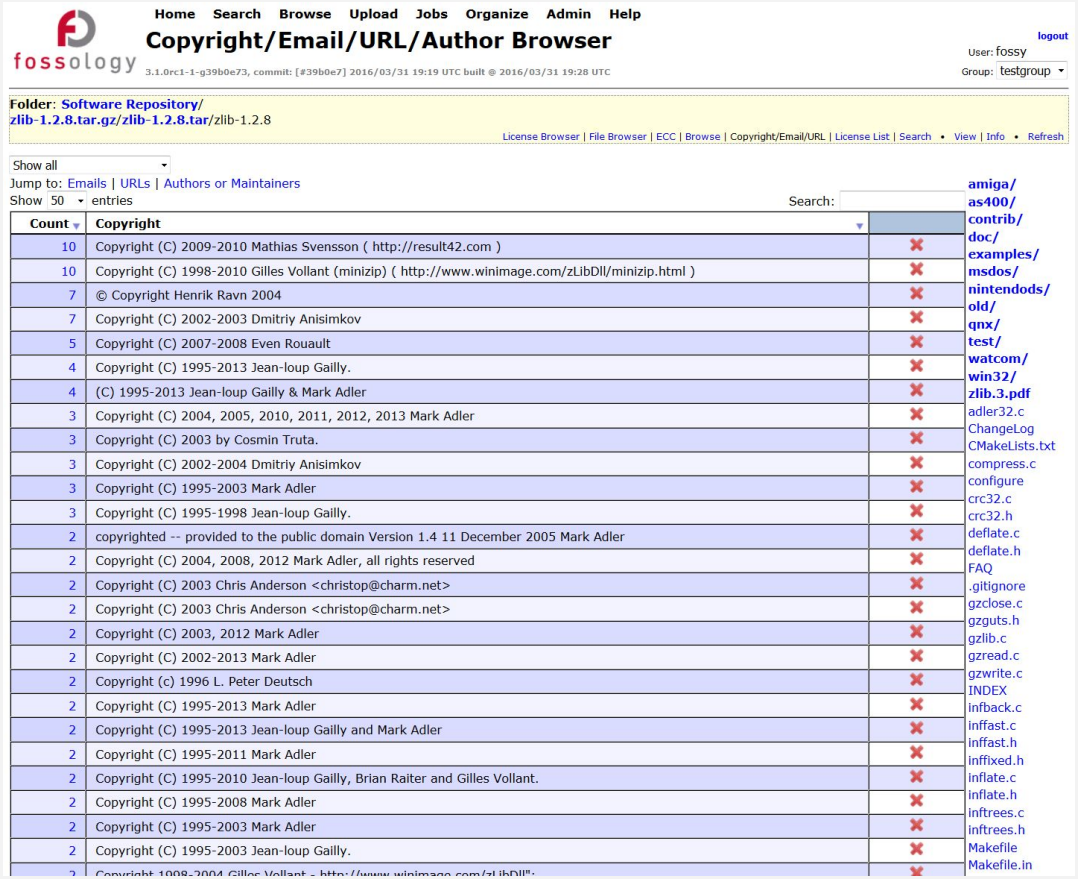
- FOSSology scans based on regular expressions
  - Copyright statements (yes, also © in UTF-8, 1152, …)
  - Authored by, contributed by, e-mail addresses
  - E-Mail addresses and http/https URLs
- Stored on file level!
  - Maintained by hashes of the file
  - Editing the copyright of a file, edits it also for future uploads
  - Recovery of deleted copyright statements possible
  - Table-sheet editing up to come

fossology

Features: Using Copyright Scanning

User Interface

- Copyright findings are presented in an aggregated view
  - For single files
  - All found items at folder level
  - All items of the upload



Features: Reporting

For distribution of OSS, the license relevant information must be provided

Use Case

- All the package has been analyzed, but to get the notice file contents: how do I get my results?

Solution

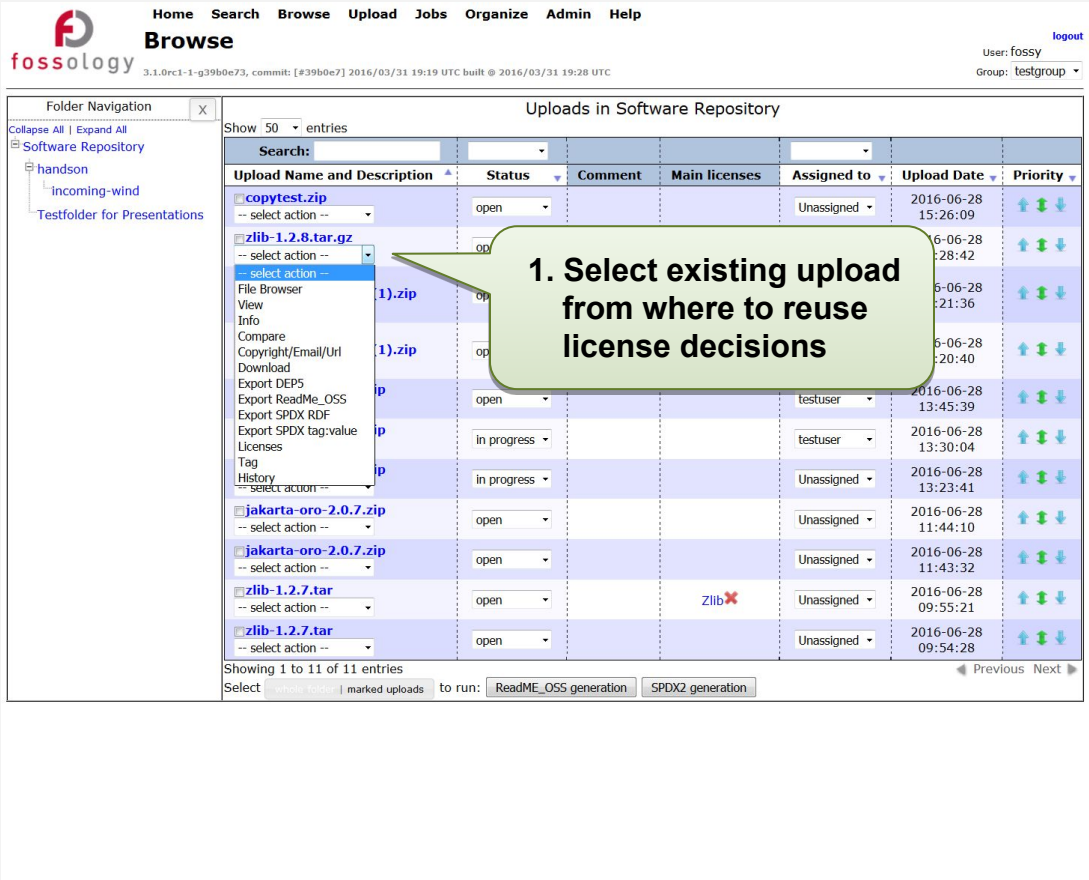
- FOSSology allows for exporting license and copyright information about the open source package in various formats:
  - **The Linux Foundation / OpenChain:**
    - SPDX 2.0 reporting in both Tag:Value and RDF/XML notations
  - **Debian ecosystem:** Debian-copyright (a.k.a. DEP5) files
  - **For your use:**
    - Plain listing of licenses with files
    - Generation of contents for Readme or notice files
    - Template based (using twig): extend with your own report

fossology

Features: Using Reporting

User Interface

- All upload items are listed in browser
- Selection of different report items in pop-up menu at every uploaded item, export
  - Debian-copyright, DEP5
  - Readme, notice file
  - SPDX RDF/XML
  - SPDX Tag:Value



Features: Auto Deciding

Review work can be tedious, if there are two scanner, why not letting them do the work?

User Case

- There are several scanners, Nomos with regular expressions, Monk with text similarity.
- If both scanners find the same license, do I need to check?

Solution

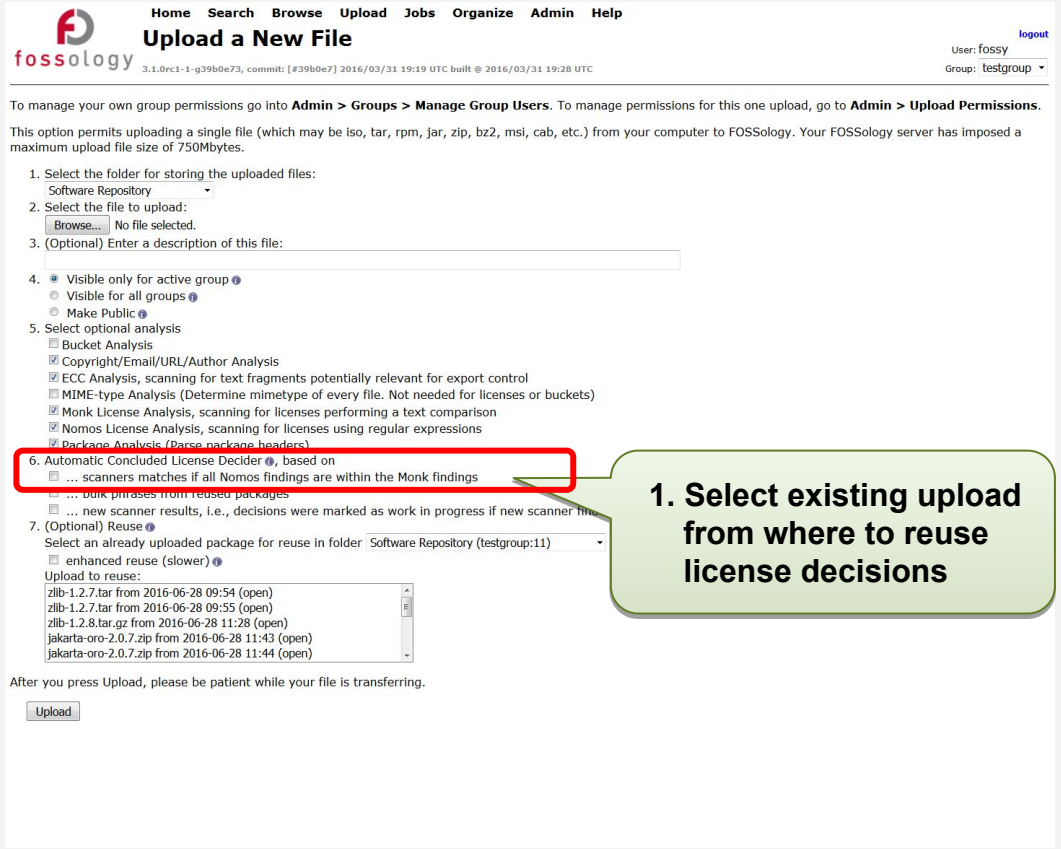
- Decider agent reviews licenses found for this file
- If there are no conflicts in the identified licenses from the two different scanners, you can apply automatically apply decisions
- Example for “no conflict situations” for a decider
  - regular expression and text comparison the same license
  - and all regex matches are inside a text match
  - no other license information detected in file -> no review necessary
- It works with the third Ninka scanner as well

fossology

Features: Using Auto Deciding

User Interface

- At upload you can select auto deciding
- Select the checkbox for the setting auto deciding for Nomos and Monk
- If you have installed Ninka, you can select auto deciding considering result from all three scanners

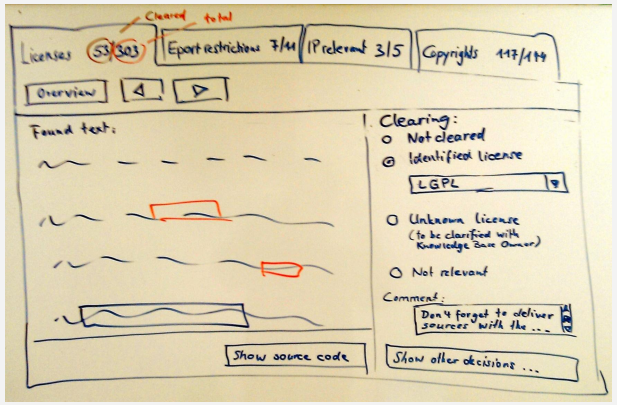


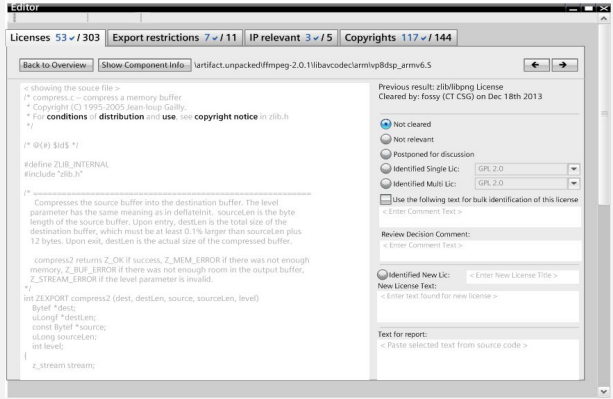
fossology

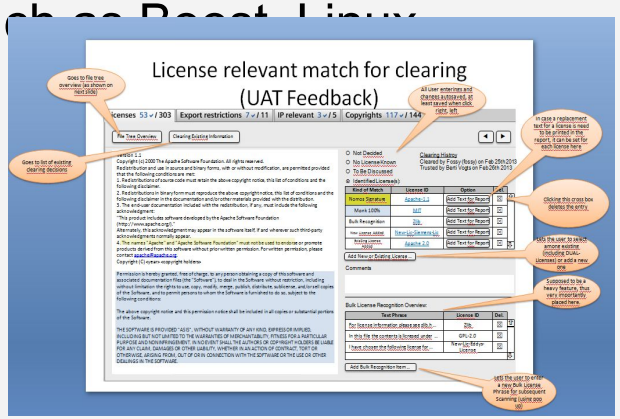
Features: Efficient User Interface

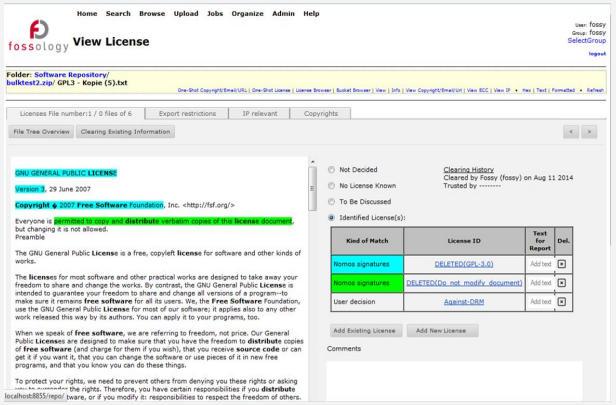
Evolution

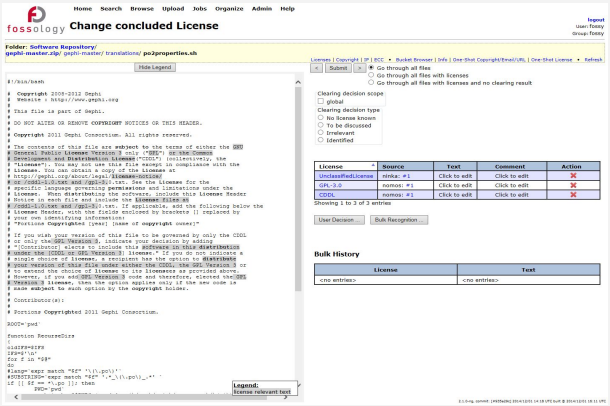
- Most work intensive part: review and correction
- User interface went over many iterations the past years
- Support for large packages (> 10.000 files, such as Boost, Linux Kernel,…)











More Features in FOSSology

- **Workflow integration with command line interfaces**
  - Do uploads and scans from the command line, right from your scripts
  - Or just run individual agents (e.g. for licenses) from command line
  - Schedule activities / integrate them into automated workflows
- **ECC (Export Control and Customs)**
  - Regular expressions searching for export control and customs
- **License data sets import and export**
  - Regular expressions searching for export control and customs
- **Buckets**
  - Define rules for files with license results for collecting them in dedicated lists / buckets

fossology

FOSSology Conclusions

Major Takeaways

- **It is open source software**
  - No vendor lock-in, it can be shared among partners
- **It is a server-based Web application**
  - Allowing for multi user operations, allowing for reuse
- **It is many features for efficiency**
  - Reducing the actual efforts for analyzing OSS components

You like it and would like to contribute?

- Put your logo on the FOSSology homepage if you use it. Send email to fossology-steering@fossology.org
- Report issues on Github: https://github.com/fossology/fossology
- Consider contributing to the code base with your extensions

fossology

Credits: Thanks to Siemens for Creating Content

Internet
https://www.fossology.org

Github
https://github.com/fossology/fossology

Further Links
https://github.com/sw360/sw360portal

fossology