



# FOSSology: Feature Overview

# What is FOSSology?

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

## FOSSology Project

<https://www.fossology.org/>

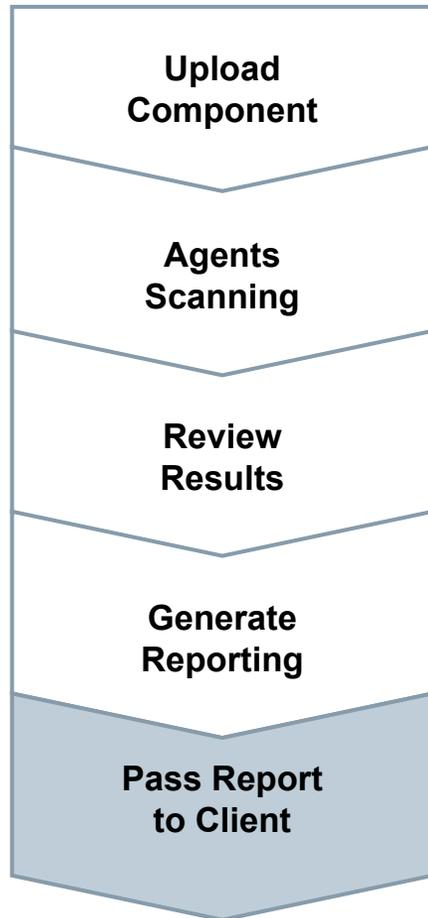
- Published first in 2008, GPL-2.0
- 2015: Linux Foundation collaboration project
- Web server based and command line interfaces
- Scanning agents searching for license and copyright relevant hits (and more ...)
- A multi-user / multi-tenant Web UI for review organizing clearing job

## 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



# 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

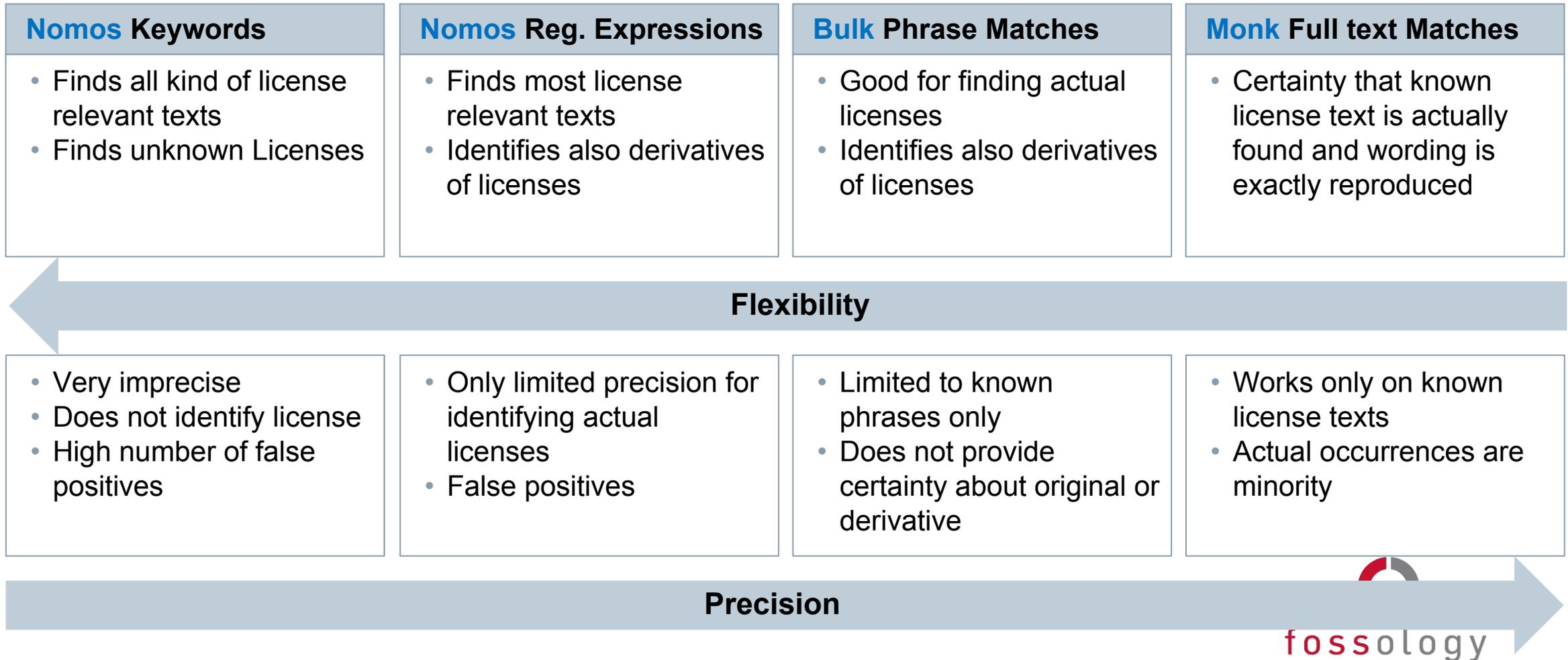
- Review copyright statements and correct in case

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



# 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 phrases ...
  - **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



# 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)

The screenshot shows the fossology web interface. The main content area displays the file view for 'zlib-1.2.7.tar/zlib-1.2.7/gzlib.c'. A red box labeled '1' highlights a copyright notice in the file content. The right sidebar shows the 'Bulk recognition' panel. A red box labeled '2' highlights the 'Bulk scan' button. A red box labeled '3' highlights the 'License' dropdown menu. A red box labeled '4' highlights the 'Schedule Bulk scan' button.

```
zlib.c -- zlib functions common to reading and writing gzip files
Copyright (c) 2004, 2010, 2011, 2012 Mark Adler
/* For conditions of distribution and use, see copyright notice in zlib.h
*/

#include "gzguts.h"

#if defined(_WIN32) && !defined(__BORLANDC__)
# define LSEEK_lseeki64
#else
#if defined(_LARGEFILE64_SOURCE) && !_LFS64_LARGEFILE-0
# define LSEEK_lseek64
#else
# define LSEEK_lseek
#endif
#endif

/* Local functions */
local void gz_reset OF((gz_statep));
local gzFile gz_open OF((const void *, int, const char *));

#if defined UNDER_CE

/* Map the Windows error number in ERROR to a locale-dependent error message string and return a pointer to it. Typically, the values for ERROR_C from GetLastError.

The string pointed to shall not be modified by the application, but may be overwritten by a subsequent call to gz_strerror

The gz_strerror function does not change the current setting of GetLastError. */
char ZLIB_INTERNAL *gz_strerror (error)
    DWORD error;
{
    static char buf[1024];

    wchar_t *msgbuf;
    DWORD lasterr = GetLastError();
    DWORD chars = FormatMessage(FORMAT_MESSAGE_FROM_SYSTEM
        | FORMAT_MESSAGE_ALLOCATE_BUFFER,
        NULL,
        error, /* Default language */
        (LPVOID)msgbuf,
        0,
    );
}

Legend:
license relevant text
```

License	Source	Text	Comment	Action
Zlib-possibility	nomos: #1	Click to edit	Click to edit	[X] [i]

Showing 1 to 1 of 1 entries

User Decision ... Bulk Recognition ...

**Bulk recognition**

Notice: Since punctuation is included in the matching process, periods needs to be included in the phrases if the word just before is included.

Int: New license candidates can be added via menu

OrganizeLicenses

Zlib

Action	License	
remove	Zlib-possibility	[remove row]
add	Zlib	[remove row]

Reference text:

For conditions of distribution and use, see copyright notice in zli

scan whole Upload  ignoreConflicts

Clean text Schedule Bulk scan

# 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



# 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)

The screenshot displays the Fossology License Browser interface. At the top, there is a navigation bar with links for Home, Search, Browse, Upload, Jobs, Organize, Admin, and Help. The user is logged in as 'fossy' with the group 'testgroup'. The current folder is 'Software Repository/zlib-1.2.8.tar.gz/zlib-1.2.8'. The interface shows a table of licenses with columns for Scanner Count, Concluded License Count, License Name, and Actions. A popup window is open, showing a list of available licenses and a list of selected licenses. A red circle labeled '2' highlights the popup, and another red circle labeled '1' highlights the [Edit] link in the Actions column of the table.

Scanner Count	Concluded License Count	License Name	Actions
45	0	Zlib-possibility	[Tag][Edit][Bulk]
16	44	Zlib	[Tag][Edit][Bulk]
10	9	BSL-1.0	[Tag][Edit][Bulk]
4	0	Trademark-ref	[Tag][Edit][Bulk]
4	0	See-file	[Tag][Edit][Bulk]
3	0	See-doc.OTHER	[Tag][Edit][Bulk]
3	0	Public-domain	[Tag][Edit][Bulk]
2	0	GPL	[Tag][Edit][Bulk]
1	0	UnclassifiedLicen	[Tag][Edit][Bulk]
1	0	Perl-possibility	[Tag][Edit][Bulk]
1	0	MIT-style	[Tag][Edit][Bulk]
1	0	LGPL	[Tag][Edit][Bulk]
1	0	info-zip	[Tag][Edit][Bulk]
1	0	GPL-2.0+	[Tag][Edit][Bulk]
1	0	BSD	[Tag][Edit][Bulk]
0	9	Apache-2.0	[Tag][Edit][Bulk]

Summary	Files
Unique licenses	16 236
Unique scanner detected licenses	16 3
Licenses found	94 62
Files with no detected licenses	141 0

# 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!)



# 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)

The screenshot shows the 'Upload a New File' page in FOSSology. The page includes a navigation bar with 'Home', 'Search', 'Browse', 'Upload', 'Jobs', 'Organize', 'Admin', and 'Help'. The user is logged in as 'fossy' in the 'testgroup'. The page title is 'Upload a New File' and the version is '3.1.0rc1-1-g39b0e73, commit: [#39b0e7] 2016/03/31 19:19 UTC built @ 2016/03/31 19:28 UTC'. The main content area contains instructions and a form for uploading a file. The form has seven numbered steps:

1. Select the folder for storing the uploaded files: Software Repository
2. Select the file to upload: [Browse...] No file selected.
3. (Optional) Enter a description of this file:
4.  Visible only for active group  Visible for all groups  Make Public
5. Select optional analysis:
  - Bucket Analysis
  - Copyright/Email/URL/Author Analysis
  - ECC Analysis, scanning for text fragments potentially relevant for export control
  - MIME-type Analysis (Determine mimetype of every file. Not needed for licenses or 1
  - Monk License Analysis, scanning for licenses performing a text comparison
  - Nomos License Analysis, scanning for licenses using regular expressions
  - Package Analysis (Parse package headers)
6. Automatic Concluded License Decider , based on
  - ... scanners matches if all Nomos findings are within the monk findings
  - ... bulk phrases from reused packages
  - ... new scanner results, i.e., decisions were marked as work in progress if new scanner finds add
7. (Optional) Reuse 
  - Select an already uploaded package for reuse in folder Software Repository (testgroup:11)
  - enhanced reuse (slower)
  - Upload to reuse:
    - zlib-1.2.7.tar from 2016-06-28 09:54 (open)
    - zlib-1.2.7.tar from 2016-06-28 09:55 (open)
    - zlib-1.2.8.tar.gz from 2016-06-28 11:28 (open)
    - jakarta-oro-2.0.7.zip from 2016-06-28 11:43 (open)
    - jakarta-oro-2.0.7.zip from 2016-06-28 11:44 (open)

After you press Upload, please be patient while your file is transferring.

Three callouts explain the reuse options:

- 1. Select existing upload from where to reuse license decisions
- 2. Choose same files either by hash-match or by diff-tool with 1-line tolerance
- 3. Choose not only clearing decision from same files, but also identified text phrases

# 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



# 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



The screenshot shows the Fossology Copyright/Email/URL/Author Browser interface. The page title is "Copyright/Email/URL/Author Browser" and the user is logged in as "fossy". The folder being scanned is "Software Repository/zlib-1.2.8.tar.gz/zlib-1.2.8". The interface displays a table of copyright findings with columns for "Count", "Copyright", and a status indicator (a red 'X'). The findings are aggregated by folder level.

Count	Copyright	Status
10	Copyright (C) 2009-2010 Mathias Svensson ( http://result42.com )	X
10	Copyright (C) 1998-2010 Gilles Vollant ( minizip ) ( http://www.winimage.com/zLibDll/minizip.html )	X
7	© Copyright Henrik Ravn 2004	X
7	Copyright (C) 2002-2003 Dmitriy Anisimkov	X
5	Copyright (C) 2007-2008 Even Rouault	X
4	Copyright (C) 1995-2013 Jean-loup Gailly.	X
4	(C) 1995-2013 Jean-loup Gailly & Mark Adler	X
3	Copyright (C) 2004, 2005, 2010, 2011, 2012, 2013 Mark Adler	X
3	Copyright (C) 2003 by Cosmin Truta.	X
3	Copyright (C) 2002-2004 Dmitriy Anisimkov	X
3	Copyright (C) 1995-2003 Mark Adler	X
3	Copyright (C) 1995-1998 Jean-loup Gailly.	X
2	copyrighted -- provided to the public domain Version 1.4 11 December 2005 Mark Adler	X
2	Copyright (C) 2004, 2008, 2012 Mark Adler, all rights reserved	X
2	Copyright (C) 2003 Chris Anderson <christop@charm.net>	X
2	Copyright (C) 2003 Chris Anderson <christop@charm.net>	X
2	Copyright (C) 2003, 2012 Mark Adler	X
2	Copyright (C) 2002-2013 Mark Adler	X
2	Copyright (c) 1996 L. Peter Deutsch	X
2	Copyright (C) 1995-2013 Mark Adler	X
2	Copyright (C) 1995-2013 Jean-loup Gailly and Mark Adler	X
2	Copyright (C) 1995-2011 Mark Adler	X
2	Copyright (C) 1995-2010 Jean-loup Gailly, Brian Raiter and Gilles Vollant.	X
2	Copyright (C) 1995-2008 Mark Adler	X
2	Copyright (C) 1995-2003 Mark Adler	X
2	Copyright (C) 1995-2003 Jean-loup Gailly.	X
2	Copyright 1998-2004 Gilles Vollant - http://www.winimage.com/zLibDll/	X

# 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



# 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

The screenshot shows the Fossology 'Browse' interface. The main content area displays a table titled 'Uploads in Software Repository'. The table has columns for 'Upload Name and Description', 'Status', 'Comment', 'Main licenses', 'Assigned to', 'Upload Date', and 'Priority'. The first row is highlighted in blue and contains the following data: 'copytest.zip', 'open', an empty comment, an empty license field, 'Unassigned', '2016-06-28 15:26:09', and a priority of 1. A callout box with a green border and a white background points to this row, containing the text '1. Select existing upload from where to reuse license decisions'. The interface also includes a 'Folder Navigation' sidebar on the left and a 'Search' bar at the top.

Upload Name and Description	Status	Comment	Main licenses	Assigned to	Upload Date	Priority
copytest.zip	open			Unassigned	2016-06-28 15:26:09	1
zlib-1.2.8.tar.gz	open			Unassigned	2016-06-28 12:28:42	1
1).zip	open			Unassigned	2016-06-28 12:21:36	1
1).zip	open			Unassigned	2016-06-28 12:20:40	1
jakarta-oro-2.0.7.zip	open			testuser	2016-06-28 13:45:39	1
jakarta-oro-2.0.7.zip	open			testuser	2016-06-28 13:30:04	1
jakarta-oro-2.0.7.zip	in progress			Unassigned	2016-06-28 13:23:41	1
jakarta-oro-2.0.7.zip	open			Unassigned	2016-06-28 11:44:10	1
jakarta-oro-2.0.7.zip	open			Unassigned	2016-06-28 11:43:32	1
zlib-1.2.7.tar	open		zlib	Unassigned	2016-06-28 09:55:21	1
zlib-1.2.7.tar	open			Unassigned	2016-06-28 09:54:28	1

# 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



# 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

Home Search Browse Upload Jobs Organize Admin Help

**Upload a New File** User: fossy  
Group: testgroup

3.1.0rc1-1-g39b0e73, commit: [#39b0e7] 2016/03/31 19:19 UTC built @ 2016/03/31 19:28 UTC

To manage your own group permissions go into **Admin > Groups > Manage Group Users**. To manage permissions for this one upload, go to **Admin > Upload Permissions**.

This option permits uploading a single file (which may be iso, tar, rpm, jar, zip, bz2, msi, cab, etc.) from your computer to FOSSology. Your FOSSology server has imposed a maximum upload file size of 750Mbytes.

1. Select the folder for storing the uploaded files:  
Software Repository
2. Select the file to upload:  
Browse... No file selected.
3. (Optional) Enter a description of this file:
4.  Visible only for active group  
 Visible for all groups  
 Make Public
5. Select optional analysis  
 Bucket Analysis  
 Copyright/Email/URL/Author Analysis  
 ECC Analysis, scanning for text fragments potentially relevant for export control  
 MIME-type Analysis (Determine mimetype of every file. Not needed for licenses or buckets)  
 Monk License Analysis, scanning for licenses performing a text comparison  
 Nomos License Analysis, scanning for licenses using regular expressions  
 Package Analysis (Parse package headers)
6. Automatic Concluded License Decider, based on  
 ... scanners matches if all Nomos findings are within the Monk findings  
 ... bulk phrases from reused packages  
 ... new scanner results, i.e., decisions were marked as work in progress if new scanner finished
7. (Optional) Reuse  
Select an already uploaded package for reuse in folder: Software Repository (testgroup:11)  
 enhanced reuse (slower)  
Upload to reuse:  
zlib-1.2.7.tar from 2016-06-28 09:54 (open)  
zlib-1.2.7.tar from 2016-06-28 09:55 (open)  
zlib-1.2.8.tar.gz from 2016-06-28 11:28 (open)  
jakarta-oro-2.0.7.zip from 2016-06-28 11:43 (open)  
jakarta-oro-2.0.7.zip from 2016-06-28 11:44 (open)

After you press Upload, please be patient while your file is transferring.

Upload

**1. Select existing upload from where to reuse license decisions**



# 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 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](mailto:fossology-steering@fossology.org)
- Report issues on Github: <https://github.com/fossology/fossology>
- Consider contributing to the code base with your extensions



# Credits: Thanks to Siemens for Creating Content

© Siemens AG 2016

CC-BY-SA 4.0

<https://creativecommons.org/licenses/by-sa/4.0/>

**Internet**

<https://www.fossology.org>

**Github**

<https://github.com/fossology/fossology>

**Further Links**

<https://github.com/sw360/sw360portal>

