Introducing FOSSology 3.0

Kate Stewart (on behalf of the FOSSology community)
November 10, 2015

Open Source License Compliance by Open Source Software

What is FOSSology:

- a system for handling compliance task?
- a framework for software analysis tools?
- a set of analysis agents that can be integrated into other tools?
- An open source software community project?

FOSSology is:

- ✓a system for handling compliance task
- ✓a framework for software analysis tools
- ✓a set of useful agents that can be integrated into other tools
- ✓an open source software community project

Complying with Open Source Licenses

- Open Source is software that has been made available for use under an open source license.
- Two basic general categories:
  - **Permissive** – minimal requirements about how the software can be redistributed.
    - Apache-2.0, BSD-2-Clause, MIT, …
  - **Copyleft** – permission is given to use, modify and distribute the software as long as the same rights are preserved down the line.
    - GPL-2.0, GPL-3.0, AGPL-3.0, LGPL-2.1, MPL-2.0, CDDL-1.0 …

Open Source Definition

open source initiative®

- Free Redistribution
- Source Code
- Derived Works
- Integrity of the Author’s Source Code
- No Discrimination Against Persons or Groups
- No Discrimination Against Fields of Endeavor
- Distribution of License
- License Must Not be Specific to a Product
- License Must Not Restrict Other Software
- License Must be Technology-Neutral

Mixing Open Source Licenses

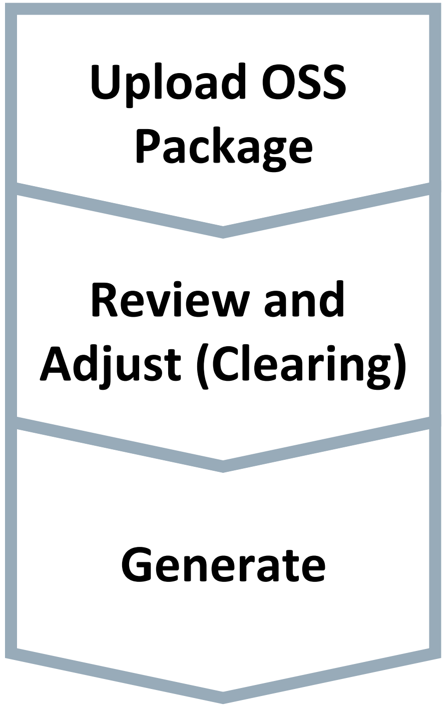
- Not all open source licenses can be used with each other!
- Various projects and distributions provide guidance as to what can be used in that project and combined together.
  - GNU
  - Apache
  - Fedora (which also influences RHEL & CentOS)
  - Debian (which also influences Ubuntu)
  - Android

Compliance?

- Comply with the terms of the license that gives you permission to use the software. Know what you must do, can do, and cannot do.
- When you are distributing software that contains multiple licenses, make sure the licenses are possible to use together.
  - For example, not all copyleft can be used together, and there are specific issues combining permissive and copyleft.
- Are there any government regulations on the content?

➔ Need to accurately know which licenses, copyrights are in the source code, and if there is any content subject to export control, etc.

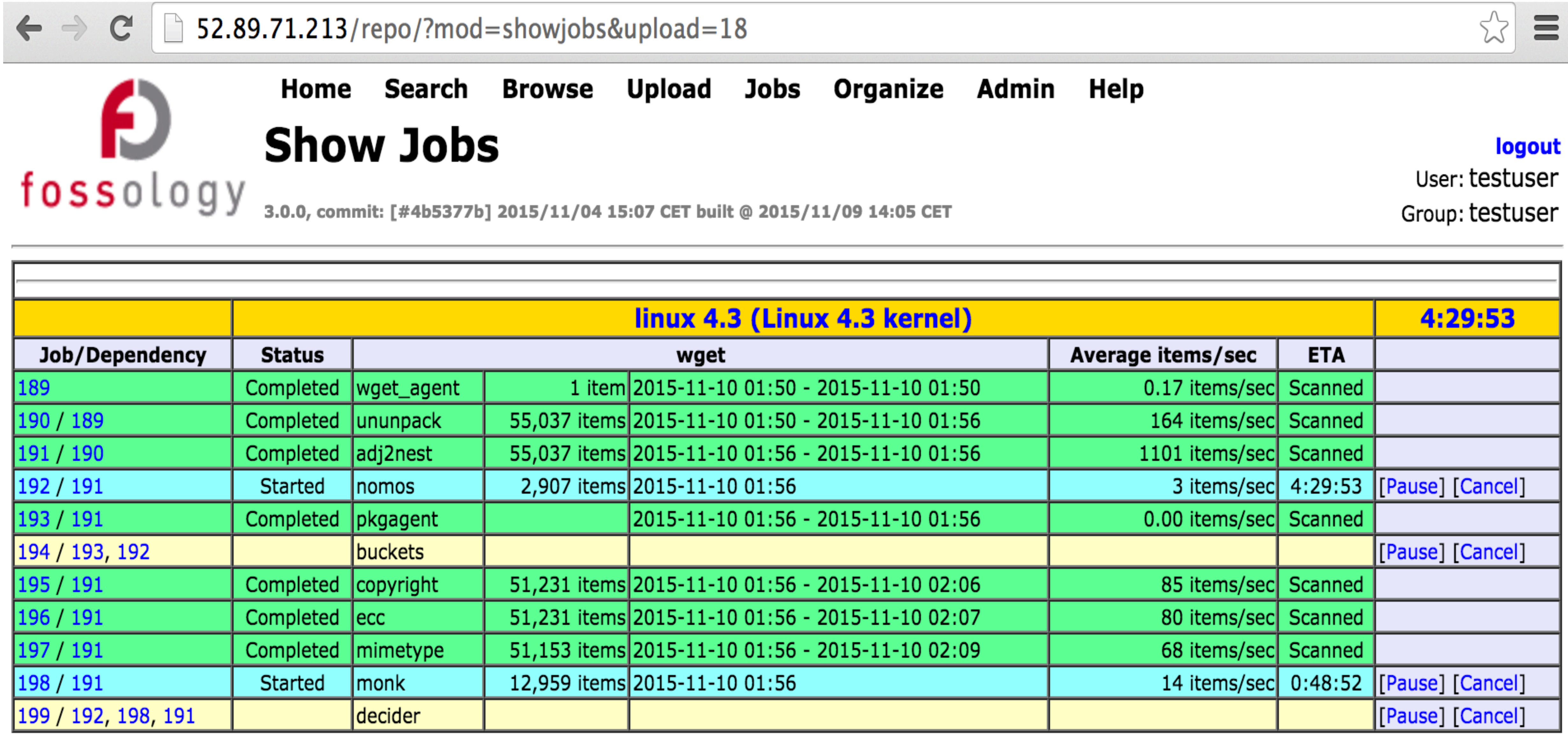
FOSSology: handing compliance tasks



- Upload an open source package to the server
- Select scan agents to analyze the software

- Review the licenses detected that scanners have found
- Compare and correct findings if necessary

- Generate report output
  - list of licenses and copyright holders (spreadsheet, SPDX, …)
  - Set of File Notices applicable
  - Export Control Information



52.89.71.213/repo/?mod=showjobs&upload=18

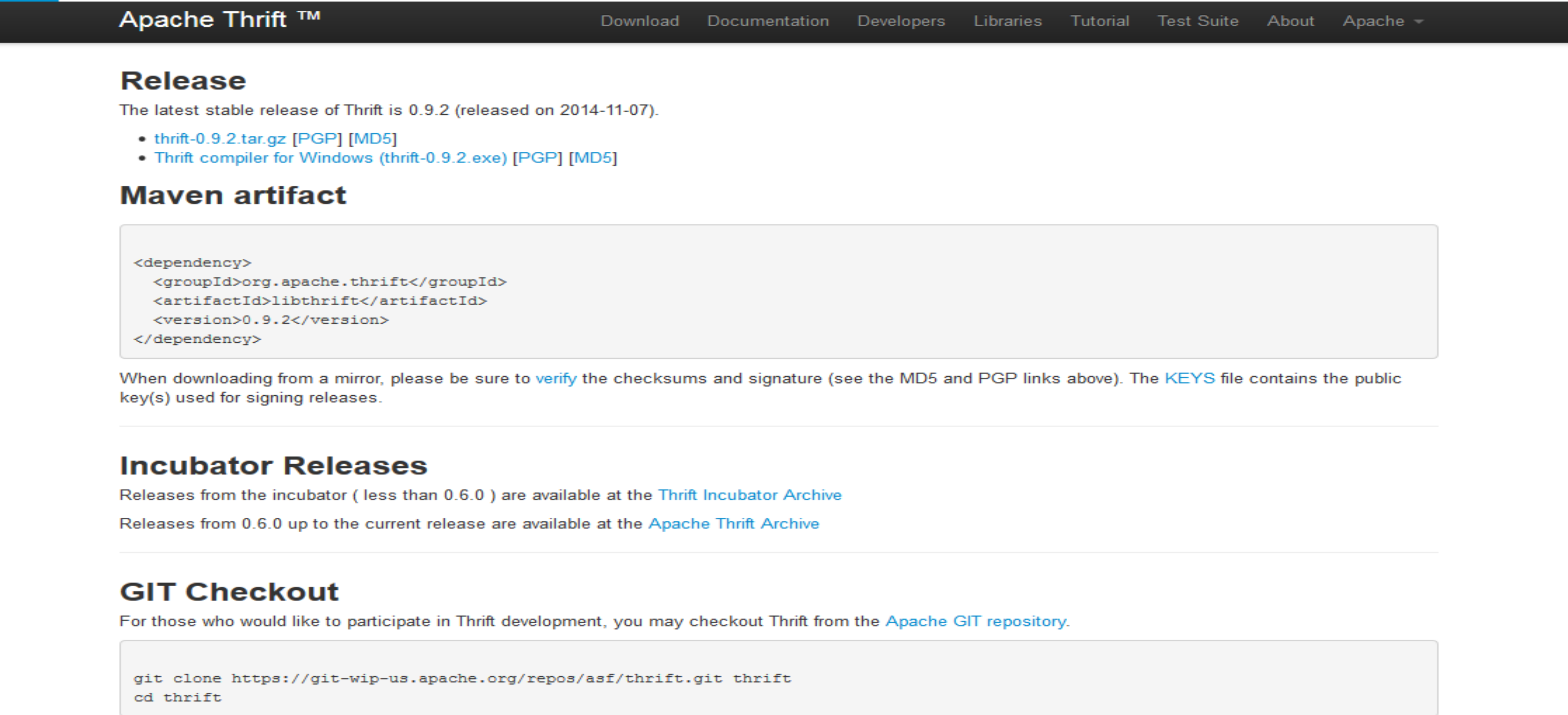
Show Jobs

3.0.0, commit: [#4b5377b] 2015/11/04 15:07 CET built @ 2015/11/09 14:05 CET

User: testuser
Group: testuser

	linux 4.3 (Linux 4.3 kernel)						4:29:53
Job/Dependency	Status	wget			Average items/sec	ETA	
189	Completed	wget_agent	1 item	2015-11-10 01:50 - 2015-11-10 01:50	0.17 items/sec	Scanned	
190 / 189	Completed	ununpack	55,037 items	2015-11-10 01:50 - 2015-11-10 01:56	164 items/sec	Scanned	
191 / 190	Completed	adj2nest	55,037 items	2015-11-10 01:56 - 2015-11-10 01:56	1101 items/sec	Scanned	
192 / 191	Started	nomos	2,907 items	2015-11-10 01:56	3 items/sec	4:29:53	[Pause] [Cancel]
193 / 191	Completed	pkgagent		2015-11-10 01:56 - 2015-11-10 01:56	0.00 items/sec	Scanned	
194 / 193, 192		buckets					[Pause] [Cancel]
195 / 191	Completed	copyright	51,231 items	2015-11-10 01:56 - 2015-11-10 02:06	85 items/sec	Scanned	
196 / 191	Completed	ecc	51,231 items	2015-11-10 01:56 - 2015-11-10 02:07	80 items/sec	Scanned	
197 / 191	Completed	mimetype	51,153 items	2015-11-10 01:56 - 2015-11-10 02:09	68 items/sec	Scanned	
198 / 191	Started	monk	12,959 items	2015-11-10 01:56	14 items/sec	0:48:52	[Pause] [Cancel]
199 / 192, 198, 191		decider					[Pause] [Cancel]

Example:



Release

The latest stable release of Thrift is 0.9.2 (released on 2014-11-07).

- thrift-0.9.2.tar.gz [PGP] [MD5]
- Thrift compiler for Windows (thrift-0.9.2.exe) [PGP] [MD5]

Maven artifact

```
<dependency>
  <groupId>org.apache.thrift</groupId>
  <artifactId>libthrift</artifactId>
  <version>0.9.2</version>
</dependency>
```

When downloading from a mirror, please be sure to verify the checksums and signature (see the MD5 and PGP links above). The KEYS file contains the public key(s) used for signing releases.

Incubator Releases

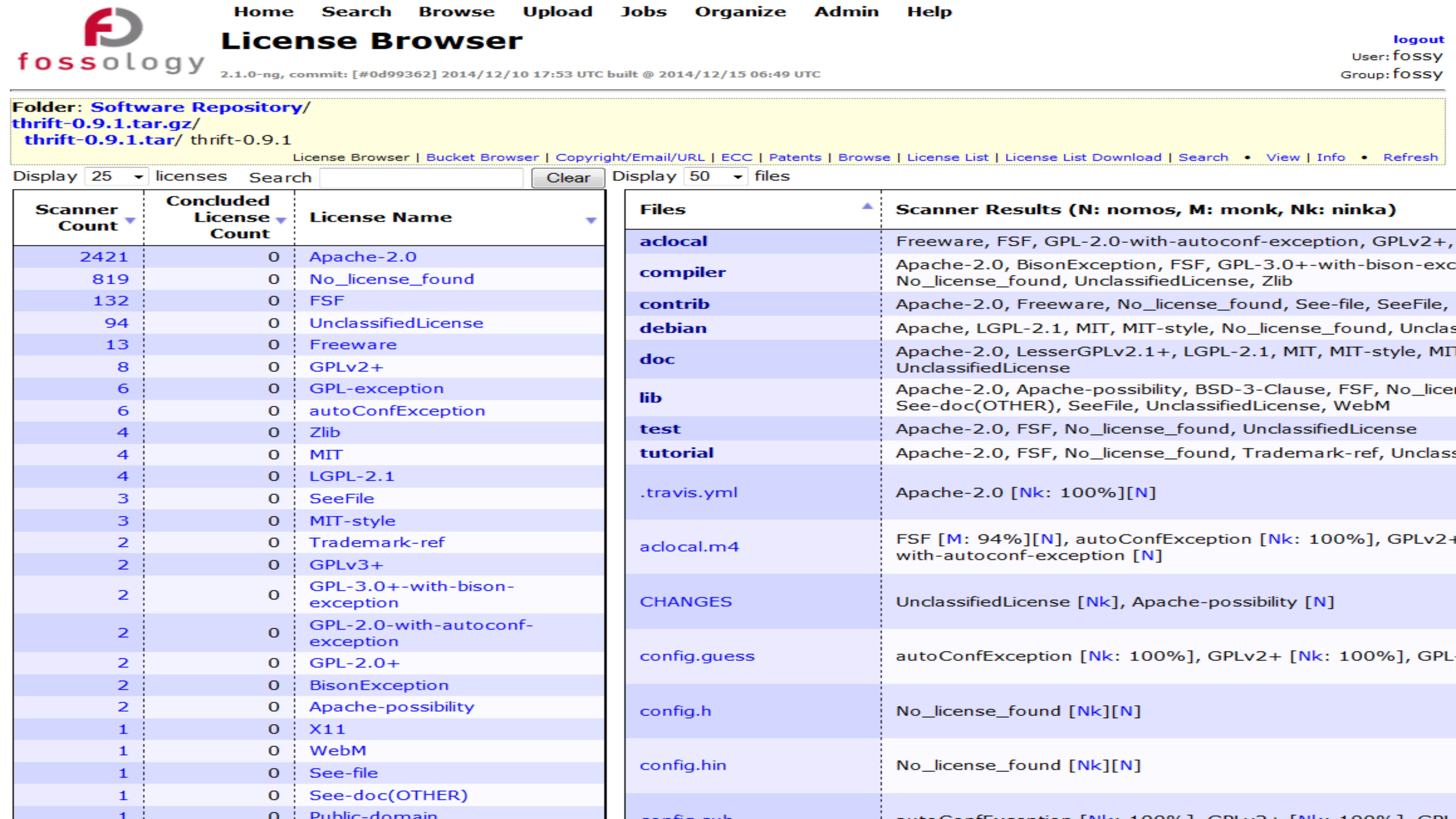
Releases from the incubator (less than 0.6.0) are available at the Thrift Incubator Archive

Releases from 0.6.0 up to the current release are available at the Apache Thrift Archive

GIT Checkout

For those who would like to participate in Thrift development, you may checkout Thrift from the Apache GIT repository.

```
git clone https://git-wip-us.apache.org/repos/asf/thrift.git thrift
cd thrift
```



fossology
Home Search Browse Upload Jobs Organize Admin Help
logout

License Browser

2.1.0-ng, commit: [#0d99362] 2014/12/10 17:53 UTC built @ 2014/12/15 06:49 UTC

User: fossy
Group: fossy

Folder: Software Repository/
thrift-0.9.1.tar.gz/
thrift-0.9.1.tar/ thrift-0.9.1

License Browser | Bucket Browser | Copyright/Email/URL | ECC | Patents | Browse | License List | License List Download | Search • View | Info • Refresh

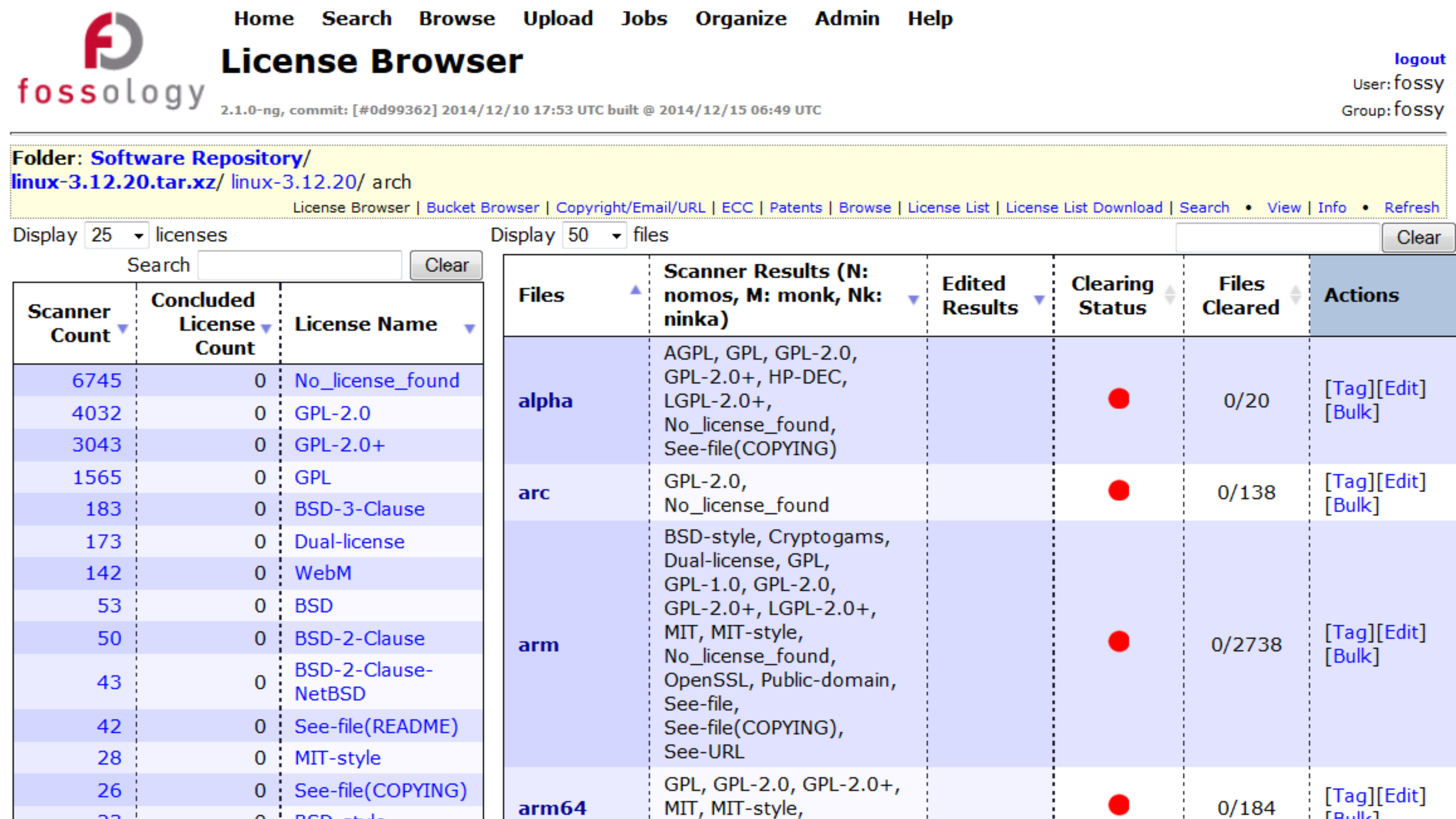
Display 25 licenses Search Clear Display 50 files

Scanner Count	Concluded License Count	License Name
2421	0	Apache-2.0
819	0	No_license_found
132	0	FSF
94	0	UnclassifiedLicense
13	0	Freeware
8	0	GPLv2+
6	0	GPL-exception
6	0	autoConfException
4	0	Zlib
4	0	MIT
4	0	LGPL-2.1
3	0	SeeFile
3	0	MIT-style
2	0	Trademark-ref
2	0	GPLv3+
2	0	GPL-3.0+-with-bison-exception
2	0	GPL-2.0-with-autoconf-exception
2	0	GPL-2.0+
2	0	BisonException
2	0	Apache-possibility
1	0	X11
1	0	WebM
1	0	See-file
1	0	See-doc(OTHER)
1	0	Public-domain

Files	Scanner Results (N: nomos, M: monk, Nk: ninka)
aclocal	Freeware, FSF, GPL-2.0-with-autoconf-exception, GPLv2+,
compiler	Apache-2.0, BisonException, FSF, GPL-3.0+-with-bison-exc No_license_found, UnclassifiedLicense, Zlib
contrib	Apache-2.0, Freeware, No_license_found, See-file, SeeFile,
debian	Apache, LGPL-2.1, MIT, MIT-style, No_license_found, Unclas
doc	Apache-2.0, LesserGPLv2.1+, LGPL-2.1, MIT, MIT-style, MI UnclassifiedLicense
lib	Apache-2.0, Apache-possibility, BSD-3-Clause, FSF, No_lice See-doc(OTHER), SeeFile, UnclassifiedLicense, WebM
test	Apache-2.0, FSF, No_license_found, UnclassifiedLicense
tutorial	Apache-2.0, FSF, No_license_found, Trademark-ref, Unclass
.travis.yml	Apache-2.0 [Nk: 100%][N]
aclocal.m4	FSF [M: 94%][N], autoConfException [Nk: 100%], GPLv2+ with-autoconf-exception [N]
CHANGES	UnclassifiedLicense [Nk], Apache-possibility [N]
config.guess	autoConfException [Nk: 100%], GPLv2+ [Nk: 100%], GPL
config.h	No_license_found [Nk][N]
config.hin	No_license_found [Nk][N]
config.sub	autoConfException [Nk: 100%], GPLv2+ [Nk: 100%], GPL

Clearing: Reviewing and Adjusting

- Why?
  - Indirect references to other files (ie. “License information can be found in README”)
  - Scanners work based on heuristics, and results may still be ambiguous as to which license applies.
  - Files may have been added from other projects under different licenses, and “surprises” may need to be addressed so the terms of the license can be respected
- How?
  - Review results in license browser, investigate “surprises” and understand context, consult with developers, legal counsel, as needed.
  - Record concluded license and save in system, so don’t need to revisit the issues again for next release of a file, effectively “clearing” the file.



License Browser

2.1.0-ng, commit: [#0d99362] 2014/12/10 17:53 UTC built @ 2014/12/15 06:49 UTC

User: fossy
Group: fossy

Folder: **Software Repository**/
linux-3.12.20.tar.xz/ linux-3.12.20/ arch

License Browser | Bucket Browser | Copyright/Email/URL | ECC | Patents | Browse | License List | License List Download | Search • View | Info • Refresh

Display 25 licenses

Search

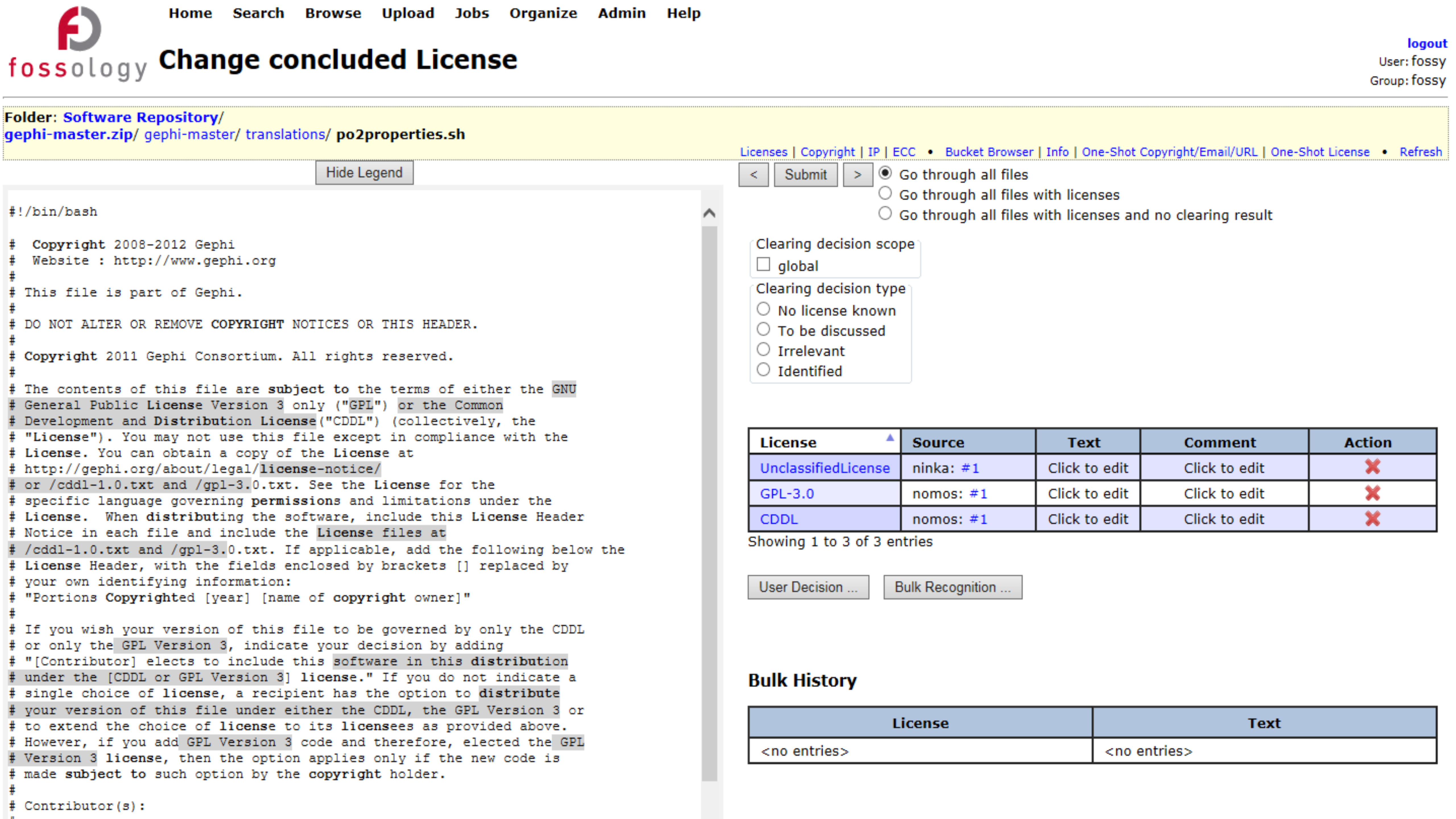
Clear

Scanner Count	Concluded License Count	License Name
6745	0	No_license_found
4032	0	GPL-2.0
3043	0	GPL-2.0+
1565	0	GPL
183	0	BSD-3-Clause
173	0	Dual-license
142	0	WebM
53	0	BSD
50	0	BSD-2-Clause
43	0	BSD-2-Clause-NetBSD
42	0	See-file(README)
28	0	MIT-style
26	0	See-file(COPYING)
23	0	BSD-style

Display 50 files

Clear

Files	Scanner Results (N: nomos, M: monk, Nk: ninka)	Edited Results	Clearing Status	Files Cleared	Actions
alpha	AGPL, GPL, GPL-2.0, GPL-2.0+, HP-DEC, LGPL-2.0+, No_license_found, See-file(COPYING)		●	0/20	[Tag][Edit] [Bulk]
arc	GPL-2.0, No_license_found		●	0/138	[Tag][Edit] [Bulk]
arm	BSD-style, Cryptogams, Dual-license, GPL, GPL-1.0, GPL-2.0, GPL-2.0+, LGPL-2.0+, MIT, MIT-style, No_license_found, OpenSSL, Public-domain, See-file, See-file(COPYING), See-URL		●	0/2738	[Tag][Edit] [Bulk]
arm64	GPL, GPL-2.0, GPL-2.0+, MIT, MIT-style,		●	0/184	[Tag][Edit] [Bulk]



Change concluded License

User: fossy
Group: fossy

Folder: Software Repository/
gephi-master.zip/ gephi-master/ translations/ **po2properties.sh**

Licenses | Copyright | IP | ECC • Bucket Browser | Info | One-Shot Copyright/Email/URL | One-Shot License • Refresh

```
#!/bin/bash

#  Copyright 2008-2012 Gephi
#  Website : http://www.gephi.org
#
# This file is part of Gephi.
#
# DO NOT ALTER OR REMOVE COPYRIGHT NOTICES OR THIS HEADER.
#
# Copyright 2011 Gephi Consortium. All rights reserved.
#
# The contents of this file are subject to the terms of either the GNU
# General Public License Version 3 only ("GPL") or the Common
# Development and Distribution License("CDDL") (collectively, the
# "License"). You may not use this file except in compliance with the
# License. You can obtain a copy of the License at
# http://gephi.org/about/legal/license-notice/
# or /cddl-1.0.txt and /gpl-3.0.txt. See the License for the
# specific language governing permissions and limitations under the
# License.  When distributing the software, include this License Header
# Notice in each file and include the License files at
# /cddl-1.0.txt and /gpl-3.0.txt. If applicable, add the following below the
# License Header, with the fields enclosed by brackets [] replaced by
# your own identifying information:
# "Portions Copyrighted [year] [name of copyright owner]"
#
# If you wish your version of this file to be governed by only the CDDL
# or only the GPL Version 3, indicate your decision by adding
# "[Contributor] elects to include this software in this distribution
# under the [CDDL or GPL Version 3] license." If you do not indicate a
# single choice of license, a recipient has the option to distribute
# your version of this file under either the CDDL, the GPL Version 3 or
# to extend the choice of license to its licensees as provided above.
# However, if you add GPL Version 3 code and therefore, elected the GPL
# Version 3 license, then the option applies only if the new code is
# made subject to such option by the copyright holder.
#
# Contributor(s):
```

● Go through all files
○ Go through all files with licenses
○ Go through all files with licenses and no clearing result

Clearing decision scope
□ global

Clearing decision type
○ No license known
○ To be discussed
○ Irrelevant
○ Identified

License	Source	Text	Comment	Action
UnclassifiedLicense	ninka: #1	Click to edit	Click to edit	✗
GPL-3.0	nomos: #1	Click to edit	Click to edit	✗
CDDL	nomos: #1	Click to edit	Click to edit	✗

Showing 1 to 3 of 3 entries

Bulk History

License	Text
<no entries>	<no entries>

FOSSology is:

- ✓a system for handling compliance task
- ✓a framework for software analysis tools
- ✓a set of useful agents that can be integrated into other tools
- ✓an open source software community project

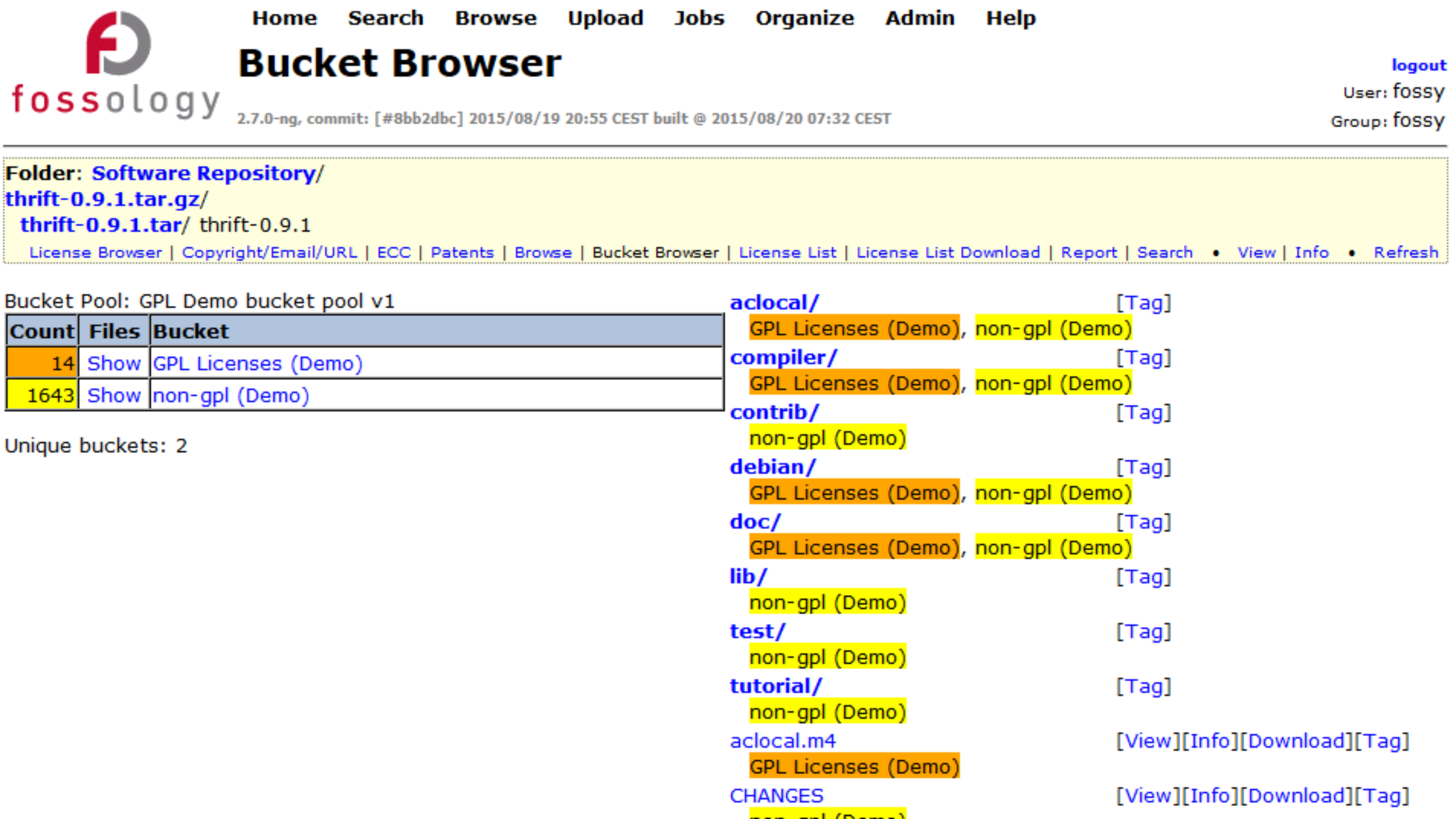
fossology

FOSSology: framework for analysis tools

- Upload files and unpack them into their components
- Browse file contents and meta data
- Scan file contents for key information with agents
- Review detected information and annotate
- Create report files based on the information you’re interested in and export them

Framework Components

- Infrastructure
  - Repository, Database, Web server, PHP
- Agents and Jobs
  - Job Scheduler
  - Plugins: file comparing, diff, ..
  - Scanners: Nomos, Monk, Ninka, Copyright, ECC
  - Other: package, mimetype, maintenance, buckets
- User Interface
  - Access control, permissions and groups, tagging & filters.



Bucket Browser

2.7.0-ng, commit: [#8bb2dbc] 2015/08/19 20:55 CEST built @ 2015/08/20 07:32 CEST

User: fossy
Group: fossy

Folder: **Software Repository/**
thrift-0.9.1.tar.gz/
thrift-0.9.1.tar/ thrift-0.9.1

License Browser | Copyright/Email/URL | ECC | Patents | Browse | Bucket Browser | License List | License List Download | Report | Search • View | Info • Refresh

Bucket Pool: GPL Demo bucket pool v1

Count	Files	Bucket
14	Show	GPL Licenses (Demo)
1643	Show	non-gpl (Demo)

Unique buckets: 2

aclocal/ [Tag]
GPL Licenses (Demo), non-gpl (Demo)

compiler/ [Tag]
GPL Licenses (Demo), non-gpl (Demo)

contrib/ [Tag]
non-gpl (Demo)

debian/ [Tag]
GPL Licenses (Demo), non-gpl (Demo)

doc/ [Tag]
GPL Licenses (Demo), non-gpl (Demo)

lib/ [Tag]
non-gpl (Demo)

test/ [Tag]
non-gpl (Demo)

tutorial/ [Tag]
non-gpl (Demo)

aclocal.m4 [View][Info][Download][Tag]
GPL Licenses (Demo)

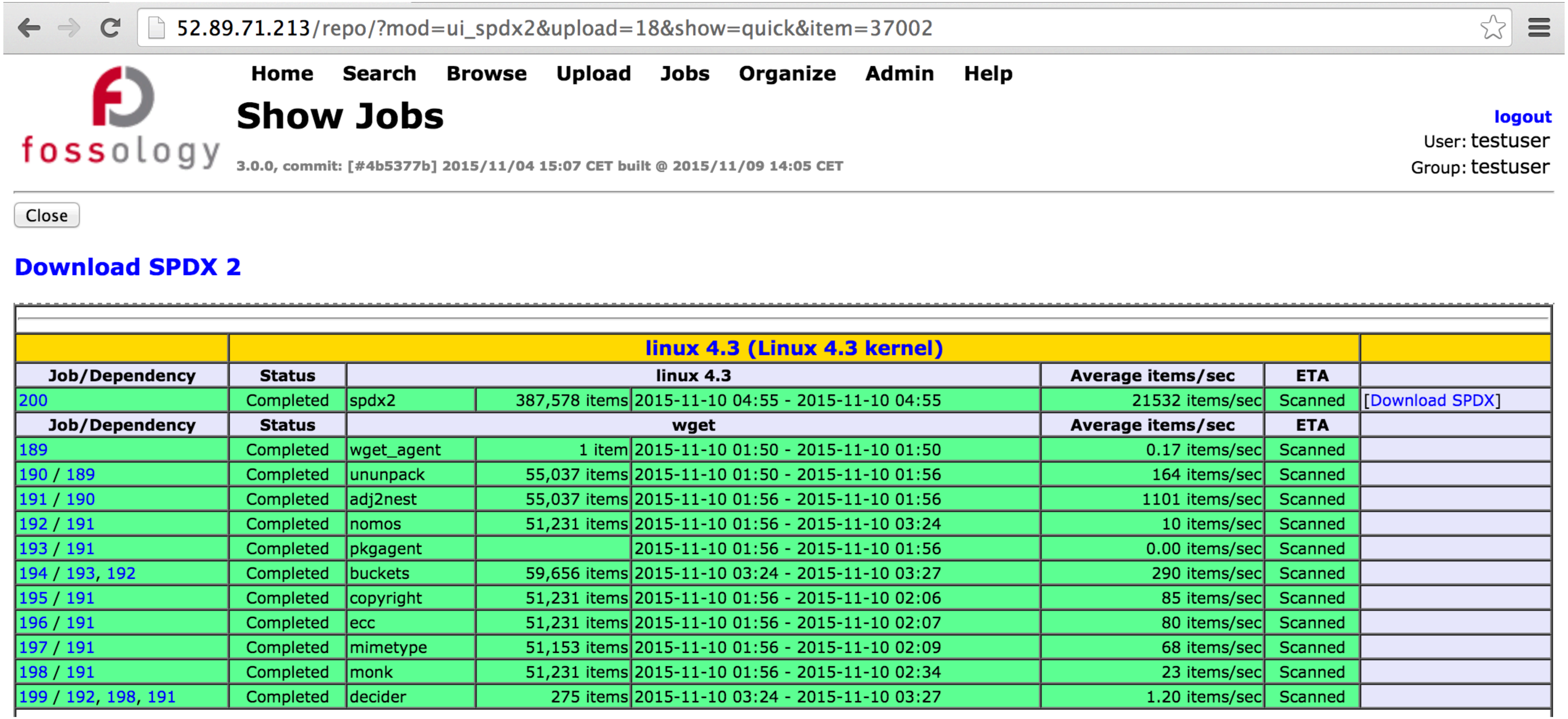
CHANGES [View][Info][Download][Tag]
non-gpl (Demo)

FOSSology 3.0: Workflow Improvements

- **User Interface**
  - Efficient reviewing of licenses for determining concluded licenses
  - Editing copyrights
  - Folder navigation
- **License Analysis Support**
  - Added the concept of candidate licenses
  - Ability to now re-use license decisions
  - Bulk assignment of license decisions based on text phrases
  - Apply license decision automatically when Monk and Nomos find the same license in same text area

FOSSology 3.0: System Interactions

- **Command line support (scriptable)**
  - Uploads and scans from the command line
  - Schedule activities / integrate them into automated workflows
  - Run individual agents (e.g. for licenses)
- **New Report Types Supported**
  - Export Bill of Material (SPDX 2.0 RDFa)
  - Generate README/COPYING text files to support distributions



52.89.71.213/repo/?mod=ui_spdx2&upload=18&show=quick&item=37002

Show Jobs

3.0.0, commit: [#4b5377b] 2015/11/04 15:07 CET built @ 2015/11/09 14:05 CET

logout
User: testuser
Group: testuser

Download SPDX 2

	linux 4.3 (Linux 4.3 kernel)						
Job/Dependency	Status	linux 4.3			Average items/sec	ETA	
200	Completed	spdx2	387,578 items	2015-11-10 04:55 - 2015-11-10 04:55	21532 items/sec	Scanned	[Download SPDX]
Job/Dependency	Status	wget			Average items/sec	ETA	
189	Completed	wget_agent	1 item	2015-11-10 01:50 - 2015-11-10 01:50	0.17 items/sec	Scanned	
190 / 189	Completed	ununpack	55,037 items	2015-11-10 01:50 - 2015-11-10 01:56	164 items/sec	Scanned	
191 / 190	Completed	adj2nest	55,037 items	2015-11-10 01:56 - 2015-11-10 01:56	1101 items/sec	Scanned	
192 / 191	Completed	nomos	51,231 items	2015-11-10 01:56 - 2015-11-10 03:24	10 items/sec	Scanned	
193 / 191	Completed	pkgagent		2015-11-10 01:56 - 2015-11-10 01:56	0.00 items/sec	Scanned	
194 / 193, 192	Completed	buckets	59,656 items	2015-11-10 03:24 - 2015-11-10 03:27	290 items/sec	Scanned	
195 / 191	Completed	copyright	51,231 items	2015-11-10 01:56 - 2015-11-10 02:06	85 items/sec	Scanned	
196 / 191	Completed	ecc	51,231 items	2015-11-10 01:56 - 2015-11-10 02:07	80 items/sec	Scanned	
197 / 191	Completed	mimetype	51,153 items	2015-11-10 01:56 - 2015-11-10 02:09	68 items/sec	Scanned	
198 / 191	Completed	monk	51,231 items	2015-11-10 01:56 - 2015-11-10 02:34	23 items/sec	Scanned	
199 / 192, 198, 191	Completed	decider	275 items	2015-11-10 03:24 - 2015-11-10 03:27	1.20 items/sec	Scanned	

FOSSology 3.0: Agent Improvements

- **New Optional Scanner - Ninka**
  - Ninka has been integrated, and can be run at upload or at scheduling
- **Detecting Export Control and Customs Information**
  - Extension to copyright agent’s regular expressions to detect ECC
- **Command Line Use**
  - Individual Scanning Agents (monk, nomos, ninka, etc.) can be invoked from the command line.

FOSSology is:

- ✓a system for handling compliance task
- ✓a framework for software analysis tools
- ✓a set of useful agents that can be integrated into other tools
- ✓An open source software community project

fossology

Agent: Nomos

Overview:

- Nomos is named after the Greek word for “law”.
- License identification is done using short phrases (regular expressions) and heuristics.   The heuristics for detecting phrases must be found in (or out of) proximity to another phrase or phrases.
- This scanner currently recognizes more than 659 licenses.

History:

- This is one of the original scanners used by HP that was the foundation for FOSSology when it was open sourced in 2007, and has been maintained and enhanced through the years

Agent: Monk

Overview:

- Monk looks for complete licenses (as defined in the license database) and reports the percentage of match to that reference version.   It is useful with license highlighting, as it allows you to see exactly what was added or removed from a license.
- Text similarity is based on the Jaccard index

History:

- This scanner was contributed by Siemens and TNGtech to the FOSSology project. It was made available as part of the 2.6 release in 2014.

Agent: Ninka

Overview:

- Ninka is sentence-based, and provides a simple way to identify open source licenses in a source code file.
- Ninka was designed to be lightweight, fast, and if it isn't sure about the license, not to guess.

History:

- This scanner was developed by a team of researchers studying automatic license detection in 2010 [1]
- Ninka has been incorporated into the FOSSology Project as part of the 3.0 release in 2015. It is used in other open source projects as well.

[1] A sentence-matching method for automatic license identification of source code files by D.M. German, Y. Manabe and K. Inoue. In Proceedings of the IEEE/ACM international Conference on Automated Software Engineering (ASE) 2010, pp: 437–446

Agent: Copyright

Overview:

- IP compliance teams also scan source to pull out copyright statements, email addresses, authorship statements and URL's helpful to identify IP ownership.
- This is a regular expression heuristic based source scanner, and false positives have been improved over time, but may still occur.
- Regular expressions used for copyright and ECC analysis can be adjusted by the configuration files

History:

- This scanner was made available as part of the 1.2 release in July 2010.
- Regular expressions for detecting export compliance information and support for a separate editing display was added in 3.0 release.

FOSSology is:

- ✓a system for handling compliance task
- ✓a framework for software analysis tools
- ✓a set of useful agents that can be integrated into other tools
- ✓a open source software community project

fossology

Mission:
Advancing Open Source License Compliance

- FOSSology is an open source license compliance software system and toolkit.
- As a toolkit you can run license, copyright and export control scans from the command line.
- As a system, a database and a web user interface are provided to give you a compliance workflow.   In one click you can generate an SPDX file, or a README with the copyright notices from your software.
- FOSSology deduplication means that you can scan an entire distro, submit a new version, and only the changed files will get rescanned. This is a big time saver for large projects.

Community Project: A bit of history

- HP created FOSSology in 2007 to help with their internal compliance activities and made it Open Source under the GPL-2.0 in December 2007.
- Periodic releases over the years, adding more features, agents (recognizing licenses), and supported platform images (Debian, Ubuntu, Fedora, RHEL, CentOS, Docker Container, Vagrant) as well as the source is available with 2.6 released on October 10th, 2014.
- January 15, 2015, with 2.6.2 release, source officially moved to github
- HP transferred stewardship of FOSSology to the Linux Foundation and announced at ELC in Dublin on October 5th.
- November 5, Release 3.0 tagged in github repository…. and the images are building now!

FOSSology Community

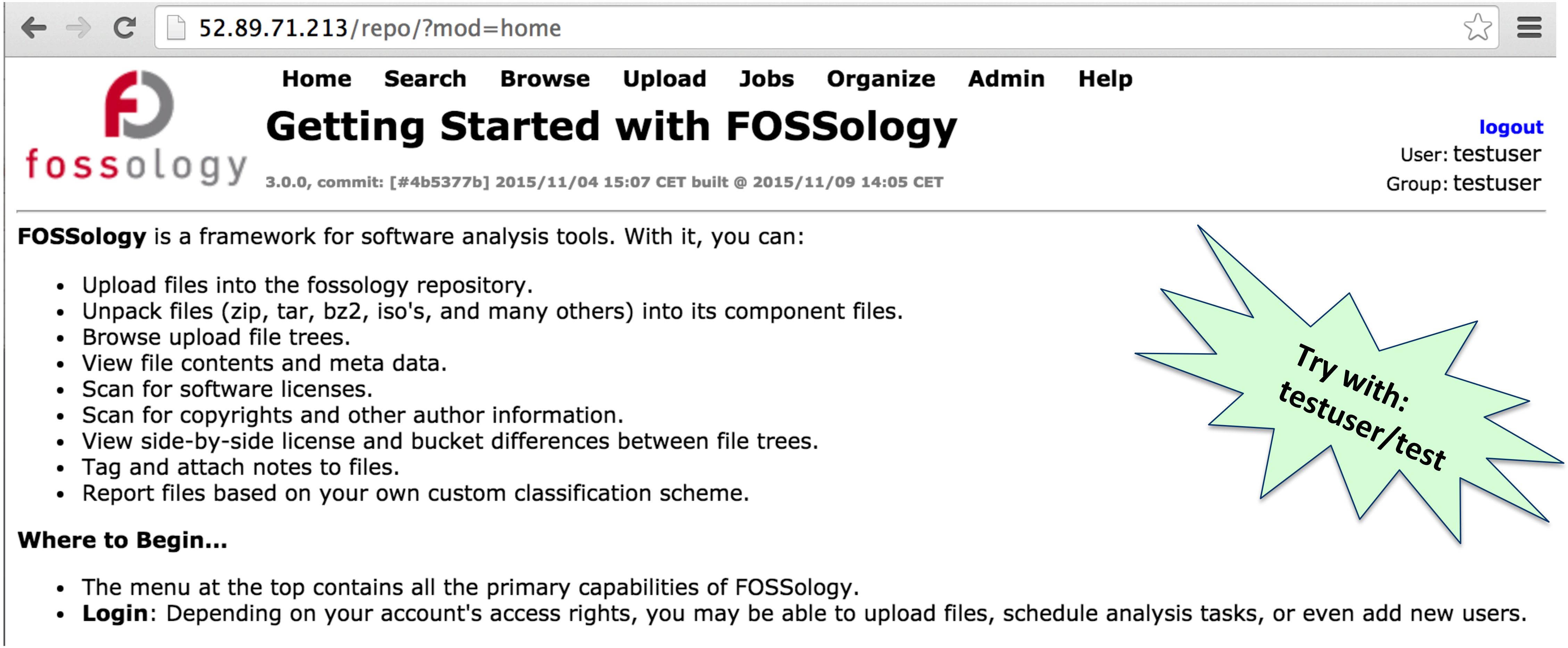
tieto
software freedom conservancy
VALIDOS
hp
BERGISCHE UNIVERSITÄT WUPPERTAL
WIND™ AN INTEL COMPANY
LINUX FOUNDATION
tl;drLegal
SIEMENS
HH PARTNERS Attorneys-at-law Ltd
Antelink
RogueWave SOFTWARE
Bittium
UNIVERSITY OF Nebraska Omaha
SULLIVAN & WORCESTER
Tjaldur Software Governance Solutions

and more to be added …

FOSSology Project Information

- **Project Home:** www.fossology.org
  - In transition to new server, will have new look & feel soon.
  - Volunteers needed to help move key web pages over to server and github wiki (github.com/fossology/fossology/wiki)
- **FOSSology Source:** github.com/fossology/fossology
  - Status: tagged 3.0 last week,
- **Mail lists:** http://lists.fossology.org/
  - Status: Will be moved to new server.

Try FOSSology



Getting Started with FOSSology

3.0.0, commit: [#4b5377b] 2015/11/04 15:07 CET built @ 2015/11/09 14:05 CET

User: testuser
Group: testuser

FOSSology is a framework for software analysis tools. With it, you can:

- Upload files into the fossology repository.
- Unpack files (zip, tar, bz2, iso's, and many others) into its component files.
- Browse upload file trees.
- View file contents and meta data.
- Scan for software licenses.
- Scan for copyrights and other author information.
- View side-by-side license and bucket differences between file trees.
- Tag and attach notes to files.
- Report files based on your own custom classification scheme.

Where to Begin...

- The menu at the top contains all the primary capabilities of FOSSology.
- **Login**: Depending on your account's access rights, you may be able to upload files, schedule analysis tasks, or even add new users.

Try with: testuser/test

Questions?

fossology