



FOSSology

shaheem.azmal@siemens.com <Shaheem Azmal M MD>

mishra.gaurav@siemens.com <Gaurav Mishra>

Thank You to Our Sponsors and Hosts!



Without them, this Conference couldn't happen.

Agenda

- **FOSSology introduction**
- **Key features**
- **FOSSology Scanning in CI**
- **Short demo**
- **Conclusion**

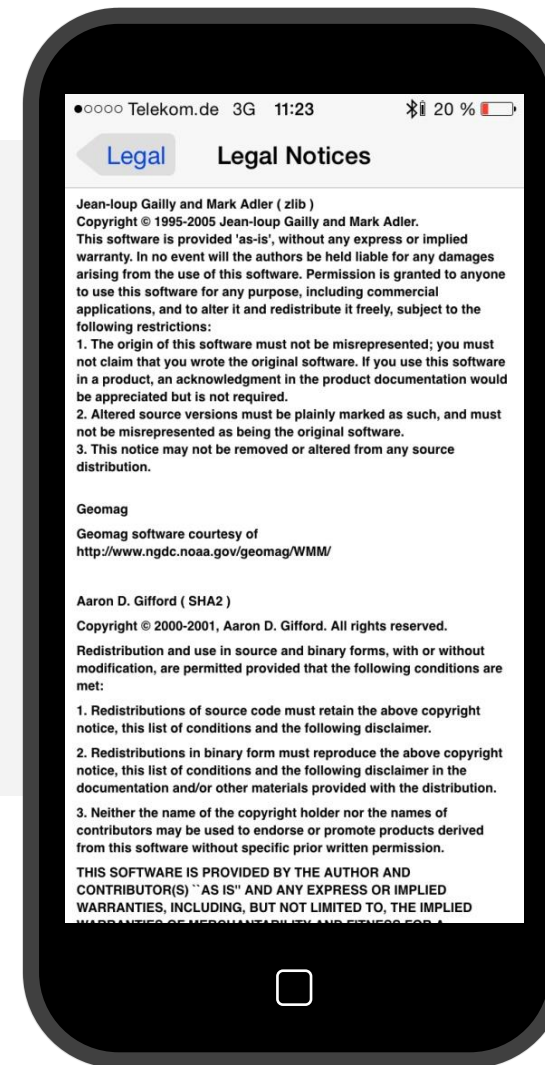


The Problem Actually

You know these examples

Distributing open source software requires to

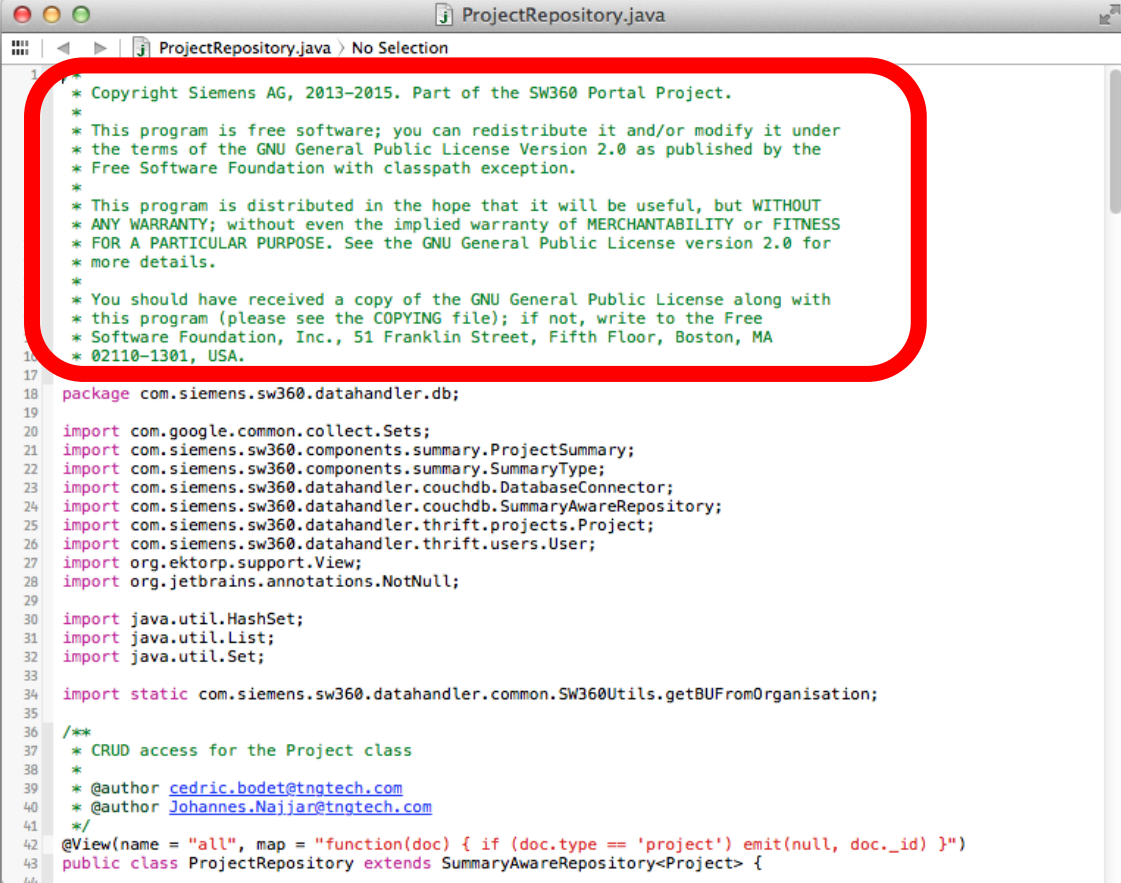
- Provide licenses of involved software
- Provide copyright statements of involved authors
- Provide disclaimers
- ... and much more



It is about finding licenses

Finding Licenses

- License texts
- References to licenses
- Written texts explaining licensing
- License relevant statements



```
ProjectRepository.java  
ProjectRepository.java > No Selection  
1  
2 * Copyright Siemens AG, 2013-2015. Part of the SW360 Portal Project.  
3 * This program is free software; you can redistribute it and/or modify it under  
4 * the terms of the GNU General Public License Version 2.0 as published by the  
5 * Free Software Foundation with classpath exception.  
6 *  
7 * This program is distributed in the hope that it will be useful, but WITHOUT  
8 * ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS  
9 * FOR A PARTICULAR PURPOSE. See the GNU General Public License version 2.0 for  
10 * more details.  
11 *  
12 * You should have received a copy of the GNU General Public License along with  
13 * this program (please see the COPYING file); if not, write to the Free  
14 * Software Foundation, Inc., 51 Franklin Street, Fifth Floor, Boston, MA  
15 * 02110-1301, USA.  
16  
17 package com.siemens.sw360.datahandler.db;  
18  
19  
20 import com.google.common.collect.Sets;  
21 import com.siemens.sw360.components.summary.ProjectSummary;  
22 import com.siemens.sw360.components.summary.SummaryType;  
23 import com.siemens.sw360.datahandler.couchdb.DatabaseConnector;  
24 import com.siemens.sw360.datahandler.couchdb.SummaryAwareRepository;  
25 import com.siemens.sw360.datahandler.thrift.projects.Project;  
26 import com.siemens.sw360.datahandler.thrift.users.User;  
27 import org.ektorp.support.View;  
28 import org.jetbrains.annotations.NotNull;  
29  
30 import java.util.HashSet;  
31 import java.util.List;  
32 import java.util.Set;  
33  
34 import static com.siemens.sw360.datahandler.common.SW360Utils.getBUFFromOrganisation;  
35  
36 /**  
37  * CRUD access for the Project class  
38  *  
39  * @author cedric.bodet@tngtech.com  
40  * @author Johannes.Najjar@tngtech.com  
41  */  
42 @View(name = "all", map = "function(doc) { if (doc.type == 'project') emit(null, doc._id) }")  
43 public class ProjectRepository extends SummaryAwareRepository<Project> {  
44
```

What is FOSSology?

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

FOSSology Project

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

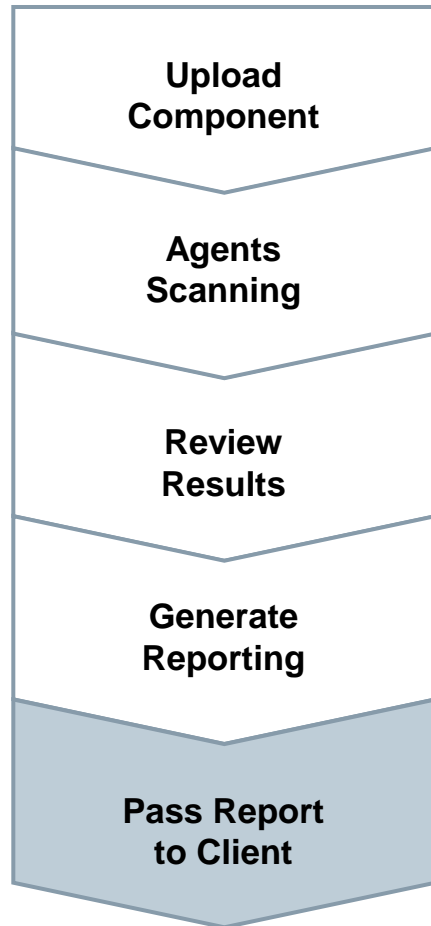
- Published first in 2008, GPL-2.0
- 2015: Linux Foundation collaboration project
- Web server based and command line interfaces
- Scanning agents searching for license and copyright relevant hits (and more ...)
- A multi-user / multi-tenant Web UI for review organizing clearing job

FOSSology Development

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

- Standard Web application stack:
 - Linux, Apache 2, PostgreSQL, PHP, Python
- Web-based UI in PHP, but scanners written in C / C++
- Two ways to interact:
 - Web user interface
 - Command line utilities

How does FOSSology work?



- Uploading source code archive (*.zip, *.tar.gz, etc)
- Agents scan for license relevant text
- Copyrights, Export Control (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
 - Word reporting
 - CycloneDX (soon)



FOSSology – It is about Overview

High Level and Drill Down

- Aggregation
 - Folder hierarchy of license findings
 - License-statement oriented view on files
 - Copyright aggregation
- Drill down
 - Navigate into folders
 - Filtering
 - Identify “the single” file

The screenshot displays the FOSSology interface. On the left, a table titled 'licenses' shows an aggregation of license findings. On the right, a table titled 'files (tree view or flat)' shows a list of files with their associated scanner results and edited results. An orange callout box points to a specific file in the 'files' table.

Scanner Count	Concluded License Count	License Name
7702	8018	EPL-1.0
2339	52	Apache-2.0
275	0	MPL-2.0
112	0	MPL-1.1
110	0	LGPL-2.0+
110	0	Dual-license
64	0	Apache-possibility
57	23	W3C
51	50	MIT
49	0	GPL
34	0	W3C-IP
24	0	W3C-possibility
18	0	Public-domain
13	11	BSD-3-Clause
12	0	WebM
8	8	Apache-1.1
7	8	Apache-1.1-variant-jakarta-oro
6	0	CPL-1.0
5	0	W3C-style
4	0	UnclassifiedLicense
4	0	CPL-0.5
4	0	BSD-style
3	0	Microsoft-possibility
2	0	libtiff
2	0	Unicode

Files	Scanner Results (N: nomos, M: monk, Nk: ninka, I: reportImport)	Edited Results
com.lowagie.text_2.1.7.v201004222200.jar	Adobe, Apache-2.0, APAFML, BSD-3-Clause, CUA-OPL-1.0, EPL-1.0, LGPL-2.0, libtiff, MIT-style, MPL-1.1, No_license_found, Permission Notice, Unicode	EPL-1.0, Apache-2.0
com.lowagie.text.source_2.1.7.v201004222200.jar	Apache-2.0, BSD-3-Clause, CUA-OPL-1.0, Dual-license, EPL-1.0, LGPL, LGPL-2.0+, libtiff, MIT, MIT-style, MPL, MPL-1.1, No_license_found, Permission Notice, Public-domain, Unicode, WebM	MIT, BSD-3-Clause, EPL-1.0
javax.wsdl_1.5.1.v201012040544.jar	GPL-0.5, GPL-1.0, EPL-1.0	
javax.xml.rpc_1.1.0.v201209140446.jar		Apache-2.0
javax.xml.soap_1.2.0.v201005080501.jar		Apache-2.0
javax.xml.stream_1.0.1.v201004272200.jar		Apache-2.0
org.apache.axis_1.4.0.v201411182030.jar	Apache-2.0, Apache-possibility, EPL-1.0, No_license_found, W3C-possibility	Apache-2.0
org.apache.batik.bridge_1.6.0.v201011041432.jar	Apache-2.0, Apache-possibility, EPL-1.0, No_license_found, Public-domain, W3C, W3C-IP	W3C, EPL-1.0, Apache-2.0
org.apache.batik.bridge.source_1.6.0.v201011041432.jar	Apache-2.0, Apache-possibility, EPL-1.0, No_license_found, Public-domain, W3C, W3C-IP	W3C, EPL-1.0, Apache-2.0

Showing 1 to 25 of 42 licenses Page 1 of 2

Hint: Click on the license name to search for where the license is found in the file listing.

Recursive unpacking of files too!

FOSSology – Review Findings

Specialized in Review

- Single file review
 - Highlighting of license relevant content
 - Reference text comparison
 - License statement decisions on statement level (“bulk scan”)

Close Cleared: 8030/11520

Hide Legend

```
/*
 * $Id: ImgJBIG2.java,v 1.1.2.1 2010/03/05 21:12:09 rbrooks Exp $
 *
 * Copyright 2009 by Nigel Kerr.
 *
 * The contents of this file are subject to the Mozilla Public License Version 1.1
 * (the "License"); you may not use this file except in compliance with the License.
 * You may obtain a copy of the License at http://www.mozilla.org/MPL/
 *
 * Software distributed under the License is distributed on an "AS IS" basis,
 * WITHOUT WARRANTY OF ANY KIND, either express or implied. See the License
 * for the specific language governing rights and limitations under the License.
 *
 * The Original Code is 'iText, a free JAVA-PDF library'.
 *
 * The Initial Developer of the Original Code is Bruno Lowagie. Portions created by
 * the Initial Developer are Copyright (C) 1999-2009 by Bruno Lowagie.
 * All Rights Reserved.
 *
 * Co-Developer of the code is Paulo Soares. Portions created by the Co-Developer
 * are Copyright (C) 2000-2009 by Paulo Soares. All Rights Reserved.
 *
 * Contributor(s): all the names of the contributors are added in the source code
 * where applicable.
 *
 * Alternatively, the contents of this file may be used under the terms of the
 * LGPL license (the "GNU LIBRARY GENERAL PUBLIC LICENSE"), in which case the
 * provisions of LGPL are applicable instead of those above. If you wish to
 * allow use of your version of this file only under the terms of the LGPL
 * License and not to allow others to use your version of this file under
 * the MPL, indicate your decision by deleting the provisions above and
 * replace them with the notice and other provisions required by the LGPL.
 * If you do not delete the provisions above, a recipient may use your version
 * of this file under either the MPL or the GNU LIBRARY GENERAL PUBLIC LICENSE.
 *
 * This library is free software; you can redistribute it and/or modify it
 * under the terms of the MPL as stated above or under the terms of the GNU
 * Library General Public License as published by the Free Software Foundation;
 * either version 2 of the License, or any later version.
 *
 * This library is distributed in the hope that it will be useful, but WITHOUT
 * ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS
 * FOR A PARTICULAR PURPOSE. See the GNU Library general Public License for more
 * details.
 *
 * If you didn't download this code from the following link, you should
 * you aren't using an obsolete version:
 * http://www.lowagie.com/iText/
 */
```

Apply decision to all future occurrences of this file

Clearing decision type

- No license known
- To be discussed
- Irrelevant
- Identified

Action	License	Source	License Text	Acknowledgement	Comment
<input type="checkbox"/>	LGPL-2.0+	nomos: #1	Click to add	Click to add	Click to add
<input type="checkbox"/>	Dual-license	nomos: #1	Click to add	Click to add	Click to add
<input type="checkbox"/>	MPL-1.1	nomos: #1	Click to add	Click to add	Click to add

Showing 1 to 3 of 3 entries

User Decision Bulk Recognition Clearing History

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.
Hint: New license candidates can be added via menu Organize>Licenses

Dual-license Show license

Action	License	License Text	Acknowledgement	Comment	
Add	MPL-1.1	Click to add	Click to add	Click to add	-
Add	LGPL-2.0+	Click to add	Click to add	Click to add	-
Remove	Dual-license	Click to add	Click to add	Click to add	-

Reference text:

Legend:
license relevant text

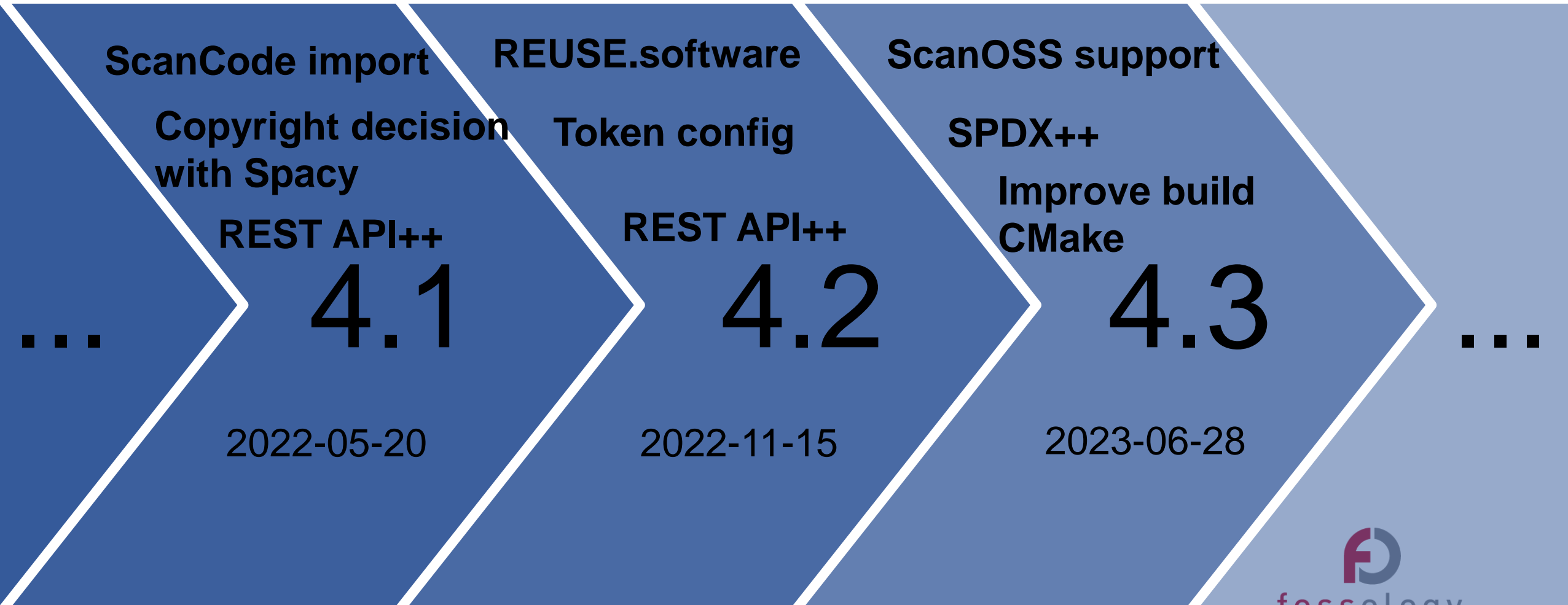
FOSSology – It is about Conclusions

Licensing Challenges

- Licensing can be simple ...
- ... or challenging:
 - Unknown Licenses
 - Written statements
 - Unclear statements
 - Ambiguous statements
 - Incomplete statements
- Depends on domain
- Can be 30% hard to decide

```
SPDXVersion: SPDX-2.0
DataLicense: CC0-1.0
##-----
## Document Information
##-----
DocumentNamespace: http://debian/repo/SPDX2TV_fossology-master-
3.zip_1490661487.spdx
...
##File
FileName: fossology-master/utils/fo-installdeps
SPDXID: SPDXRef-item361
FileChecksum: SHA1: 3fc0aa4a4face8a0d317e0272c5e28e43f44c45a
FileChecksum: MD5: 1576b827a8b28ce1513a490fe2fecdc
LicenseConcluded: GPL-2.0
LicenseInfoInFile: GPL-2.0
FileCopyrightText: <text> Copyright (C) 2008-2014 Hewlett-Packard
Development Company, L.P. </text>
...
```

New Versions mean new Features



FOSSology Scanning In CI

Power of Open Source, benefits of automation



WHY?

Is the current way good enough?



Preparation for release

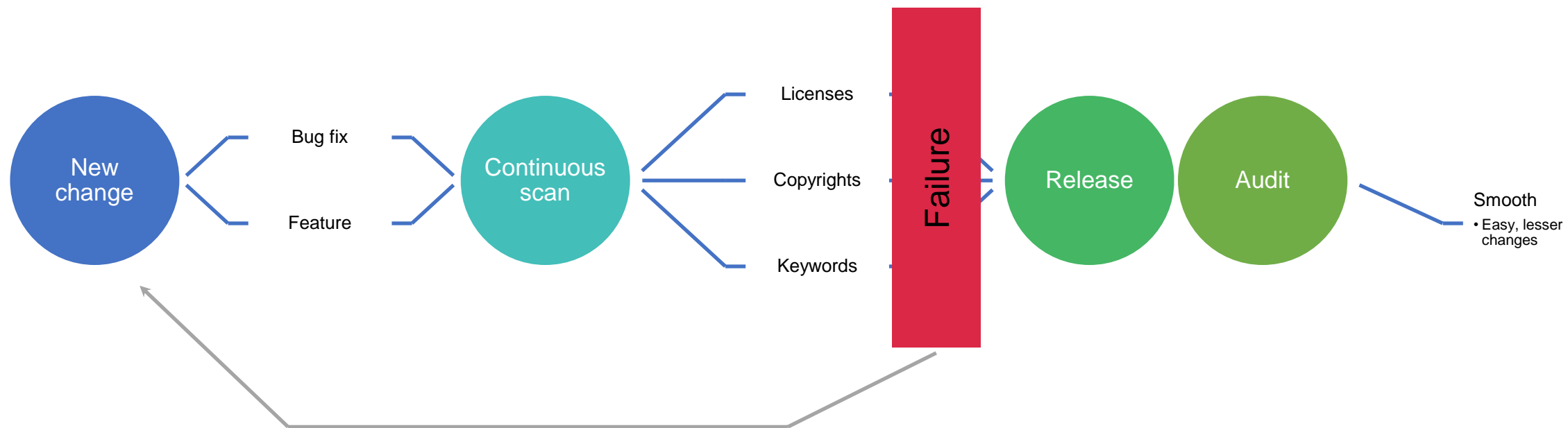
Perform license and copyright scanning

Go or no-go decision



New way

Ease the load with automation



Changes required

```
stages:  
  - license  
  
.gitlab-ci.yml
```

```
license_check:  
  stage: license  
  image: fossology/fossology:scanner  
  script:  
    - /bin/fossologyscanner nomos ojo  
  only: [merge_requests]  
  artifacts:  
    paths:  
    - results  
    expire_in: 1 week  
    when: on_failure
```

```
copyright_check:  
  stage: license  
  image: fossology/fossology:scanner  
  script:  
    - /bin/fossologyscanner copyright keyword  
  only: [merge_requests]  
  artifacts:  
    paths:  
    - results  
    expire_in: 1 week  
    when: on_failure  
  
{  
  "licenses": [  
    "GPL-2.0+",  
    "GPL-2.0",  
    "LGPL-2.1+"  
  ],  
  "exclude": [  
    "*/agent_tests/*",  
    "src/vendor/*"  
  ]  
}
```

whitelist.json

```
.travis.yml
```

```
- stage: Compliance  
  name: License  
  addons: {}  
  services: docker  
  script:  
    - >-  
      if [ "$TRAVIS_PULL_REQUEST" != "false" ]; then  
        docker pull fossology/fossology:scanner  
        && docker run --name "fossologyscanner" -w "/opt/repo" -v ${PWD}:/opt/repo  
        -e TRAVIS=${TRAVIS} -e TRAVIS_REPO_SLUG=${TRAVIS_REPO_SLUG}  
        -e TRAVIS_PULL_REQUEST=${TRAVIS_PULL_REQUEST}  
        fossology/fossology:scanner "/bin/fossologyscanner" nomos ojo ;  
      fi
```

Checkout the documentation:
<https://github.com/fossology/fossology/wiki/FOSSology-as-CI-scanner>



Pipeline status

GitLab

License check failure

Request to merge `test/wrong-license` into `master`

Open in Web IDE | Check out branch | Download

Detached merge request pipeline #161155718 failed for `d4a5c510`

No approval required

View eligible approvers

Merge Delete source branch

1 commit and 1 merge commit will be added to master. [Modify merge commit](#)

You can merge this merge request manually using the [command line](#)

Open | Opened 1 day ago by | Edit | Close merge request

feat(rest): Get file info from hash

Overview 0 | Commits 1 | Pipelines 2 | Changes 14

Status	Pipeline	Triggerer	Commit	Stages	Run Pipeline
	#161155718 detached		<code>d4a5c510</code> feat(rest): Get file in...	00:00:57 21 hours ago	Download
	#161154915 latest		<code>d4a5c510</code> feat(rest): Get file in...	00:05:32 1 day ago	

Oll Korrekt

Request to merge `test/correct-license` into `master`

Open in Web IDE | Check out branch | Download

Detached merge request pipeline #161156143 passed for `6749f42a`

No approval required

View eligible approvers

Merge Delete source branch

1 commit and 1 merge commit will be added to master. [Modify merge commit](#)

You can merge this merge request manually using the [command line](#)

Open | Opened 1 day ago by | Edit | Close merge request

fix(delagent): Remove clearing_decision and lrb

Overview 0 | Commits 1 | Pipelines 2 | Changes 2

Status	Pipeline	Triggerer	Commit	Stages	Run Pipeline
	#161156143 detached		<code>6749f42a</code> fix(delagent): Remo...	00:00:55 21 hours ago	
	#161156051 latest		<code>6749f42a</code> fix(delagent): Remo...	00:05:26 1 day ago	

Pipeline status

Travis

License check failure

Oll Korrekt

Build jobs		View config	
✔ Build			🕒 5 min 9 sec
✔ # 358.1	AMD64 Bionic Build		🕒 5 min 9 sec
✘ Compliance			🕒 6 min 59 sec
✘ # 358.2	AMD64 Bionic Copyright		🕒 6 min 4 sec
✘ # 358.3	AMD64 Bionic License		🕒 6 min 58 sec

Build jobs		View config	
✔ Build			🕒 4 min 57 sec
✔ # 360.1	AMD64 Bionic Build		🕒 4 min 57 sec
✔ Compliance			🕒 6 min 46 sec
✔ # 360.2	AMD64 Bionic Copyright		🕒 1 min 11 sec
✔ # 360.3	AMD64 Bionic License		🕒 6 min 46 sec



Output

License failure

```
3 Preparing the "docker+machine" executor 00:18
4 Using Docker executor with image fossology/fossology:scanner ...
5 Pulling docker image fossology/fossology:scanner ...
6 Using docker image sha256:fafc4a6ebd4a42bcac13d4ff386a0b2c9a56876d8f1d377657f5f85cc2b72f6a for fossology/fossology:scanner ...
7
8 Preparing environment 00:03
9
10
11 Getting source from Git repository 00:06
12
13
14
15
16
17
18
19 Executing "step_script" stage of the job script 00:04
20 $ /bin/fossologyscanner nomos ojo
21 x Following licenses found which are not whitelisted:
22 File: src/www/ui/api/Controllers/FileSearchController.php
23 License:
24     OSL-3.0
25 File: src/lib/php/Dao/PfileDao.php
26 License:
27     Apache-2.0
28 File: src/www/ui/api/Models/Findings.php
29 License:
30     OSL-3.0
31 File: src/lib/php/Dao/test/PfileDaoTest.php
32 License:
33     Apache-2.0
34
35 Uploading artifacts for failed job 00:02
36 Uploading artifacts...
37 results: found 2 matching files and directories
38 Uploading artifacts as "archive" to coordinator... ok id=615646538 responseStatus=201 Created token=soLUiAye
39
40 ERROR: Job failed: exit code 1
```

Output

Copyright failure

```
23 Running before_script and script
24 Authenticating with credentials from job payload (GitLab Registry)
25 $ /bin/fossologscanner copyright keyword
26 ✘ Following copyrights found:
27 File: tags
28 Copyrights:
29 copyright__ my_module/__init__.py /^__copyright__ = metadata.copyright$/" v email__ my_module/licenseDownloader.py /^__email__ = "aman
jain5221@gmail.com"$/" v enter__ pavement.py /^ def __enter__(self):$/" m class:cwd file: exit__ pavement.py /^ def __exit__(sel
f, type_, value, traceb
30 copyright docs/source/conf.py /^copyright = metadata.copyright$/" v
31 copyright my_module/metadata.py /^copyright = '2016 ' + authors_string$/" v coverage pavement.py /^def coverage():$/" f csvColumn
s my_module/licenseDownloader.py /^csvColumns = ["shortname", "fullname", "text", "license_header", "url", "deprecated", "osi_approved", "isException"]$/"
v cwd pavemen
32 File: my_module/licenseDownloader.py
33 Copyright:
34 Copyright 2018 Aman Jain (amanjain5221@gmail.com)
35 ✘ Following keywords found:
36 File: my_module/licenseDownloader.py
37 Keyword:
38 modify it under
39 Running after_script
40 Uploading artifacts for failed job
41 ERROR: Job failed: exit code 1
```

00:11

Potential whitelist file

00:02

Scanners availability

Following scanners are shipped with the runner

nomos

- Most trusted scanner in FOSSology
- Uses regular expression and heuristics

copyright

- Very low false negative findings
- Can find email and URLs too
- Uses regular expressions

ojo

- SPDX License Identifier scanner
- Can find licenses attached using **WITH, AND, OR**
- Uses regular expressions
- Lightning fast

keyword

- Helps in finding potential harmful keywords like:
- licensed, modify it under, etc.



Scanning modes

Diff scanning

- Default scanning mode
- Scan only the diff created by the merge request
- Reduced set of data to scan
- Faster feedback at commit level for developers creating the changes
- Good for build CI pipeline

Repo scan

- Can be used using repo flag
- Scan the complete repo at that commit
- Provides a good overview of the repo for audit works
- Can be scheduled to run at set interval cron jobs
- Good for release/tag pipeline



Allow listing

```
{
  "licenses": [
    "GPL-2.0+",
    "GPL-2.0",
    "LGPL-2.1+"
  ],
  "exclude": [
    "**/agent_tests/**",
    "src/vendor/**"
  ]
}
```

licenses

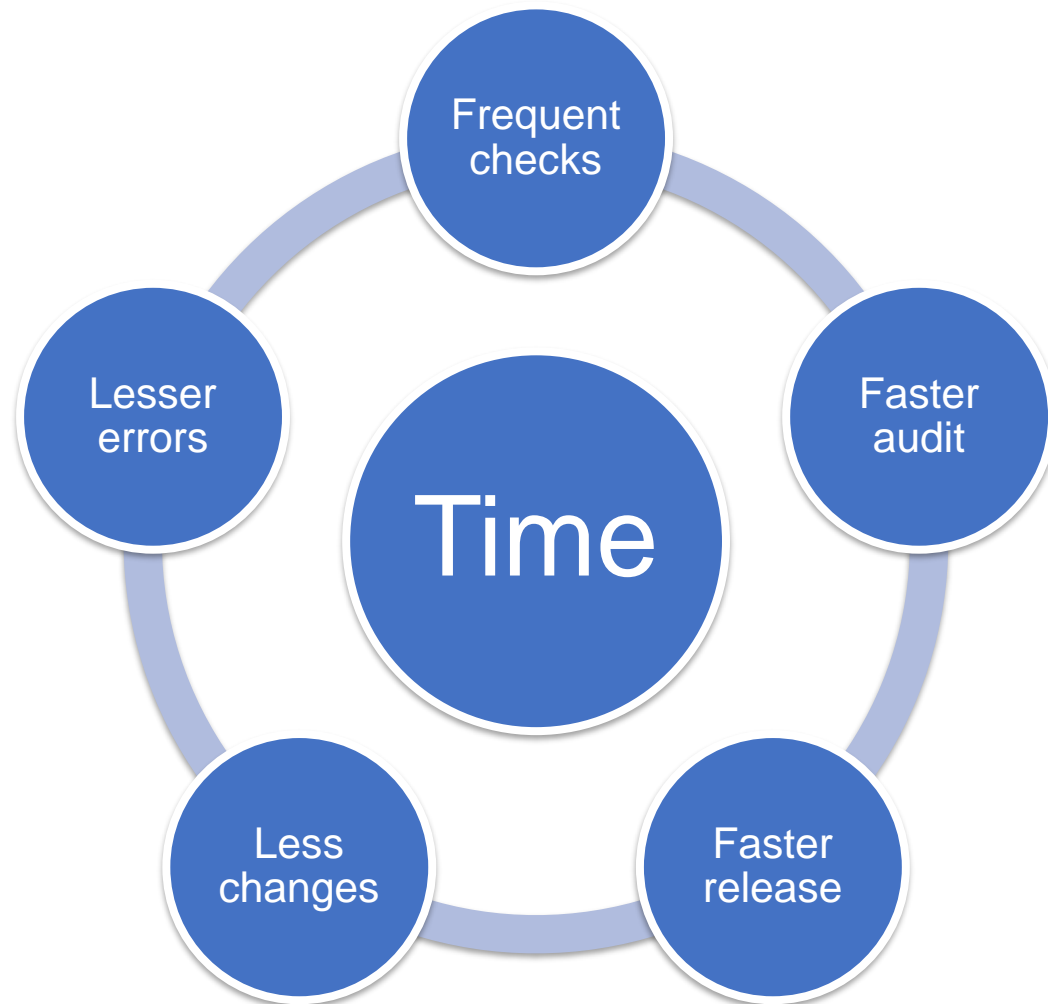
- List of licenses which are allowed
- Each licenses needs to be explicitly mentioned to avoid false negative

exclude

- Files to exclude from scan
- Configuration or test folders
- Understands file glob wild characters



Benefits

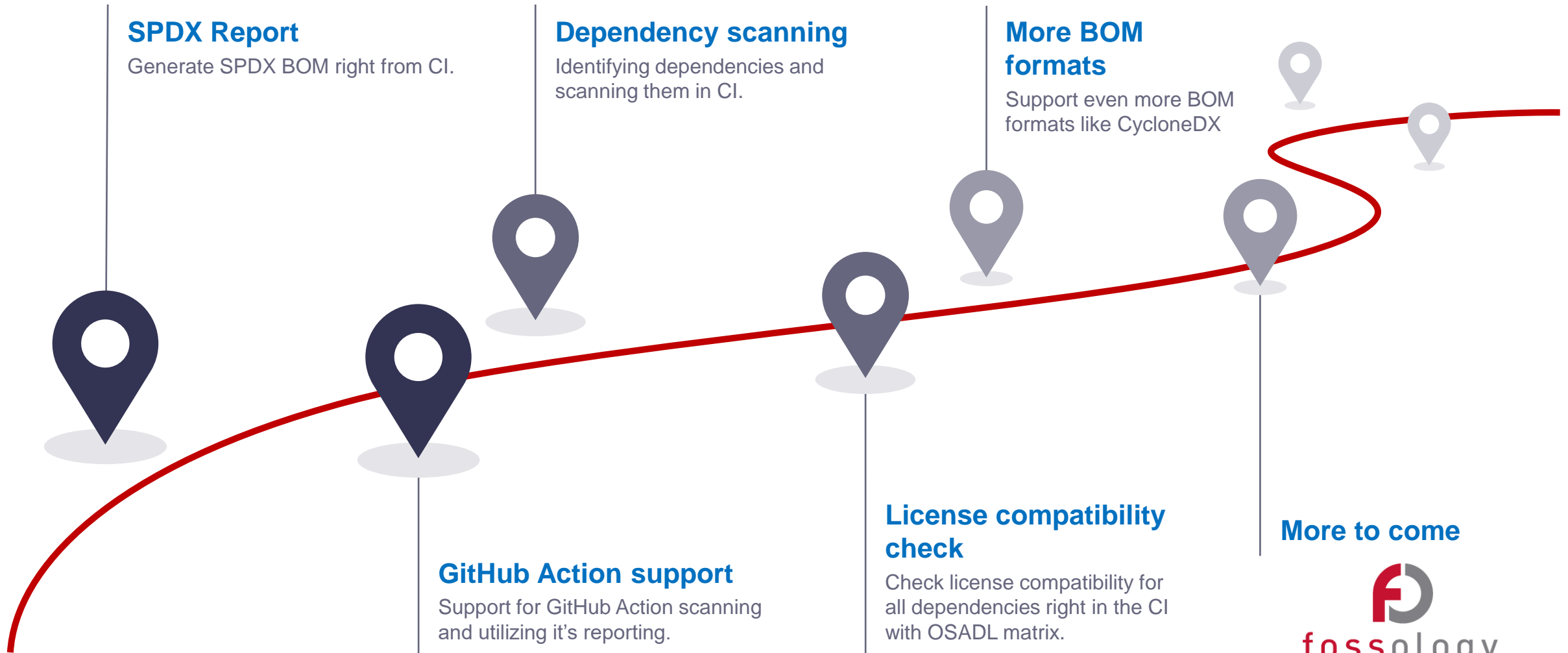


Easier management and automation

- If you manage packages with SW360 (software component catalogue app), FOSSology can receive packages from it and allow product clearing.
- For more info on SW360 : <https://www.eclipse.org/sw360/>



Planned features



There is more at github.com/fossology

Atarashi

- Standalone license scanner
- Written in Python
- Implement multiple text statistics and information retrieval algorithms
- More info at: <https://github.com/fossology/atarashi>

FOSSologySlides

- Slides for Presenting FOSSology
- Make your presentation with FOSSology
- Material for a 1-day training
- More info at: <https://github.com/fossology/FOSSologySlides>

Nirjas

- Python library and tool
- Extract source code and comments
- Supports 25 languages
- Differentiate single line, multiline and continued comments
- More info at: <https://github.com/fossology/Nirjas>



Thank you for your attention!

© 2016-2023 Siemens AG, The Linux Foundation

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://www.spdx.org>

<https://www.openchainproject.org>

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

Contact :

FOSSology Mailing list

- fossology@fossology.org

Email us

- shaheem.azmal@siemens.com
- mishra.gaurav@siemens.com

