



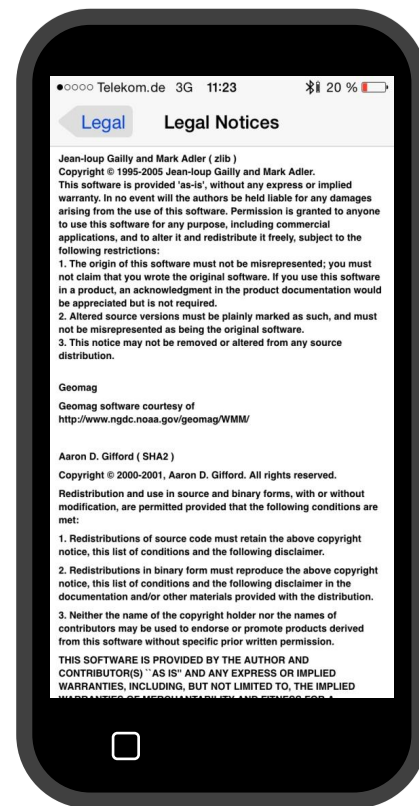
Making Compliance Easy: Filling in the Missing Pieces

Kate Stewart, Sr. Director of Strategic Projects
Feb 15, 2017

Product Distribution

Requires:

- Provide licenses of involved open source software
- Provide copyright statements of involved authors
- Provide disclaimers, etc.



Why is License Compliance **still** a problem?

- Sharing source code between projects is needed for rapid development of new features.
- Scale of open source software available!
- Product companies may have different focus than open source code developers.
- Focus on licensing **after** development done.

Identifying Licenses: Software Archaeology!

- License text at project level may not apply to all files in project.
- Written text found “explaining” licensing
- License relevant statements unclear



```
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

/*
 * Copyright Siemens AG, 2013-2015. Part of the SW360 Portal Project.
 * This program is free software; you can redistribute it and/or modify it under
 * the terms of the GNU General Public License Version 2.0 as published by the
 * Free Software Foundation with classpath exception.
 *
 * This program 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 General Public License version 2.0 for
 * more details.
 *
 * You should have received a copy of the GNU General Public License along with
 * this program (please see the COPYING file); if not, write to the Free
 * Software Foundation, Inc., 51 Franklin Street, Fifth Floor, Boston, MA
 */
package com.siemens.sw360.datahandler.db;

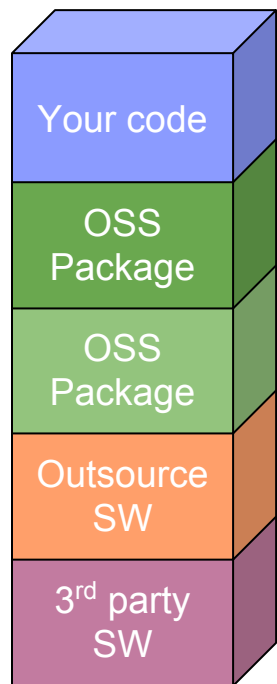
import com.google.common.collect.Sets;
import com.siemens.sw360.components.summary.ProjectSummary;
import com.siemens.sw360.components.summary.SummaryType;
import com.siemens.sw360.datahandler.couchdb.DatabaseConnector;
import com.siemens.sw360.datahandler.couchdb.SummaryAwareRepository;
import com.siemens.sw360.datahandler.thrift.projects.Project;
import com.siemens.sw360.datahandler.thrift.users.User;
import org.ektorp.support.View;
import org.jetbrains.annotations.NotNull;

import java.util.HashSet;
import java.util.List;
import java.util.Set;

import static com.siemens.sw360.datahandler.common.SW360Utils.getBUFFromOrganisation;

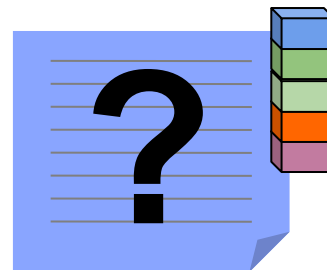
/**
 * CRUD access for the Project class
 *
 * @author cedric.bodet@ngtech.com
 * @author Johannes.Najjar@ngtech.com
 */
@View(name = "all", map = "function(doc) { if (doc.type == 'project') emit(null, doc._id) }")
public class ProjectRepository extends SummaryAwareRepository<Project> {
```

Open Source Compliance: The Challenge



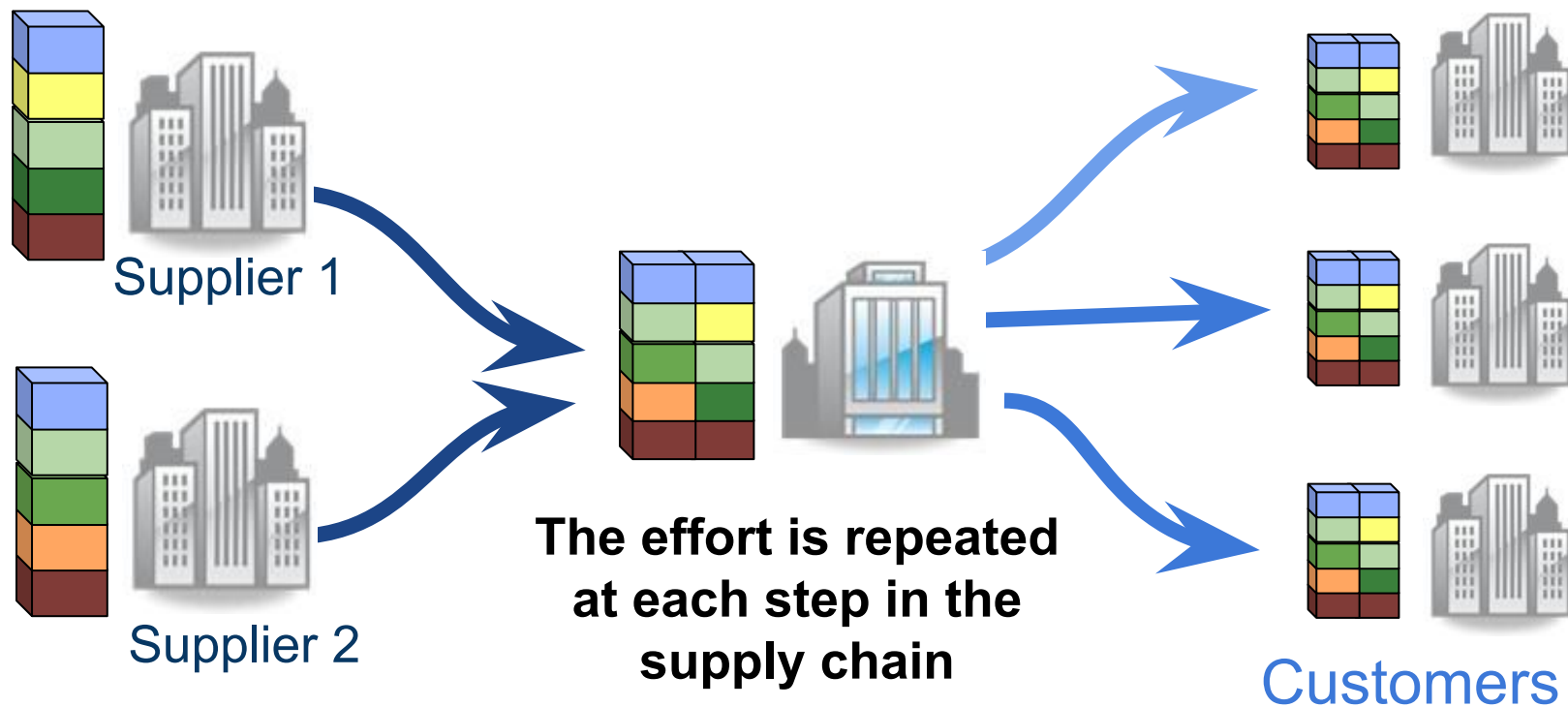
Companies combine
Open Source Software
with other software

**Software Bill of
Materials (BOM)**

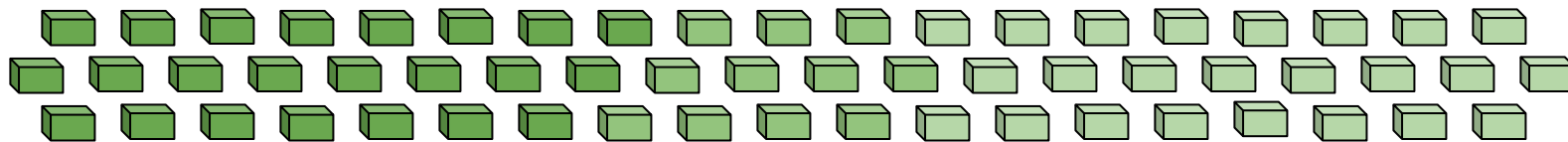


Creating an accurate bill
of materials and notices
requires effort & research

Open Source Compliance: The Challenge

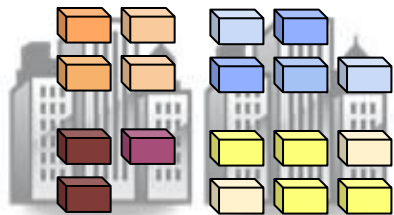


“Open Source”-scape

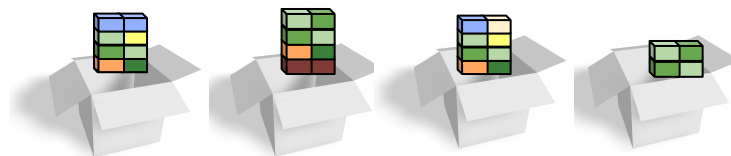


Upstream Projects

Useful “Collections” of Open Source

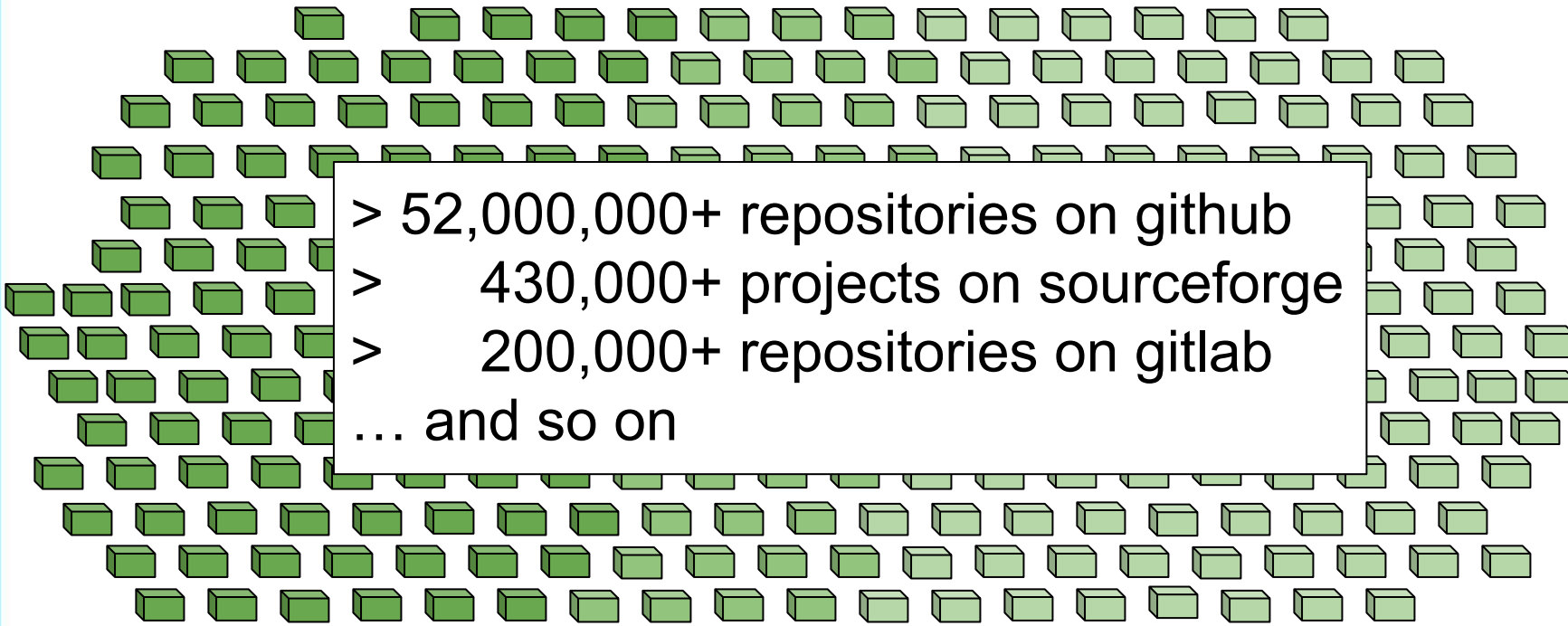


Added-value Software

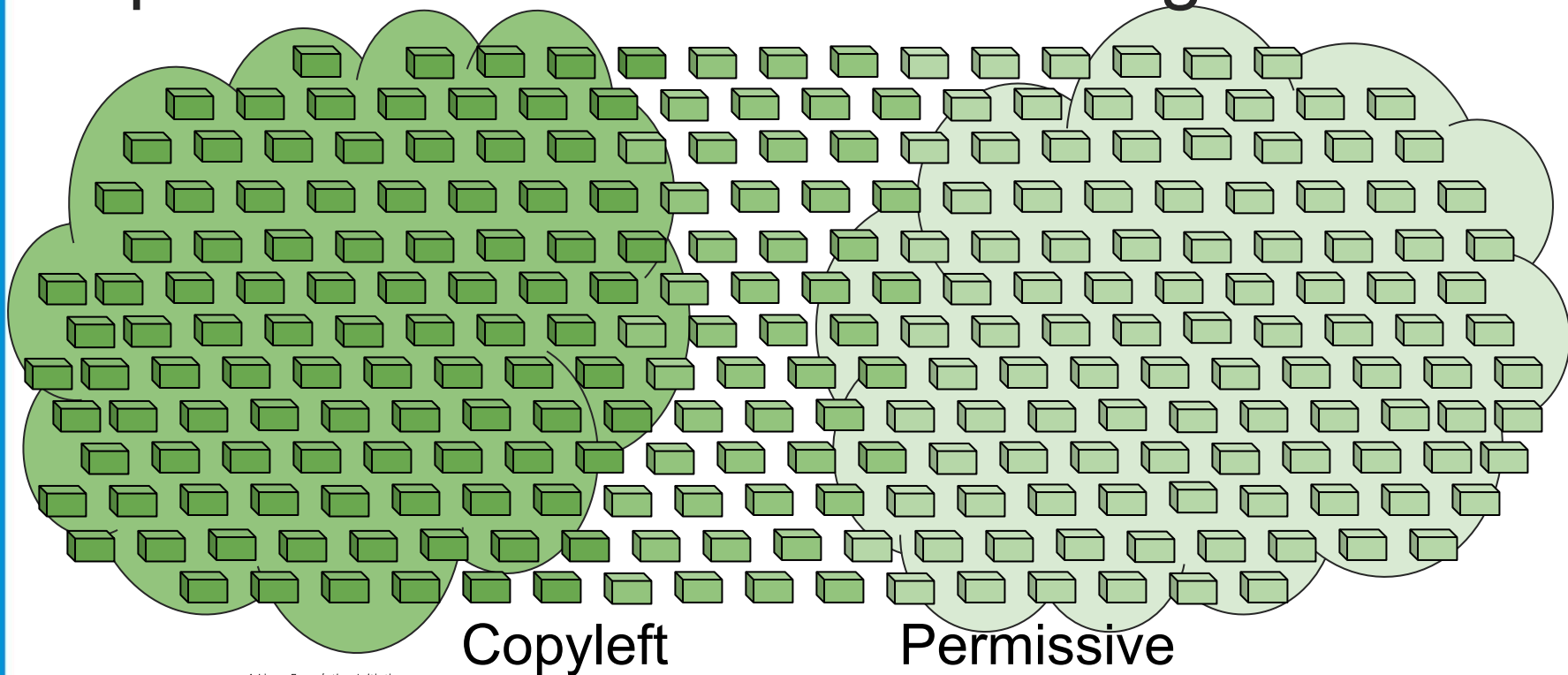


Products

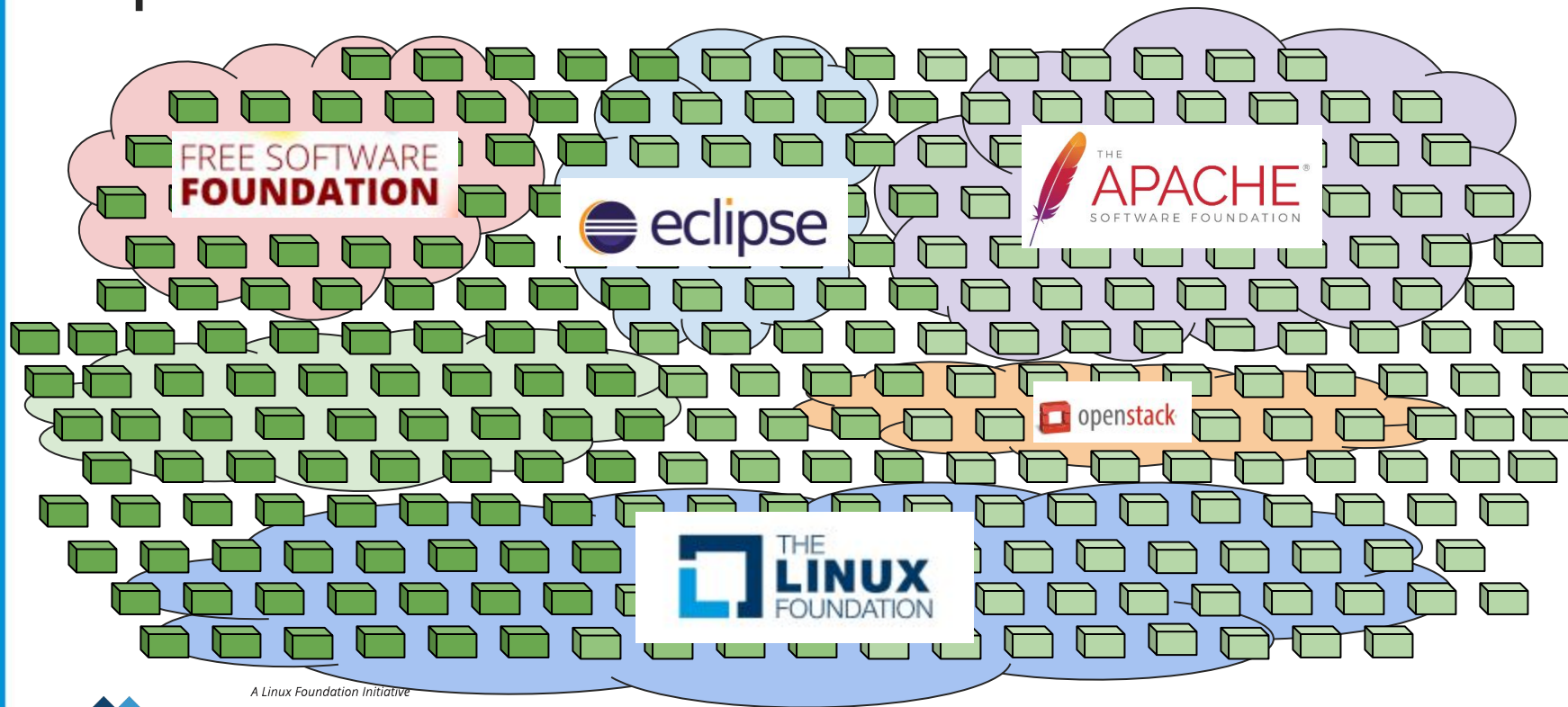
Open Source Upstream Projects

- 
- > 52,000,000+ repositories on github
 - > 430,000+ projects on sourceforge
 - > 200,000+ repositories on gitlab
 - ... and so on

Open Source Software Licensing



Open Source Communities Governance

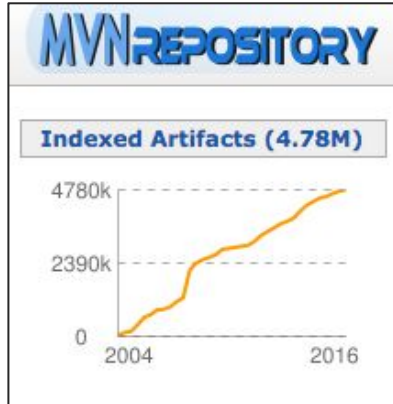


Open Source Distributions & Packaging

- Debian / Ubuntu
- Fedora / Red Hat / CentOS
- Android / Chrome
- Open SUSE / SLES
- FreeBSD / NetBSD
- Yocto / Open Embedded
- ...

Many different policies and practices on packaging open source projects and how licensing information is expressed.

Code Repositories and Package Managers



What can you make with 350,000 building blocks?

The npm registry hosts over a quarter million packages of reusable code — the largest code registry in the world.



Find

Popular libraries like jQuery, Bootstrap, React, and Angular, and components from frameworks including Polymer.



Discover

Packages for mobile, IoT, front end, back end, robotics... everything you need to start building amazing things.



Build

Assemble packages like building blocks to quickly develop awesome new projects.

PyPI - the Python Package Index

The Python Package Index is a repository of software for the Python programming language. There are currently **92741** packages here.

... and others
per language

Everyone does things differently....



Source: <http://clipartix.com/questions-clipart-image-29259/>

.... how can we automate?

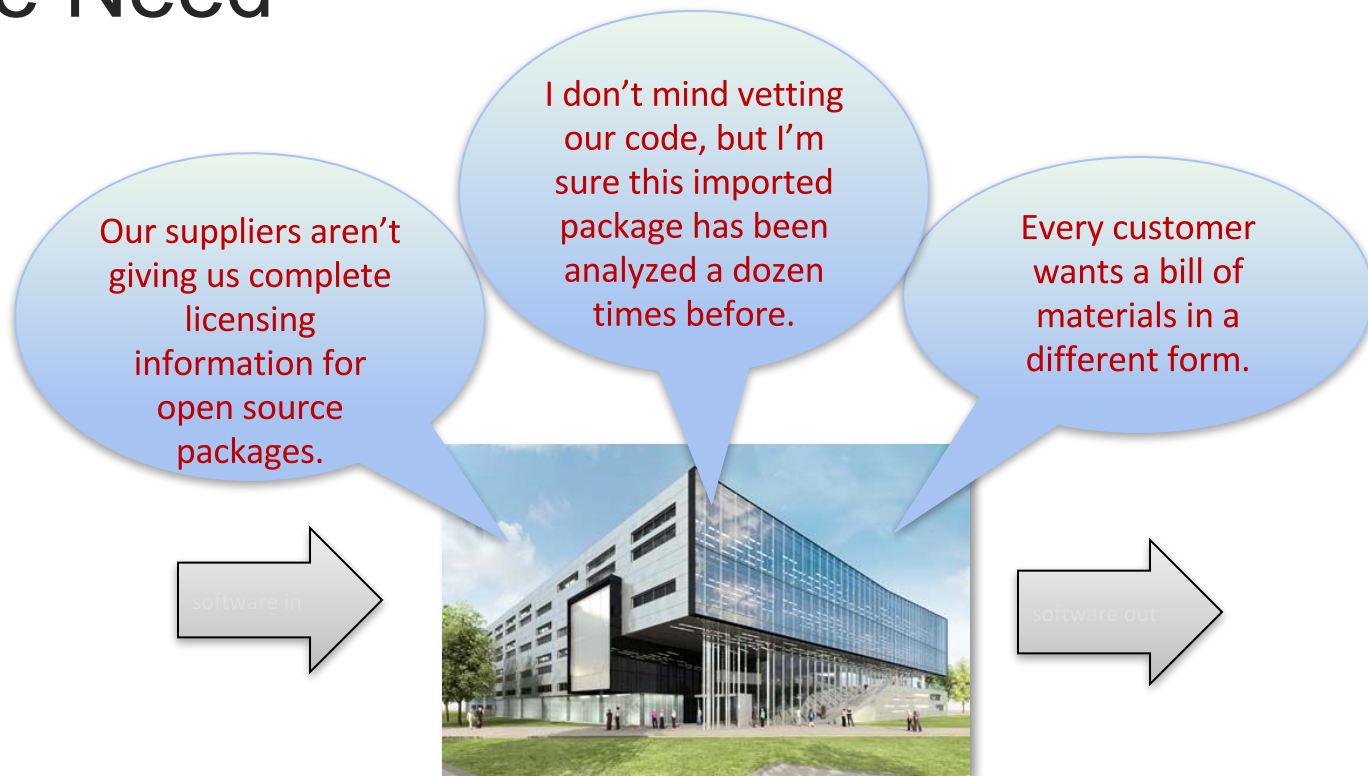
One Step at a Time...

- 1) Common language to communicate licensing data
- 2) Open Source tools to generate licensing data summaries
- 3) Keep licensing data current with every source change
- 4) Transparency on software's licensing data
- 5) Common processes to pass licensing data with software
- 6) Adoption in key projects, distributions, repositories, ...

Ecosystem Automation Scorecard

Status	Goal	Notes
	Common language to communicate licensing data	
	Open Source tools to generate licensing data summaries	
	Keep licensing data current with every source change	
	Transparency of software's licensing data	
	Common processes to pass licensing data with software	
	Adoption by key projects, distributions, repositories	

The Need



Software Package Data eXchange

Open Standard:

- A standard format for communicating the licenses and copyrights associated with software packages

Vision:

- To help reduce redundant work in determining software license information and facilitate compliance

Guiding principles:

- Human and machine readable
- Focus on capturing facts; avoid interpretations

SPDX 2.1

Latest version published 10/2016, addresses all original use-cases.

- Use SPDX License List short identifiers to refer to common licenses found in Open Source efficiently
- Tag source files with SPDX license list short identifiers
- Provide an SPDX document to summarize the licenses in any software you distribute

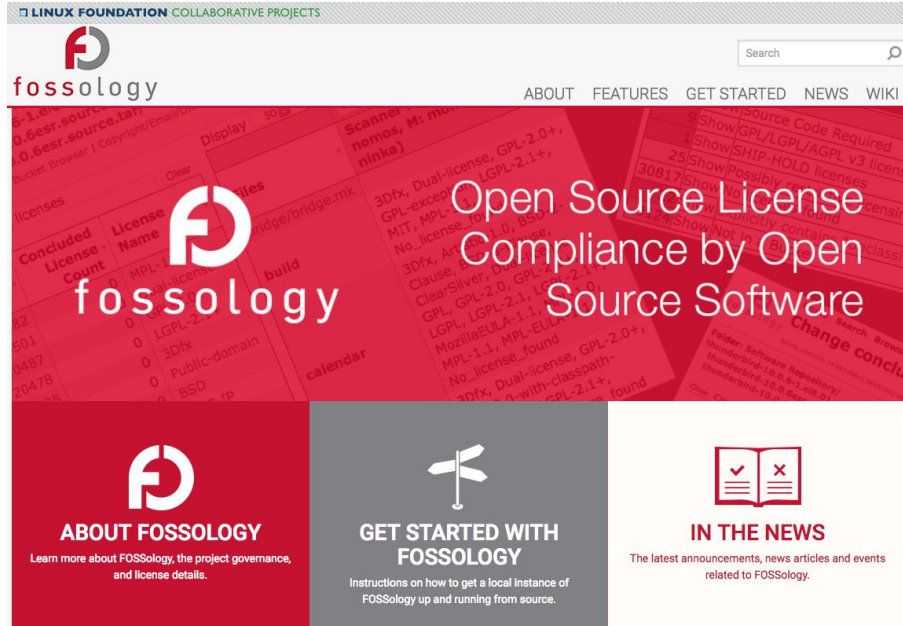
2016 Ecosystem Automation

Status	Goal	Notes
	Common language to communicate licensing data	SPDX
	Open Source tools to generate licensing data summaries	
	Keep licensing data current with every source change	
	Transparency of software's licensing data	
	Common processes to pass licensing data with software	
	Adoption by key projects, distributions, repositories	

Open Source SPDX Document Creation

- SPDX-Tools:
 - <https://github.com/spdx/tools>
- FOSSology
 - <https://github.com/fossology/fossology>
- DoSOCSv2
 - <https://github.com/DoSOCSv2/DoSOCSv2>

www.fossology.org



3.1 release
generates: both
SPDX tag:value
& SPDX RDF
documents.

Open Source tools for Summarizing Licensing

- Auditing existing code and generating SPDX document
 - FOSSology 3.1
- Command line generate SPDX with build scripts
 - DoSOCSv2 project used with Yocto
 - Prototype FOSSology with ELBE with Debian
 - LiD (announced this week).
- Dependency tracking in repositories
 - Maven POM, Eclipse Plugin prototypes

2016 Ecosystem Automation

Status	Goal	Notes
	Common language to communicate licensing data	SPDX
	Open Source tools to generate licensing data summaries	FOSSology, SPDX-tools...
	Keep licensing data current with every source change	
	Transparency of software's licensing data	
	Common processes to pass licensing data with software	
	Adoption by key projects, distributions, repositories	

Keep licensing data current with every change

- **Command line tools able to generate SPDX documents**
 - For upstream project to use for releases
 - For inclusion in check-patch utilities (stop garbage in)
 - For packaging & build scripts to run
 - For code composers from building blocks and libraries.

Some starting points exist but need to make robust.

Command Line SPDX Tools...

- DoSOCSv2
 - <https://github.com/DoSOCSv2/DoSOCSv2>
- **Coming Soon:** FOSSology command line improvements & wrapper scripts with ELBE
- **Coming Soon:** LiD code License Scanner
 - <https://www.codeaurora.org/qualcomm-ostg-lid>
- **Coming Soon:** ScanCode is having SPDX added
 - <https://github.com/nexB/scancode-toolkit>

2016 Ecosystem Automation

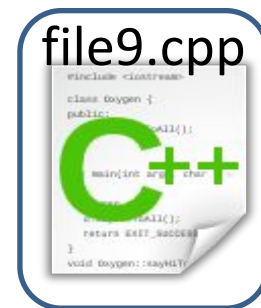
Status	Goal	Notes
	Common language to communicate licensing data	SPDX
	Open Source tools to generate licensing data summaries	FOSSology, SPDX-tools
	Keep licensing data current with every source change	DoSOCSv2, LiD, ...
	Transparency of software's licensing data	
	Common processes to pass licensing data with software	
	Adoption by key projects, distributions, repositories...	

Evolving From Package to Source File Licensing

- With widespread sharing of source files (composable repositories, etc.), the package level license may not be complete
- The licenses of source files need to be reviewed for distribution obligations



vs.



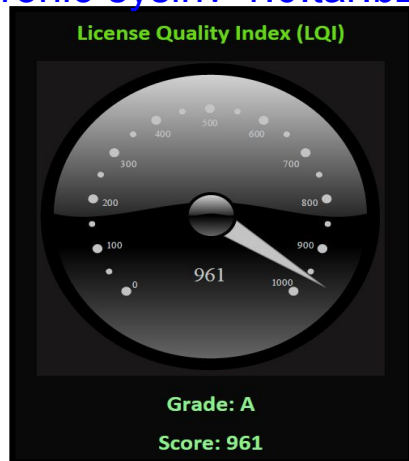
Establishing a License Coverage “Grade”

- Grade = % of copyrightable source files with clear licensing terms *contained within the file*.
- A license notice per file should be standard header (if it exists) but can be as simple as:
“**SPDX-License-Identifier: GPL-2.0**”
- Grade bump from F to D if LICENSE.txt exists
- Although a top level license often exists, the emphasis is on individual source file licenses

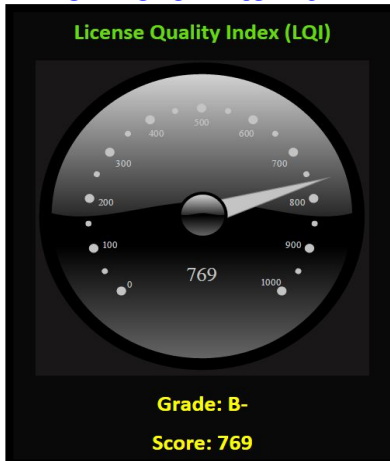
Grade	% Licenses Detected
A+	97 - 100
A	93 - 96
A-	90 - 92
B+	85 - 89
B	80 - 84
B-	75 - 79
C+	70 - 74
C	65 - 69
C-	55 - 64
D+	50 - 54
D	40 - 49
D-	30 - 39
F	<= 29

Example: OpenStack Packages*

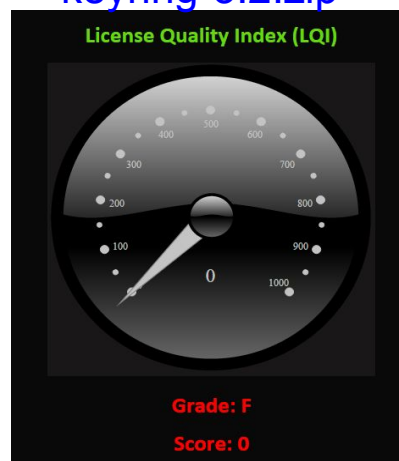
ironic-sysinv-1.0.tar.bz2



novnc-0.4.tar.bz2



keyring-3.2.zip



*source: Mark Gisi, Windriver
http://spdx.windriver.com/pkg_upload.aspx

Building on Best Practices in Communities

Package	Foundation	Grade	% Detected
apache-tomcat-7.0.47-src.tar.gz	Apache	A+	99.5
httpd-2.4.6.tar.bz2	Apache	A+	98.4
autoconf-2.69.tar.gz	FSF	A+	100
grep-2.20.tar.xz	FSF	A+	94
org.eclipse.datatools.sqltools-master.tar.gz	Eclipse	A-	92.3
org.eclipse.dltk.core-master.tar.gz	Eclipse	A-	90.4

- It is a best practice to include a license notice in every file.
- Apache & FSF packages are generally getting it right.
 - Key is stopping problems at the source! :-)

Projects tracking license at file level?

Dependent on Community Governance

- Apache Software Foundation
- Free Software Foundation
- The Linux Foundation
- Eclipse Foundation
- OpenStack Foundation
- ...

Started - initial focus area

Also depends on Distribution's project packaging.

Goal:

Get same level of **automatically** detectable information through entire ecosystem by increasing **transparency** on licensing.

Transparency of Software's Licensing Data

Standard method for summarizing licensing at file level

- Need simple “License Coverage Grade” per project
 - provides grading A+-F based on transparent heuristics.
 - Simple to generate from SPDX document & sources.
- Need to develop open source command line tool to implement.
- Work with projects to “self score” (code authors).
- Work with foundations and distributions to adopt as part of packaging and distribution.

2016 Ecosystem Automation

Status	Goal	Notes
	Common language to communicate licensing data	SPDX
	Open Source tools to generate licensing data summaries	FOSSology, SPDX-tools
	Keep licensing data current with every source change	DoSOCSv2, LiD, ...
	Transparency of software's licensing data	?
	Common processes to pass licensing data with software	
	Adoption by key projects, distributions, repositories...	

Software Supply Chain Information Needs

Products today are built on many, many layers of software packages interacting together.

Product creators need to:

- understand which security vulnerabilities may be relevant
- understand who may be able to fix them
- understand distribution obligations associate with software's licensing terms

Supporting Supply Chain Requirements

For each package:

- understand which security vulnerabilities may be relevant
⇒ link it to NIST Common Platform Enumeration (**CPE**), which will permit lookups to CVEs & CWEs as they change, via NISTs databases.
- understand who may be able to fix them
⇒ who are the **copyright** holders of all the files?
- understand distribution obligations
⇒ what are all the **licenses** in use for the package?

SPDX 2.1:

- supports licensing & copyright at file level
- support summaries at package level and links to NIST CPE

www.openchainproject.org



OpenChain: Building the Business Processes

- Identification of the origin and license of FOSS software
- Tracking FOSS software within the development process
- Performing FOSS review and identifying license obligations
- Fulfillment of license obligations when product ships
- Oversight for Open Source Compliance Program, creation of policy, and compliance decisions
- Training

Common processes to pass licensing data

Supply chain Processes: OpenChain Project

- Specification (lead: Mark Gisi): 1.0 release in October 2016.
- Curriculum (lead: Shane Coughlan):
 - [175 contributed slides](#) from ARM, Qualcomm, Philips, Samsung
 - Curated down to 75 slides in 7 sections
- Conformance (lead: Miriam Ballhausen): Self-Conformance to online Questionnaire is first phase.

Community Project Processes: varied, based on community

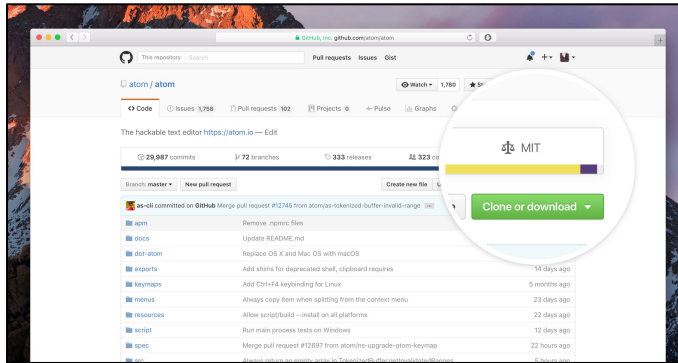
- Need to interface community practices (1 license file per package) better with supply chain needs (licensing information at source file level).

2016 Ecosystem Automation

Status	Goal	Notes
	Common language to communicate licensing data	SPDX
	Open Source tools to generate licensing data summaries	FOSSology, SPDX-tools
	Keep licensing data current with every source change	DoSOCSv2, LiD, ...
	Transparency of software's licensing data	?
	Common processes to pass licensing data with software	OpenChain
	Adoption by key projects, distributions, repositories...	

Adoption in Ecosystem

- **Adoption SPDX License Identifiers:**
 - Debian recognized since DEP5 adopted, Fedora transitioning.
 - Linux Foundation transitioning, Eclipse considering.
 - New project in Package Manager Repositories adopting
 - Github adopted for projects in September 2016 (see Licenses API)!



A Linux Foundation Initiative



SPDX current adoption

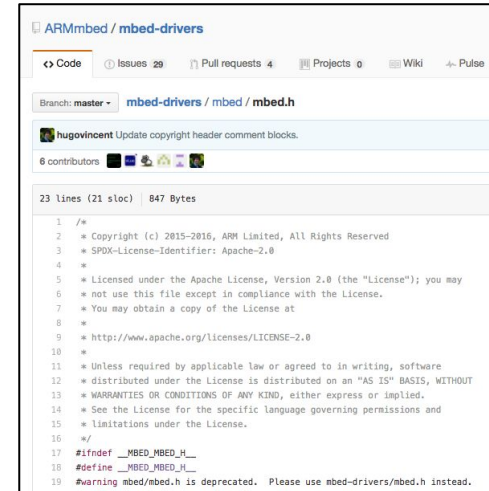
Package Manager	Licence in md	Dep. in md	SPDX IDs	Expressions	Non proliferation	Comment
NPM	✓	✓	✓	✓	✓	
Composer	✓	✓	✓	✓	✓	
Cargo	✓	✓	✓	!	✓	Developers pushing for SPDX 2
RubyGems	✓	✓	✓	✗	✗	
Bower	✓	✓	✓	✗	✗	
Maven	✓	✓	✗	✗	✗	
(Meta)CPAN	✓	✓	✗	✗	✗	
Pip / PyPi	✓	✗	✗	✗	✗	Page / Discussion / warehouse github.com/open-source-project-templates/46 python.org/en/news/2016-09-01
Opam	!	✓	!	✗	✗	Sanitizing effect of SPDX (Thanks to legal team)

SPDX General meeting 20160303

CURE inno

Adoption in Ecosystem

- **Use of SPDX License Tags in Source Files:**
 - Developer initiated in U-Boot in 2013 for efficiency and to help with automatic processing.
 - Selective upstream projects adopt based on developer preferences.
 - Linux Foundation projects adopting: started adding to Linux in November.
 - “[Open Government Partnership](#)” created a [best practices template](#) for Open Source Policy that includes SPDX-License-Identifiers in December, France adopting “as is”.



```
1 /*
2  * Copyright (c) 2015-2016, ARM Limited, All Rights Reserved
3  * SPDX-License-Identifier: Apache-2.0
4  *
5  * Licensed under the Apache License, Version 2.0 (the "License"); you may
6  * not use this file except in compliance with the License.
7  * You may obtain a copy of the License at
8  *
9  * http://www.apache.org/licenses/LICENSE-2.0
10  *
11  * Unless required by applicable law or agreed to in writing, software
12  * distributed under the License is distributed on an "AS IS" BASIS, WITHOUT
13  * WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
14  * See the License for the specific language governing permissions and
15  * limitations under the License.
16  */
17 #ifndef __MBED_MBED_H_
18 #define __MBED_MBED_H_
19 #warning mbed/mbed.h is deprecated. Please use mbed-drivers/mbed.h instead.
```

Examples:

Original software

<AUTHORS>,

Copyright © 2013-2015 ENTITY_NAME

SPDX-License-Identifier: GPL-3.0

v1.0

Adoption in Ecosystem

- **SPDX Specification:**
 - Windriver releases with all their products with SPDX documents. Hosts free service to generate documents.
 - Companies able to use commercial tools able to generate documents (BlackDuck, Palamida, SourceAuditor, etc.) as well as open source tools (FOSSology, homegrown, etc.).
 - Upstream projects need open-source based command line tools to integrate into CI loops (DoSOCSv2, LiD, ...)
 - Used to structure internal databases in large companies (Samsung, TI, ARM, Intel, Siemens ...)

2016 Ecosystem Automation

Status	Goal	Notes
	Common language to communicate licensing data	SPDX
	Open Source tools to generate licensing data summaries	FOSSology, SPDX-tools
	Keep licensing data current with every source change	DoSOCSv2, LiD, ...
	Transparency of software's licensing data	?
	Common processes to pass licensing data with software	OpenChain
	Adoption by key projects, distributions, repositories...	Github, Debian,

How Can You Help?

- Add “SPDX-License-Identifier” tags to open source files where you have commit rights if they do not already have standard licenses.
 - If the license is common, but not on the SPDX license list, ask to be added.
- Participate (develop, test, report bugs, document) FOSSology creating command line interfaces to generate SPDX files and incorporate better agents.
- Generate SPDX documents for the projects you participate in
 - Make sure licenses are consistent ;-)
- Participate in defining policies and open source tools for industry wide standard on a “License Coverage Grade” based on analyzing SPDX documents and source code for projects in 2017.

Closing thoughts...

If everyone does a bit,

- we can make easy to understand which license apply for products,
- we can respect the open source developers intent when they contributed code

Step by step, together we can get this automated!



Source:
<https://catalog.archives.gov/id/535413>



Thank you! Questions?

kstewart@linuxfoundation.org