SPDX: Bridging the Compliance Tooling Gap

Gary O’Neall
Source Auditor Inc. and SPDX Tech Working Group co-lead

Steve Winslow
Director of Strategic Programs, The Linux Foundation
The Journey to Open Source Compliance

• A Trail Map for Compliance
• How tools can fit in
• The tools landscape
• A common language to enable tools integrations
Compliance Trail Map

What is Open Source?
- obligations?
- risks?

Talking about the same license
- Text variations
- License choices

Knowing what software we have
- Source
- Dependencies
- Build artifacts

Knowing attributes about the software
- Licenses
- Copyrights
- Notices
- Vulnerabilities

Communicating what we have
- Attribution
- Required source
- Meet downstream consumer requirements

Understanding interactions
- What is linked
- What is distributed
- What is hosted

Create and refine policies

Source: https://commons.wikimedia.org/wiki/File:NPS_guadalupe-peak-trail-map.jpg CC-BY-SA
# How Tools Fit In

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talking about the same license</td>
<td>License compare tools, license matching tools, scanners</td>
</tr>
<tr>
<td>Knowing what software we have</td>
<td>Dependency tools and plugins, scanners, security analysis tools, source snippet matching tools, binary analysis tools</td>
</tr>
<tr>
<td>Knowing attributes about the software</td>
<td>Scanners, binary analysis tools, vulnerability analysis tools</td>
</tr>
<tr>
<td>Understanding interactions</td>
<td>Build analysis tools, software interaction analysis tools, some scanners</td>
</tr>
<tr>
<td>Communicating what we have</td>
<td>Spreadsheets, PDFs, web apps, plain text</td>
</tr>
</tbody>
</table>
Tooling Landscape (apologies for everything we’re missing!)

License Info
- diff
- SPDX-License-Identifier

Analysis Tools
- binaryanalysis-ng (BANG)
- grep
- licensee
- licensechecker (lc)
- ScanCode
- cregit

Tooling Frameworks
- Automating Compliance Tooling (ACT)
- Eclipse Oscano
- TERN
- OSS Review Toolkit

Proprietary Tools
- FOSSID
- BLACK DIsc
- Sonatype
- WhiteSource
- SYNOPSYS
- SOURCE Auditor
- FOSSA
- THE LINUX FOUNDATION
Tooling Landscape (apologies for everything we’re missing!)

Best Practices and Initiatives

Sharing-creates-value

Information Sources

More Landscapes!

DoubleOpen survey
Why not One Tool to Rule Them All?

Material Design icons by Google; used under Apache 2.0; https://material.io/tools/icons/

“One Ring interpretation” image by Jorge Arimany; used under CC-BY-SA-3.0; https://commons.wikimedia.org/wiki/File:One_ring.png
The Need for an Interchange Standard

- We will have multiple tools
- We will have multiple frameworks for how the tools fit together
- We will have multiple organizational policies, processes and best practices for what “compliance” means
- The information exchanged is very similar
- We need a vocabulary of common terms the tools can use to communicate
SPDX is a **vocabulary** and a **language**

It is not a tool (though the SPDX project provides some tools)

It is not a framework

It is not a policy or a process

It is a human-readable, machine-readable way to get tools, frameworks, policies and processes to fit together
SPDX: Talking About the Same License

What license is this?

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
**SPDX: Talking About the Same License**

**Or this?**

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

* Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
* Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
* Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE LIABILITY OF THE COPYRIGHT HOLDERS AND CONTRIBUTORS EXCEED $250. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
How about this one?

Redistribution and use in source and binary forms, with or without modification, are permitted (subject to the limitations in the disclaimer below) provided that the following conditions are met:

* Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
* Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
* Neither the name of [Owner Organization] nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

NO EXPRESS OR IMPLIED LICENSES TO ANY PARTY'S PATENT RIGHTS ARE GRANTED BY THIS LICENSE. THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- Redistributions in any form must be accompanied by information on how to obtain complete source code for the DB software and any accompanying software that uses the DB software. The source code must either be included in the distribution or be available for no more than the cost of distribution plus a nominal fee, and must be freely redistributable under reasonable conditions. For an executable file, complete source code means the source code for all modules it contains. It does not include source code for modules or files that typically accompany the major components of the operating system on which the executable file runs.

THIS SOFTWARE IS PROVIDED BY SLEEPYCAT SOFTWARE “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT, ARE DISCLAIMED. IN NO EVENT SHALL SLEEPYCAT SOFTWARE BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
Or we could talk about them this way instead:

- BSD-3-Clause
- LicenseRef-BSD-3-Clause-custom-disclaimer
- BSD-3-Clause-Clear
- Sleepycat
SPDX: Talking About the Same License

- How SPDX helps
  - License ID
  - License Expressions
  - Templated text in License List

- Tooling Examples
  - SPDX License Identifiers (spdx.org/ids)
  - spdx-license-diff (courtesy of Alan Tse)
  - REUSE Software from FSFE (reuse.software)
  - Check License (spdxtools.sourceforge.com)
  - licensechecker / lc (https://github.com/boyter/lc)
In source code:

// SPDX-License-Identifier: MIT
In an SPDX document:

For a package’s license:

PackageLicenseConcluded: MIT AND BSD-3-Clause
PackageLicenseDeclared: MIT
In an SPDX document:
For a file’s license:

LicenseConcluded: MIT AND BSD-3-Clause
LicenseDeclared: MIT
In an SPDX document:

For a file’s license, not on the License List:

LicenseConcluded: LicenseRef-Acme-custom
...

LicenseID: LicenseRef-Acme-custom
LicenseName: Acme Source License
ExtractedText: This is the Acme Source License. ...
LicenseCrossReference: https://example.com/acme/license.txt
SPDX: Understanding Your Software

• How SPDX helps
  • Simple and broad concepts. A **package** is a collection of software. Can be a collection of files pulled from a repo; a .tar.gz file; a container image; a dependency without reference to individual files; ...

• Tooling Examples
  • [github.com/spdx/tools](http://github.com/spdx/tools), tools-python, tools-golang

Knowing what software we have

• Source
• Dependencies
• Build artifacts
In an SPDX document:

Referring to a dependency:

PackageName: requests
SPDXID: SPDXRef-requests
PackageDownloadLocation: git+ssh:kennethreitz/requests.git
FilesAnalyzed: false
In an SPDX document:

OIN Linux System Definition packages, github.com/swinslow/spdx-oin:

```
PackageName: at-spi
SPDXID: SPDXRef-31
PackageVersion: 1.6.6
PackageDownloadLocation: http://ftp.gnome.org/pub/GNOME/sources/at-spi/1.6/...
FilesAnalyzed: false
PackageHomePage: http://www.gtk.org/
PackageSourceInfo: Project Download URL was http://ftp.gnome.org/pub/GNOME/...
PackageLicenseConcluded: NOASSERTION
PackageLicenseDeclared: NOASSERTION
PackageCopyrightText: NOASSERTION
PackageDescription: Assistive Technology Service Provider Interface
```
SPDX: Understanding Your Software

In an SPDX document:

Referring to a file:

FileName: /src/main.go
SPDXID: SPDXRef-file1
FileChecksum: SHA1: ...
In an SPDX document:

Describing a file’s attributes:

FileName: /src/main.c  FileName: /bin/mainApp
SPDXID: SPDXRef-main-c  SPDXID: SPDXRef-mainApp
FileType: SOURCE  FileType: BINARY
              FileType: APPLICATION
In an SPDX document:

Describing build artifacts:

FileName: /src/main.c  FileName: /bin/mainApp
SPDXID: SPDXRef-main-c          SPDXID: SPDXRef-mainApp

Relationship: SPDXRef-main-c GENERATES SPDXRef-mainApp
SPDX: Gathering Software Information

• How SPDX helps
  • Provides a structured, **machine-readable** and **human-readable** format for this information

• Tooling Examples
  • FOSsology
  • ScanCode
  • SPDX Maven Plugin
  • Proprietary tools

Knowing attributes about the software

• Licenses
• Copyrights
• Notices
• Vulnerabilities
In an SPDX document:

Licenses and copyright notices in a dependency:

PackageName: requests
SPDXID: SPDXRef-requests
Filesanalyzed: false
PackageLicenseConcluded: Apache-2.0 AND BSD-3-Clause
PackageLicenseDeclared: Apache-2.0
PackageCopyrightText: Copyright 2018 Kenneth Reitz
In an SPDX document:

Referring to a dependency, omitting this info:

- PackageName: requests
- SPDXID: SPDXRef-requests
- FilesAnalyzed: false
- PackageLicenseConcluded: NOASSERTION
- PackageLicenseDeclared: NOASSERTION
- PackageCopyrightText: NOASSERTION
In an SPDX document:

Licenses and copyright notices in a file:

FileName: /src/main.go
SPDXID: SPDXRef-file1
FileChecksum: SHA1: ...
LicenseConcluded: Apache-2.0 AND MIT
LicenseInfoInFile: MIT
FileCopyrightText: Copyright Steve Winslow
In an SPDX document:

Referring to a file, omitting this info:

FileName: /src/main.go
SPDXID: SPDXRef-file1
FileChecksum: SHA1: ...
LicenseConcluded: NOASSERTION
LicenseInfoInFile: NOASSERTION
FileCopyrightText: NOASSERTION
In an SPDX document:

Describing security vulnerabilities:

PackageName: spring_framework
SPDXID: SPDXRef-springFramework
FilesAnalyzed: false
ExternalRef: SECURITY cpe23Type
cpe:2.3:a:pivotal_software:spring_framework:
4.1.0:*:*:*:*:*:*:*:*:*
SPDX: Understanding Software Interactions

• How SPDX helps
  • The Relationship field provides a rich vocabulary for describing ways that files and packages relate and interact with one another

• Tooling Examples
  • SPDX Maven plugin
  • opportunities to improve here...
In an SPDX document:
Describing build process:

FileName: /src/main.c  
SPDXID: SPDXRef-main-c  
Relationship: SPDXRef-main-c GENERATES SPDXRef-main-o

FileName: /bin/main.o  
SPDXID: SPDXRef-main-o
In an SPDX document:

Describing patches:

FileName: /src/patch001
SPDXID: SPDXRef-patch001
Relationship: SPDXRef-patch001 PATCH_FOR SPDXRef-driver.c

FileName: /src/driver.c
SPDXID: SPDXRef-driver.c
In an SPDX document:

Describing linked files:

FileName: /obj/LGPL-lib.o  
SPDXID: SPDXRef-LGPL-lib

Relationship: SPDXRef-LGPL-lib DYNAMIC_LINK SPDXRef-Proprietary
or
Relationship: SPDXRef-LGPL-lib STATIC_LINK SPDXRef-Proprietary
In an SPDX document:

Describing files that might be less relevant for compliance:

<table>
<thead>
<tr>
<th>FileName</th>
<th>SPDXID</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>/src/project.c</td>
<td>SPDXRef-projectCode</td>
<td>TEST_CASE_OF SPDXRef-projectCode</td>
</tr>
<tr>
<td>/src/otherFile</td>
<td>SPDXRef-otherFile</td>
<td>BUILD_TOOL_OF SPDXRef-projectCode</td>
</tr>
</tbody>
</table>
In an SPDX document:

Describing dependency relationships:

FileName: /src/myCode.py
SPDXID: SPDXRef-myCode

PackageName: requests
SPDXID: SPDXRef-requests

Relationship: SPDXRef-myCode HAS_PREREQUISITE SPDXRef-requests

Relationship: SPDXRef-requests HAS_PREREQUISITE SPDXRef-urllib3
SPDX: Communicating Open Source Info

• How SPDX helps
  • An SPDX document lays out the license, attribution and copyright notices you need to provide, licenses requiring source, ...
  • Provide the SPDX document itself, enable downstream users to use for their own compliance processes

• Tooling Examples
  • SPDX to Spreadsheet (spdxtools.sourceauditor.com)
  • License coverage grader (https://github.com/spdx/license-coverage-grader)
In an SPDX document:

Communicating licenses and copyright notices:

PackageLicenseConcluded: Apache-2.0 AND BSD-3-Clause AND LicenseRef-Acme-custom

PackageCopyrightText: Copyright 2018 Acme Software
In an SPDX document:

Tracking corresponding source code:

Filename: /bin/binary
SPDXID: SPDXRef-binary

Filename: /src/sources.tar.gz
SPDXID: SPDXRef-sources

Relationship: SPDXRef-binary DISTRIBUTION_ARTIFACT SPDXRef-sources
In an SPDX document:

Providing other metadata about your software:

FileName: /bin(binary
SPDXID: SPDXRef-binary
Relationship: SPDXRef-pom METAFILE_OF SPDXRef-binary

FileName: /pom.xml
SPDXID: SPDXRef-pom

Create and Refine Policies

• How SPDX helps
  • It doesn’t, at least not directly
  • SPDX documents carry information that your policies and processes can act upon
Create and Refine Policies

From an SPDX document:

Deciding whether to fail a build because of licenses:

- LicenseConcluded: NONE
- LicenseConcluded: JSON
  - LicenseRef-Weird-Proprietary ✗
- LicenseConcluded: GPL-3.0-only
- LicenseConcluded: Apache-2.0 ✓
Create and Refine Policies

From an SPDX document:

Deciding whether to fail a build because of vulnerabilities:

PackageName: LibFoo
PackageVersion: 1.2.17a
Relationship: SPDXRef-mycode
HAS_PREREQUISITE SPDXRef-Libfoo
Create and Refine Policies

From an SPDX document:

Or automate any other kind of findings and actions:

FileName: /src/myCode.c
FileCopyrightText: Copyright OldCompanyName

Copyright notices to update
SPDX – Bridging the Gaps in Compliance