Describing License Information in SPDX

It’s Easier than you think

Gary O’Neall, Source Auditor Inc.
Agenda

- What is SPDX, who is using it, and why?
- The complex – examples of SPDX documents
- A simple start – using SPDX license ID’s
  - Overview of the SPDX license list
  - Including license identifiers in your source code
  - Tools to support using license IDs
- A bit more work – SPDX documents for original source
  - Example of an SPDX document
  - Tools to support creating an SPDX document
- Advanced SPDX usage – automating your supply chain
- Questions
What is SPDX?

- **Standard:**
  - A standard format for communicating the components, licenses and copyrights associated with a software package

- **Guiding principles:**
  - Human and machine readable
  - Focus on capturing facts; avoid interpretations

- **Vision:**
  - To help reduce redundant work in determining software license information and facilitate compliance
Who is using SPDX?

- SPDX has over 20 active participants
- Active participation by organizations in the middle of the software supply chain (e.g. TI, HP, Wind River, Samsung)
- Tools providers are incorporating SPDX for import and export of software “Bill of Materials”
- Consultants using the format as a standard output
- Open source projects using the SPDX license identifiers (e.g. Das U-Boot, Ruby projects, GitHub)
Our suppliers aren’t giving us complete licensing information for open source packages.

I don’t mind vetting our code, but I’m sure this imported package has been analyzed a dozen times before.

Every customer wants a bill of materials in a different form.
Example SPDX document

Formats

- **Tag/Value**

```xml
<rdf:Description rdf:nodeID="A0">
  <copyrightText>
    * <br>(C) 1999 by Randolph Chung &<br>
    <copyrightText>
    <licenseInFile rdf:resource="http://spdx.org/license">
    <licenseConcluded rdf:resource="http://spdx.org/license">
    <fileType rdf:resource="http://spdx.org/rdf/terms#file">
    <checkSum rdf:nodeID="A1">$
    <fileName networking/hostname c e <filename>
    <rdfType rdf:resource="http://spdx.org/rdf/terms#file">
  </copyrightText>
</rdf:Description>
```

- **RDF**

```xml
<rdf:Description rdf:nodeID="A2">
  <checkSumValue>
    dc9a437e03f31ab0b4c705996da3f88b288d6777d
  </checkSumValue>
  <algorithm rdf:resource="http://spdx.org/rdf/terms#checksumAlgorithm_sha1" />
  <description rdf:resource="http://spdx.org/rdf/terms#checkSum" />
</rdf:Description>
```

- **Spreadsheet**

![Spreadsheet Table Image]
Starting Simple: SPDX License List Identifiers

```plaintext
projects / u-boot.git / blob

summary | shortlog | log | commit | commitdiff | tree
history | raw | HEAD

Merge branch 'master' of git://git.denx.de/u-boot-mpc85xx

[u-boot.git] / post / post.c

1 /*
2  *(C) Copyright 2002
3  * Wolfgang Denk, DENX Software Engineering, wd@denx.de.
4  *
5  * SPDX-License-Identifier: GPL-2.0+
6 */
7```
Advantages to License ID’s

- Reduces effort to determine the actual license used in a source file
- List of (common) open source* licenses
  - Over 300 licenses and over 20 common exceptions
- SPDX.org/licenses contains the text, reference URL’s, templates for matching, whether OSI approved and headers
- Backed by an active organization which maintains the license list
- Matching guidelines to help determine if the license text matches the text
More Info on the License List

- Human and machine readable at http://spdx.org/licenses
- RDFa machine readable access
- JSON file at http://spdx.org/licenses/licenses.json
- For tools support, see the tech report “Accessing SPDX Licenses”: http://spdx.org/publications/tool-documentation/accessing-spdx-licenses
- License ID’s in Source: http://wiki.spdx.org/view/Technical_Team/SPDX_Meta_Tags
A bit more work: Using License Expressions

- Expresses a combination of licenses (e.g. “Dual Licensing”)
- AND – Conjunctive licenses
- OR – Disjunctive licenses
- With – Exceptions
- + - Or later versions
Examples

- Example: GPL-2.0+ OR MIT – Early versions jQuery

- Example: GPL-2.0-with-autoconf-exception AND GPL-3.0+ AND LGPL-2.1+ AND GPL-2.0+

- Example: (GPL-2.0+ OR MIT) AND BSD-3-Clause
More Initial Effort: SPDX for Original Source

- Requires some tooling since the verification code is generated by checksums
Tools to Support SPDX Creation

- Wind River website (spdx.windriver.org)
- FOSSology for analyzing licenses (https://fosslogyspdx.ist.unomaha.edu/)
- Maven plug-in (https://github.com/goneall/spdx-maven-plugin)
- Yocto – UNO built SPDX extension
- FOSSosology (SPDX support in beta) (http://www.fossology.org/projects/fossology/wiki/WhoUsesFOSSology)
- Others… - See spdx.org/tools
Example: Maven Plug-in

- Leverages existing meta data
- Additional configuration parameters for SPDX specifics
- Builds relationships based on dependencies
- Can be included in a “parent POM” for a larger open source project

```xml
<plugins>
  <plugin>
    <groupId>org.spdx</groupId>
    <artifactId>spdx-maven-plugin</artifactId>
    <executions>
      <execution>
        <id>build-spdx</id>
        <goals>
          <goal>createSPDX</goal>
        </goals>
      </execution>
    </executions>
    <configuration>
      <spdxDocumentNamespace>
        http://spdx.org/documents/spdx-toolsv2.0-rc1
      </spdxDocumentNamespace>
      <defaultFileComment>test file comment</defaultFileComment>
      <documentComment>Initial submission for the SPDX document bake-off</documentComment>
      <documentAnnotations>
        <param>
          <annotator>Person: Gary O'Neill</annotator>
          <annotationDate>2015-07-23T18:30:22Z</annotationDate>
          <annotationType>OTHER</annotationType>
          <spdxComment>Initial submission for the SPDX 2.0 document bake-off</spdxComment>
        </param>
      </documentAnnotations>
    </configuration>
  </plugin>
</plugins>
```
Advantages to SPDX Documents

- Can be very descriptive
  - In addition to licenses, contains copyrights, relationships, creation and origin information
- Can be easily translated to human readable form (e.g. a spreadsheet)
- Some verification supported (package verification code, checksums)
- Integrates with various license compliance tools
- REALLY helps the downstream consumers – changes from identification to (maybe) verification
Tips for Keeping it Simple: Not all fields are required

Document Creation Information
2.1 SPDX Version.
2.2 Data License
2.3 SPDX Identifier
2.4 Document Name
2.5 SPDX Document Namespace
2.8 Creator
2.9 Created

Package Information
3.1 Package Name
3.2 Package SPDX Identifier
3.7 Package Download Location
3.8 Package Verification Code
3.12 Concluded License
3.13 All Licenses Information from Files.
3.14 Declared License.
3.16 Copyright Text

File Information
4.1 File Name
4.2 File SPDX Identifier
4.4 File Checksum
4.5 Concluded License
4.6 License Information in File
4.8 Copyright Text
Tips for Keeping it Simple

- Beware of over-complicating the relationships


- Refer to other example SPDX documents, such as the results of the SPDX “bake-off” results

- Monitor the spdx tech mailing lists
Example – SPDX Tools:

```xml
<rdf:RDF ...
<referencesFile>
  <File rdf:nodeID="A2">
    <fileName>resources/stdlib licenses/GFDL</fileName>
    <fileType rdf:resource="http://spdx.org/rdf/terms#fileType_other"/>
    <rdfs:comment></rdfs:comment>
    <licenseComments></licenseComments>
    <licenseInfoInFile rdf:resource="http://spdx.org/licenses/Apache-2.0"/>
    <checksum>
      <Checksum>
        <checksumValue>922fa7cdc2b9a02aa6e71b82e564f3c53625e73c</checksumValue>
        <algorithm rdf:resource="http://spdx.org/rdf/terms#checksumAlgorithm_sha1"/>
      </Checksum>
    </checksum>
    <copyrightText>Copyright (c) 2012, 2013, Source Auditor Inc.</copyrightText>
    <noticeText>* Copyright (c) 2013 Source Auditor Inc.
* Licensed under the Apache License, Version 2.0 (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.apache.org/licenses/LICENSE-2.0
* Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.</noticeText>
  </File>
</referencesFile>
```
Using SPDX in the Software Supply Chain

- Importing upstream SPDX files
- Validating upstream SPDX files
- Scanning and generating SPDX for packages without SPDX files
- Creating output SPDX files

- Will typically involve multiple packages, multiple documents and relationships
Advantages to using SPDX in the Software Supply Chain

- Maintain fidelity of the licensing data
- Reduce the effort for downstream consumers
- Reduce the effort for suppliers who are using SPDX
Package Relationships Example

Package

MQ Apache

Jetty Web Container

javax.servlet

SPDX Doc

Apache MQ

contains

Jetty Web Container

contains

javax.servlet

contains

continue'd by

continue'd by
Tips on keeping it simple in the supply chain

- Reference existing SPDX documents using external SPDX document relationships
- Choose a common “namespace” for your SPDX documents
- Use relationships and annotations to provide any corrections
More Information

- spdx.org
  - General information
- SPDX License list: spdx.org/licenses
- SPDX Tools: spdx.org/tools
- wiki.spdx.org
  - Workgroup wiki’s for Technical, Business, and legal teams
  - Contains information on joining the mailing lists and calls
Open source (Apache 2.0 licensed) Java code providing access the SPDX Standard Libraries

Binary .jar files available:

Source code available:
- git.spdx.org – spdx-tools project
- [https://github.com/spdx/tools](https://github.com/spdx/tools)

Bugs / enhancement ideas can be reported:
- bugs.linuxfoundation.org – project SPDX/tools
- Github issues tracking

Contributions welcome!
More Information

- See the tech report “Accessing SPDX Licenses”: http://spdx.org/publications/tool-documentation/accessing-spdx-licenses

- Join the spdx-tech mailing list

- Email me at gary@sourceauditor.com