Sharing Open Source License and Copyright Data with SPDX

Martin Michlmayr, Hewlett-Packard

tbm@hp.com
Agenda

- The Problem
- Overview of the SPDX Spec
- Tools to Help
- About the Beta Program
- SPDX and Distributions
- The Future
Open Source Compliance: The Challenge

License info for OSS is not provided in a consistent, easy-to-use format.

One OSS Package = Many Licenses
Open Source Compliance: The Challenge

Creating an accurate bill of materials requires effort & research.

Companies combine OSS with other software.

Software Bill of Materials (BOM)
Open Source Compliance: The Challenge

Customers

Supplier 1

Supplier 2

The effort is repeated at each step in the supply chain
Scope of the Problem

- Prior to distributing a collection of software the contents of each package to be included needs to be reviewed, to ensure compliance with all the licenses in the code being redistributed.

- Supply chain for products now requires software copyright and licensing information for lawsuit avoidance and risk mitigation.

- A package’s declared license may not always match the licenses of individual files inside the package itself.

- A package may consist of thousands of files with different licenses in the files.

- Need a standard way of referring to the legal compliance “bill-of-materials” of a software package and be able exchange information with others efficiently and accurately.
1000s of packages and complex dependencies between them...

- Package contents evolve over time
  - Different versions can have different licenses
  - Declared license of a package is not always accurate
  - Package with different license has “useful” routines (that potentially get included)

- Package dependency/requisite hierarchy can have incompatibilities
  - Hidden/enveloped package in dependency chain
  - Incidental packages get included by accident
  - Not all OSS licenses compatible with each other
Software Package Data Exchange (SPDX)

- SPDX - A standard format for communicating the components, licenses and copyrights associated with a software package.

- Key pillar in Linux Foundation’s Open Compliance Program
What it Means to You

Embedded & SW Supply Chains
- Save Time/Money
- Better Compliance

Open Source Developers
- Help Users Comply With Your Licenses

Consumers of SW & OSS
- Understand Licensing of the Code You Use
Participants

OSS Organizations
End-Users
Integration & Services
Device OEMs
Applications
OS Distributions
Systems
Semiconductor Vendors

Participation is from a range of organizations and across various roles

Copyright Linux Foundation 2011
Working Group Operation

- The working group runs similarly to an open source project without centralized constitution or bylaws
- Intellectual property contributed by participants members is covered under the Creative Commons license (CC-BY-3.0)

Structure
- General Meeting and mailing list
- Teams: Technical, Business, Legal

Very inclusive process
- Self-subscription for interested participants
- Those willing to “do” can influence direction
- Mailing lists, wiki, phone calls, BOFs…
- http://spdx.org
Specification Goals

- A file format for license and copyright information to accompany packages
  - Guiding Principle: **Just the facts – no interpretations**

- A standardized short form to refer to the official version of common licenses

Benefits
- Allows easy exchange of license information between companies reducing burden on both suppliers and consumers
- Avoids due diligence redundancy where the same source code package is analyzed multiple times by different receivers
- Provides a unified method for exchanging license information
The SPDX File

- **Analysis Info**: Info about source of this SPDX™ file
- **Package Info**: Associates SPDX™ file with a software package (tarball, zip, archive)
- **Licensing Info**: Text of licenses that are not in SPDX™ standard list
- **File Info**: License associated with each file. Refers to std.license list on SPDX™ or a non-standard license included with the file
- **Review**: Log of 3rd party reviews

*File is in RDF/XML form; can be converted to tag value or spreadsheet.*
Analysis Information

- SPDX Version (used in creation of SPDX file)
- How this info was generated
  - Manual review (who, when)
  - Tool (id, version, when)
- Creator (who created the SPDX file)
- Creator License (license of data in file)
Package Information

- Formal Name of Package (Full name given by originator and version information)
- Package File Name (Name package obtained under (.tar, .rpm, etc.))
- Package Checksum (to unambiguously map file to a package)
- Package Download Location (download URL)
- Licensing for Package
  - Declared License- License that has been asserted for the package
  - Concluded License- License that Creator has concluded
- Copyright Text
- Description of Package (optional)
This section is for licenses not on the standard list.
Aim for ~90% coverage with standard short forms; no plan to be exhaustive
Background: ~20 licenses cover most code; OSI recognizes 67 licenses as open source
File Information

- File Name
- File Type (source, binary, archive)
- File CheckSum
- Concluded License (license determined by SPDX file creator)
- Copyright Text
- Artifact Project Name (from which project it came)
Review

- Reviewer
- Review Date
- Review Comment
The SPDX List of “Standard Licenses”

**SPDX license repo**

<table>
<thead>
<tr>
<th>License Identifier</th>
<th>Recognized Exceptions</th>
<th>Full name of License</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFL-3.0</td>
<td></td>
<td>Academic License 3.0</td>
</tr>
<tr>
<td>ACPL-3.0</td>
<td></td>
<td>(GNU) Affero General Public License v3</td>
</tr>
<tr>
<td>APL</td>
<td></td>
<td>Adaptive Public License</td>
</tr>
<tr>
<td>ASL-2.0</td>
<td></td>
<td>Apache License 2.0</td>
</tr>
<tr>
<td>APSL-2.0</td>
<td></td>
<td>Apple Public Source License 2.0</td>
</tr>
<tr>
<td>Artistic-2.0</td>
<td></td>
<td>Artistic license 2.0</td>
</tr>
<tr>
<td>AAL</td>
<td></td>
<td>Attribution Assurance License</td>
</tr>
<tr>
<td>BSD-4-Clause</td>
<td></td>
<td>BSD 4-clause “Original” or “Old” License</td>
</tr>
<tr>
<td>BSD-3-Clause</td>
<td></td>
<td>BSD 3-clause “New” or “Revised” License</td>
</tr>
<tr>
<td>BSD-2-Clause</td>
<td></td>
<td>BSD 2-clause “Simplified” or “FreeBSD” License</td>
</tr>
<tr>
<td>BSL-1.0</td>
<td></td>
<td>Boost Software License 1.0</td>
</tr>
<tr>
<td>CATOSL-1.1</td>
<td></td>
<td>Computer Associates Trusted Open Source License 1.1</td>
</tr>
<tr>
<td>CC-BY-1.0</td>
<td></td>
<td>Creative Commons Attribution 1.0</td>
</tr>
<tr>
<td>CC-BY-NC-1.0</td>
<td></td>
<td>Creative Commons Attribution Non Commercial 1.0</td>
</tr>
<tr>
<td>CC-BY-ND-1.0</td>
<td></td>
<td>Creative Commons Attribution No Derivatives 1.0</td>
</tr>
<tr>
<td>CC-BY-SA-1.0</td>
<td></td>
<td>Creative Commons Attribution Share Alike 1.0</td>
</tr>
<tr>
<td>CC-BY-NC-SA-1.0</td>
<td></td>
<td>Creative Commons Attribution Non Commercial Share Alike 1.0</td>
</tr>
<tr>
<td>CC-BY-ND-2.0</td>
<td></td>
<td>Creative Commons Attribution No Derivatives 2.0</td>
</tr>
<tr>
<td>CC-BY-SA-2.0</td>
<td></td>
<td>Creative Commons Attribution Share Alike 2.0</td>
</tr>
<tr>
<td>CC-BY-NC-SA-2.0</td>
<td></td>
<td>Creative Commons Attribution Non Commercial Share Alike 2.0</td>
</tr>
</tbody>
</table>

- List of most common licenses (100+)
- Include common exceptions
- Standardized license names
- Exact text of licenses
- Available on SPDX website – URLs won’t change
SPDX Tools

- Objectives
  - Reduce the effort of creating, consuming and validating SPDX documents
  - Provide a translation from the technical document (e.g. RDF/XML or tag-value format) and a more readable format
  - Provide a mechanism for validating SPDX documents
  - Enable contributions and review of the tool implementation by the broader technical community through open source licensing
SPDX Tools

- Tooling Needs
  - Viewers
  - Validators
  - Create SPDX file
  - Read SPDX file

- Open Source Tools
  - FOSSology
  - “Pretty Print” viewer for SPDX

- Commercial Tools
  - Scanning tools expected to provide SPDX support
SPDX Viewer

- Command line driven Java application which formats a valid SPDX document into a text file
- Validates the SPDX document and provides somewhat detailed error messages when there are parsing errors
- Status: Currently being updated to the latest specifications.
SPDX Translator

- Converts a spreadsheet containing SPDX information into a valid SPDX/RDF file
- Converts a valid SPDX/RDF file into a spreadsheet
- Java based command line tool
- Requires the spreadsheet be in a very specific format
- A spreadsheet template with the specific column names and order will be provided as part of the tool
- Status: In development.
The SPDX Beta Program is designed to ensure the SPDX specification meets users needs
Timeline

Jan 2010 - Spec started
Jun 2010 - Spec v1 Beta
Mar 2011 - Spec v1 RC
Apr 2011 - Start Beta
Jun 2011 - Beta Feedback
Aug 2011 - Spec v1 Final

Beta Testing between April and June 2011
Beta Process

Company 1

• Create your SW BOM
• Convert to an SPDX file

Company 2

• Convert from SPDX
• Review SW BOM
Feedback We Need

- How well did the SPDX format cover information you need to exchange?
- What was effort involved in converting to/from SPDX format?
- Did the SPDX format help you understand what was in the software?
- How did the standard license information work for you?
- Can you see how using SPDX could save your organization time/money?
- Are any changes needed to make SPDX meet your needs?
- What other tooling is needed for SPDX?
SPDX and Distros

- Distros do a lot of license review when adding a new open source package
- No unified format used by distros (e.g. Debian uses free-form debian/copyright; DEP5 proposal for machine-readable debian/copyright)
- SPDX (or derived format) would allow more collaboration between distros: would reduce work for everyone, make it easier to find potential problems, etc
- Distros are a good way of working with upstream: getting SPDX adopted by open source projects directly.
The Future

- Successful SPDX Beta Test
- Final release of SPDX
- Adoption of SPDX: SPDX as required standard in supply chain
- Adoption of SPDX in upstream projects
- Development of tagging mechanism for source code