Software Package Data Exchange (SPDX™) Specification

Kate Stewart,[a] Phil Odence,[b] Esteban Rockett[c]

(a) Ubuntu Release Manager, Canonical, Inc.; (b) Senior Vice President of Business Development, Black Duck Software; (c) Senior Counsel, Motorola Mobility, Inc.

DOI: 10.5033/ifosslr.v2i2.45

Abstract
The goal of the Software Package Data Exchange (SPDX™) specification is to enable companies and organizations to share license and component information (metadata) for a software package and related content with the aim of facilitating license and other policy compliance. The specification is being developed through collaboration between technical, business and legal professionals from a range of organizations to create a standard that addresses the needs of various participants in the software supply chain.

Keywords
License Compliance; Copyright Identification; Specification Format; Software Package Metadata; Software Supply Chain.

Info
This item is part of the **Tech Watch** section of IFOSS L. Rev. For more information, please consult the relevant section policies statement. This article has been independently peer-reviewed.

Background

Companies at all points in the software supply chain are becoming conscious of the need to treat open source just like any other third party code. They need to know and document the components in the products and software they are consuming and distributing. There are a variety of reasons for this, not the least of which is to make sure they understand their legal obligations. Thus the need for a common approach to sharing information about software packages and their related content has never been greater. Breaking down information silos is still a work in progress. The Software Package Data Exchange working group[1] was formed originally as a FOSSBazaar

1 http://www.spdx.org/

sponsored effort[2] and is now a part of the Linux Foundation's Open Compliance Program[3]. The working group's goal is to define a way to share copyright and license information about a software packages and common licenses in that package, down to the file level.

Why is a standardized specification needed?

Innovation happens very rapidly in the open source ecosystem, often by developers by building on top of the work of other developers. To do this, source code files, that have been created as part of one project under a specific license may be copied and reused in another project that may be under a different license. This mixing and matching of licenses, creates problems for those companies reusing and redistributing software packages that contain this combined software. It becomes a real challenge for them to figure out what they need to do to comply with the licenses that govern their software packages. By creating a standard way of summarizing the licensing and copyright information to the file level and providing a way to double check that the summary actually matches the code, a standard makes the task of figuring out what license are in effect much easier. This permits creation of a software "bill of materials" that can be passed with the actual software, throughout the supply chain, saving considerable analysis effort at every step. Simply saying your company is doing the right thing is not enough: savvy consumers in the supply chain want proof to limit the risk of non-compliance with licenses. Suppliers themselves welcome a single standard format for disclosing open source rather than having to respond to each customer's request using a unique format.

What does the SPDX™ specification consist of?

The SPDX™ effort has focused on coming up with a way to summarize the discoverable facts about code content and ownership. By providing a 'defined format of file to accompany any software package,' the effort eases the burden of exchange of license information between companies. The standard defines a format for sharing: facts that deal with identification, facts that provide overview information, and facts that provide file-specific information about the software package.

Facts that deal with a software package's identification (metadata) included in the specification are:

- Version of the SPDX™ specification is in use
- Unique identifier (based on a cryptographic hash algorithm) representing a unique identifier that correlates with this specific software package
- Method by which information was generated (who, when, tools used, etc.)
- Independent audit information (sign-off/reviewed by)

Facts that provide overview information about a software package's content include:

2 https://fossbazaar.org/
3 http://www.linuxfoundation.org/programs/legal/compliance

- Formal Name
- Package File Name
- Download Location
- Declared License(s)
- Detected License(s)
- Copyrights and Dates

Facts that are specific to a software package's file-specific properties:

- File Name (including subdirectory)
- File Type (source or binary)
- Detected license(s) governing file (from file)
- Copyright owners and dates (if listed)

Because of the license orientation of the specification, the working group is also committed to providing standardized license references. The specification includes:

- License names
- Unique identifiers for common open source licenses
- Mechanisms for handling non-standard licenses.

The SPDX™ specification does not attempt to transmit legal judgement, but rather provides a format for a summary of the facts from which professionals (perhaps using other tools) may make judgements.

How far along is the development and what are the next steps?

The Version 1.0 beta form of the specification is available for download[4], but it is just a starting point. It has had some road testing, but has not been driven by the public, so the group's focus is shifting to driving practical applications and incorporating the inevitable feedback before we release the official version 1.0. The group is assembling a list of key projects for which to create SPDX™ reports, and to get create those reports by any method possible. Initially we expect members of the group to roll up their sleeves on this live testing, but we are also working hard with tool vendors (proprietary and open source) to create other options for generating these reports. We anticipate the need to develop new tools (e.g. syntax checkers and reading and displaying tools) to enable this development, as well as training materials for educating others on using the standard.

4 http://www.spdx.org/spec/current

Interested in learning more and helping out?

If you want to join our volunteer effort, and help make it better, there is information on how to participate available on our web site[5]. Sub-groups with their own mail lists have recently formed around technical, business, and legal issues, and depending where your interests are, all are open and welcome new members to collaborate on the specific topic areas.

Conclusion

Getting the SPDX™ specification adopted across the ecosystem will be a challenge. We need participation and support from key Linux distros and package maintainers, legal experts, tool developers (commercial and open source) and package consuming organizations as well. With major players in all those categories already on board, and with the support from FOSSbazaar and the Linux Foundation, the pieces are finally coming together to let us achieve our goal of a useful specification.

About the authors

Kate Stewart *is Ubuntu's Release Manager at Canonical, Inc. After reviewing way too many standard projects for license and copyright info in her prior job at Freescale Semiconductor, Inc., she found a group of folk equally frustrated, and set out collaborating with them on defining a specification for sharing package licensing and copyright facts between projects and organizations.*

Phil Odence *is Vice President of Business Development for Black Duck Software, makers of enterprise application development tools that address management, compliance and security challenges associated with open source. In that role, he is responsible for expanding Black Duck's reach, image and product breadth by developing partnerships in the multi-source development, legal and open source ecosystem.*

Esteban Rockett *is a Senior Counsel at Motorola Mobility, Inc. He is Mobility's lead software intellectual property counsel, including serving as legal lead for its open source contributions and compliance, mobile application stores, and digital content licensing and delivery.*

5 http://www.spdx.org/wiki/spdx/participation-guidelines