



» The Open Compliance Program

A Common Software Package Data Exchange Format:

1.0 Release Update and Discussion

.....
Phil Odenice, VP, Black Duck Software

Kate Stewart, Ubuntu Release Manager, Canonical

SPDX™ Working Group

A White Paper By The Linux Foundation
<http://www.linuxfoundation.org>

The debate about the increasing role of open source in the software community is over. At LinuxCon North America 2010, Forrester Research's Jeff Hammond announced that open source had "crossed the chasm" into the mainstream. Mark Driver of Gartner went further in a November 2010 report¹, stating: "Open source is ubiquitous, it's unavoidable... having a policy against open source is impractical and places you at a competitive disadvantage."

Certainly there exists a vast and growing pool of open source-licensed applications, but open source code is even more pervasive through components embedded in almost every piece of software developed today. Over half a million open source projects can be found on the Internet². Superimpose that on the overall ubiquity of software in products from cars to handsets to power plants to medical devices and it becomes clear that open source code is flowing through countless supply chains in almost every industry.

Companies at all points in the supply chain are becoming aware of the need to treat open source just like any other third party code. They need to track and document the components in the products and software they are consuming and distributing for a variety of reasons, 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 the content and licensing of software packages has never been greater. Breaking down information silos is still a work in progress. For more than a year, the SPDX working group has been tackling one of the toughest obstacles: the difficulty of collaborating and sharing information about software packages and their licenses.

Do the right thing

It's hard. With over 2,000 different software licenses for software freely available on the Internet, license proliferation is a major headache for software development organizations as well as for companies redistributing software packages in their products. Scope is one problem: From the Free Beer license to the GPL family of licenses to platform-specific licenses such as Apache and Eclipse, the sheer number and variety of licenses makes it difficult for companies to "do the right thing" with respect to the software components in their products and applications.

Each license carries within it the author's definition of how the software can be used and re-used. Permissive licenses like BSD and MIT make it easy; software can be redistributed and developers can modify code without the requirement of making changes publicly available. Reciprocal licenses, on the other hand, place varying restrictions on re-use and redistribution. Woe to the developer who snags a bit of code after a simple web search without understanding the ramifications of license restrictions. Woe to the company that doesn't provide their developers the education, guidance and tools to avoid such issues.

License compliance-a first step to doing the right thing

While most companies want to do the right thing with regard to license compliance, the lack of a standard format for key license information complicates matters. Many approaches to ensuring license compliance exist — from handcrafted spreadsheets to free software options to commercial tools — so an overarching standard for exchanging package data has been elusive.

1. Available to Gartner clients.

2. From Black Duck Software KnowledgeBase <http://www.blackducksoftware.com/protex/kbase>



The problem is exacerbated as software development crosses organizational boundaries. Software supply chains are not set up to communicate package content and licensing information. Suppliers, if they are cataloging the information at all, have their own formats and conventions. Consumers in the chain are increasingly asking for this information, but again, there seem to be as many different formats as entities asking for it.

That situation is changing, however. First, through the education efforts of many people and organizations, companies are becoming more conscious of the issues and the need to pay attention. The Linux Foundation, for example, now offers the [Open Compliance Program](#), whose goals are to:

- Boost adoption of Linux and other FOSS by making license compliance ever-easier to achieve
- Increase awareness and understanding of FOSS compliance responsibilities
- Make available free resources that can help companies establish their FOSS compliance programs

An element of that program, the [Software Package Data Exchange® \(SPDX\)](#) tackles the format problem head-on by defining a standard for exchanging package content. The SPDX working group, an on-going grass-roots effort sponsored by the Linux Foundation, includes representatives from dozens of organizations — software, systems and tool vendors, foundations and systems integrators — committed to creating a useful standard.

Really, though, who cares about software package data exchange® formats?

It's not just software development managers and lawyers who care about having a standardized approach to software license compliance. Any corporation that uses and/or distributes software packages has a stake in the outcome. IT managers concerned with compliance care, executives at companies buying software packages care, and software development organizations care—especially distributed global development teams for whom it is especially difficult to get visibility into licenses and their obligations. Some of this interest is being driven by more and more companies demanding that suppliers provide them with a Bill of Materials (that states clearly which software components are in a specific package and which licenses are involved). Simply asserting that your company is doing the right thing is not enough: Savvy consumers of software want proof to limit their own risk of non-compliance with license obligations. Suppliers welcome a single standard format for disclosing open source rather than having to respond to each customer's request in a unique way.

With the release of V1.0 of the SPDX standard, the SPDX group achieved its initial goal: to create a common software package data exchange format to simplify the discovery, collection and sharing of information about software packages and related content. The standard promises to save time, improve the accuracy of license data collection, and simplify compliance with software licenses. To fully realize this promise, the group will continue to focus on adoption and improving the standard to better meet the needs of its users.



The scope of the problem

Most companies have well-established practices that govern the release and distribution of software. In bygone days, the only path for third-party software into an organization was via its purchasing organization. But open source use has created additional wrinkles... lots of them! Any developer with a browser has complete access to half a million projects. Because most software products developed today are composed of mixed code acquired from many different sources (in many cases, without the knowledge of product and development managers and executives) the software supply chain has become extremely complex.

Breaking the problem down into its component pieces gives a sense of its scope:

- Prior to distributing a collection of software, the contents of each package to be included need to be reviewed to ensure compliance with all the licenses in the code being redistributed.
- Therefore, the supply chain for products requires developers to create a 'software pedigree' that includes information necessary to avoid misuse and mitigate risk.
- A software package's declared license may not always match the licenses of individual files inside the package. In fact, a typical software package may consist of thousands of files with different licenses.
- Code re-use may have introduced code fragments and components covered by a range of incompatible licenses.

Adding to the urgency of this problem, software packages with more than one version have complex interdependencies. As software evolves over time, new code components may be included that have different licenses, conceivably at any level of the software. Code re-use is a great way to speed up development, but it can introduce license conflicts over time. After all, with more than 2,000 licenses out there, it stands to reason that not all permutations of licenses will be compatible.

This is why the industry has needed a standard way of referring to the legal compliance "bill of materials" of a software package. It's necessary to standardize a way to exchange information about the licenses contained in a software package efficiently and accurately, so as to minimize the collective overhead involved in compliance along the supply chain.

Just the facts, please

Although most software licenses convey intent and may require interpretation, the SPDX effort has focused on rendering facts versus judgment calls (which might be made differently by different organizations). The SPDX working group does not attempt to apply legal judgment, for example, by classifying a license as "BSD-like." The standard errs on the side of being very explicit about the exact license in question, and when there isn't a visible fact available, provides a mechanism for explaining how the conclusion was made.

Version 1.0 of the SPDX standard provides a format for identifying the package, the package content, and file level information as well. What follows is more detail on the kind of information in an SPDX file.

Software package identification and analysis data include:

- Which version of the SPDX specification is in use and how the information in the file can be

shared.

- How the information was created; the SPDX specification defines a way to provide:
 - Manual/visual analysis of code (who, when)
 - Tools used (id, version)
- Reviews of the information. SPDX includes the possibility of a multi-person "signoff/reviewed by" process including when the review was performed

Information about the software package being described includes:

- Formal name of the package
- Version of the package
- Name of package file
- Description of the package
- Download location
- Unique identifier (to tie the file to a specific package and formulated so the SPDX file can be included inside the package without affecting the value)
- License declared within the package
- License concluded by the creator of the SPDX file, which could be different
- List of all licenses found in the package at the file level
- Copyright text and dates

Properties of files within the software package include:

- File Name (including subdirectory)
- File Type (source, binary, archive, other)
- Checksum
- License Information contained in file
- License concluded by SPDX creator to apply to the file (if, for example, there is no license in the file, but there is a copying file in the same directory)
- Copyright owners (if listed)
- Copyright dates (if listed)
- Associated project from which the file may have come

To streamline the file size while allowing for unambiguous license identification, the specification also includes a set of short licenses identifiers for popular licenses (e.g. "Apache-2.0" for the Apache 2 license). Each identifier is tied to specific license text on the spdx.org website with a unique, permanent URL. There is also a mechanism to efficiently include full license text for non-standard licenses. As of the release of V1.0, the standard list comprises almost 200 licenses, including all of the OSI approved licenses. Recently, OSI adopted the same naming conventions and there is also consistency with the Debian conventions.

The SPDX file can be expressed using a tag-value format or using standard Resource Description Framework (RDF). Both formats are human and machine-readable and can be translated to each other and to a spreadsheet format.

So where is SPDX™ now?

On spdx.org, of course! Along with a wealth of supporting information and tools.

Version 1.0 of the specification was released in August 2011. It has undergone "road testing" through a beta process in which pairs of supply chain partners tried handing off package information as they would in practice. Valuable feedback was incorporated back into the specification before the initial release. In addition, the release includes an extensive set of supporting materials (including this whitepaper), version 1 of the standard license list and some



initial tools, all housed on the website.

The tools are open source and available under the Apache 2.0 license. They enable reading from and writing to a spreadsheet or tag value format. The group's expectation and hope is that SPDX-adopters will integrate these tools into their existing infrastructure and processes.

Where does it go from here?

Version 1.0 is a very solid starting point, but expectations are that it will evolve rapidly as adoption broadens. Early adopters will have great influence on the evolution of the spec.

The group expects some 1.x releases to add new features and correct newly identified issues. Work has already begun on incorporating more hierarchy into the specification. Today, the spec handles package level and file level information, but nothing in between. Oftentimes packages are composed of other packages, so we anticipate the ability to contain SPDX files within SPDX files to be a valuable simplifying addition to the 2.0 version of the specification.

Getting SPDX adopted across the ecosystem is a challenge. We need participation and support from key Linux distros and package maintainers, tool developers (commercial and open source) and package consuming organizations as well. With major players in all those categories already on board, along with strong support from the Linux Foundation, we are confident SPDX will become a critical industry asset.

Participate

If you're interested in participating in the SPDX group, the website, <http://spdx.org>, should provide all of the information you need, but feel free to contact the authors. Work is divided between three teams: Technical, Business and Legal. It's very easy to get on our mailing lists to monitor activity, and there are plenty of opportunities to participate, contribute to and steer the evolution of SPDX.

About the Authors

Phil Oden (podence@blackducksoftware.com) is Vice President of Business Development for Black Duck Software, makers of enterprise app 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. Prior to Black Duck, Phil served in senior marketing, sales and business development positions with Empirix, High Performance Systems and Teradyne. He has an AB in Engineering Science and an MS in System Simulation from Dartmouth College.

Kate Stewart (stewart@linux.com) is Canonical's Ubuntu Release Manager. Prior to Canonical she managed the team doing the open source development for the Power architecture at Freescale Semiconductor, Inc. for the last 10 years. In this role, she managed the development of Linux board support packages and basic enablement for new silicon. This required that she understand the open source licenses and pioneer policies with the worldwide developers, corporate legal team, senior management, and third party partners to permit software distribution as well as enable code contributions to the community Linux, gnu compiler, and U-Boot projects. She has a BS



in Computer Science from University of Manitoba, and a MM in Computer Science from University of Waterloo.

About the Open Compliance Program

The Linux Foundation's Open Compliance Program is the industry's only neutral, comprehensive software compliance initiative. By marshaling the resources of its members and leaders in the compliance community, the Linux Foundation brings together the individuals, companies and legal entities needed to expand the use of open source software while decreasing legal costs and FUD. The Open Compliance Program offers comprehensive training and informational materials, open source tools, an online community (FOSSBazaar), a best practices checklist, a rapid alert directory of company's compliance officers and a standard to help companies uniformly tag and report software used in their products. The Open Compliance Program is led by experts in the compliance industry and backed by such organizations as the Adobe, AMD, ARM Limited, Cisco Systems, Google, HP, IBM, Intel, Motorola, NEC, Novell, Samsung, Software Freedom Law Center, Sony Electronics and many more. More information can be found at <http://www.linuxfoundation.org/programs/legal/compliance>.



The Linux Foundation promotes, protects and standardizes Linux by providing unified resources and services needed for open source to successfully compete with closed platforms.

To learn more about The Linux Foundation, the Open Compliance Program or our other initiatives please visit us at <http://www.linuxfoundation.org/>.

