Advancing the Software Package Data Exchange: An Update on SPDX

Phil Odence,[a] Scott Lamons,[b] Jilayne Lovejoy[c]

(a) Vice President of Corporate and Business Development, Black Duck Software; (b) Program Manager, HP Open Source Program Office; (c) Corporate Counsel, OpenLogic, Inc.

DOI: 10.5033/ifosslr.v5i2.89

Abstract
Since 2010, the Software Package Data Exchange, a Linux Foundation work group, has made great progress. This article provides an overview of advancements on the specification itself, survey results on use, adoption by corporate users and FOSS communities, and future plans and initiatives.

Keywords
Law; information technology; Free and Open Source Software; SPDX; Software Package Data Exchange; software licensing; copyright; bill of materials.

Introduction

SPDX® (or Software Package Data Exchange®) is a specification for exchanging package content, copyright, and licensing information between software supply chain partners. Organized under the Linux Foundation, the SPDX work group introduced SPDX to the international legal community in an article in the *International Free and Open Source Software Law Review* Vol. 2, Issue 2 when the specification was going through beta testing.[1] This article is an update on the current state of the work and future direction, focusing on a look at current attitudes regarding SPDX adoption, tooling, and plans for version 2.0.

While the specification has evolved since the original publication, the work group's mission has remained constant:

> *Develop and promote adoption of a specification to enable any party in a software supply chain, from the original author to the final end user, to accurately communicate the licensing information for any piece of copyrightable material that such party may create, alter, combine, pass on, or receive, and to make such information available in a consistent, understandable, and re-usable fashion, with the*

1 Stewart, K., Odence P., Rockett, E. (2010) 'Software Package Data Exchange (SPDX™) Specification', IFOSS L. Rev., 2(2), pp 191 – 196 DOI: 10.5033/ifosslr.v2i2.45

aim of facilitating license and other policy compliance.[2]

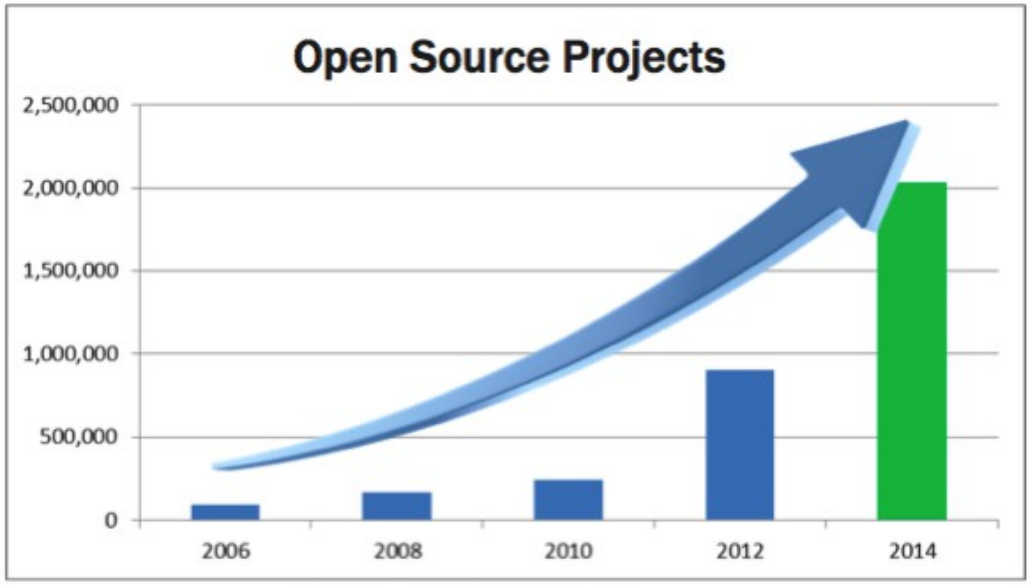
Establishing a common data format enables producers and consumers of software (and the tool vendors that support them) to build processes and tooling that reduce the initial effort and rework involved in understanding and communicating what is in a software package. Thus, a standard format allows more effort be expended on licence compliance. After all, license compliance can only begin once all software and associated licenses have been identified in a particular code base.

The content of an SPDX document comprises, among other things, information definitively identifying the software package, and package level and file level licensing and copyright information. It also provides metadata about the analysis itself: who created the file, when, and how.

The SPDX work group consists of representatives from companies and organizations who use or are considering using the SPDX standard. The work group operates much like a meritocratic, consensus-based community project; that is, anyone with an interest in the project can join the community, contribute to the specification, and participate in the decision-making process.

State of the System

Free and open source software (FOSS) projects continue to multiply at an accelerating rate. Since 2010, the number of freely available software projects on the Internet has climbed from about a quarter of a million to over a million and is projected to top two million in 2014.



Source: Black Duck Software

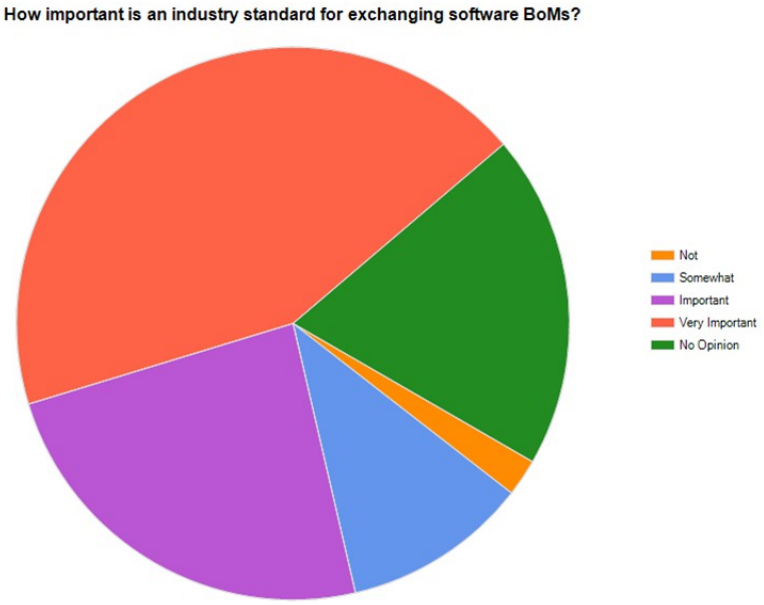
At the same time, while the overall awareness of the need to manage open source software and licensing is clearly on the rise, adoption of some kind of governance program lags far behind. In a late 2012 study of the European automotive industry, BearingPoint found that while 85% of

2 http://spdx.org/about-spdx

respondents reported that their companies were deploying FOSS, only 2.3% had open source compliance tooling in place.[3] The SPDX work group's hypothesis is that at least part of the problem is the lack of an industry standard. A standard would allow for the consistent and common exchange of license information, protect tooling investments, spur a broader range of tools, and allow tooling to interoperate with each other.

Survey of SPDX Awareness and Adoption

The SPDX work group conducted a survey in spring of 2013 to collect information regarding understanding and adoption of SPDX by corporate and community members and organizations. The survey was publicized via posting online (with the link provided on the SPDX website), Linux Foundation events, various open source mailing lists, and word of mouth. About 100 people completed the survey with a majority of responses coming from technical resources at a mixture of small and large companies worldwide.[4] Most notably, about two-thirds of the respondents said that "an industry standard for exchanging software bill of materials (BoMs)" was *very important* or *important*, thus validating the over-arching goal of SPDX.



Source: SPDX Survey[4] conducted during May 2013

More notable points from of the results of the survey are discussed in the subsequent sections of this article.

Adoption

As with other standards, adoption is often slower than expected, but interest is recently on the rise from both open source projects and companies. The SPDX survey cited above revealed that as of

3 http://www.bearingpoint.com/en-uk/7-5601/study-foss-management/
4 See http://wiki.spdx.org/view/Business_Team/Surveys for a summary of the responses and download of complete survey.

yet only a handful of organizations are producing or requiring SPDX documents from suppliers and most were, at best, experimenting internally. On the other hand, the need was clear; many more were intending to use SPDX in the future and only a tiny fraction expected to use another format. The longest journey starts with a single step, however, and SPDX is clearly beyond that.

Corporate Adoption

Companies tend to be private about their contractual arrangements, which makes it hard to comprehensively track who is using SPDX, planning to use it, experimenting internally, and so forth. As to date, only a few companies have come forward publicly regarding their adoption or use.

Wind River, a supplier of a Linux-based embedded systems platform, has been a proactive advocate and early adopter of SPDX. Wind River Linux 5, a commercial grade version of Yocto, ships over 700 SPDX files, one for each of the packages in its distribution. Wind River also asks all ISVs to include SPDX files with their software deliverables and have assisted several ISVs in creating an SPDX file for their offering. Additionally, the company uses SPDX data in its IP Compliance Review process and distributes SPDX files to its customers to meet any open source disclosure requirements. Taking this all one step further, Wind River hosts a website that provides free high quality samples of SPDX files, as well as a free cloud service to enable anyone to generate an SPDX file for any uploaded package. The main purpose of these efforts is to promote the adoption of SPDX among Wind River customers and the software community at large.[5]

At LinuxCon North America in September 2013, an engineer from Samsung Electronics delivered a talk titled, *Piloting SPDX in Samsung: Case Studies and Experiences,* which discussed Samsung's internal experimentation and development around the use of the SPDX standard and provided feedback to the work group.[6] Texas Instruments and Alcatel-Lucent are also using SPDX for internal communications.[7] Other large companies like HP and Cisco are heavily involved in the development of the specification, presumably with adoption on the horizon.

While few companies are yet taking a public position, the survey indicates that a number of companies have plans to both require and offer SPDX documents to accompany exchanged software packages. Discussions held under the Chatham House Rule[8] at LinuxCon Japan and the Linux Collaboration Summit this past year, as well as inquiries to the work group, indicate increasing interest and experimentation. Interest regarding adoption has come from company representatives in a wide variety of industry sectors. One large auto manufacturer has started requiring SPDX documents from suppliers and a large telecom company is doing the same.

Community Adoption

The SPDX survey reinforced the “chicken and egg” nature of starting a standard; that is, adoption breeds adoption. Upstream FOSS projects are one of the keys to getting the cycle rolling in the right direction. Working with them provides an opportunity to improve the reach of the standard, fostering a broad adoption base with downstream consumers.

The SPDX work group is communicating with a number of projects and foundations about adoption of the standard. Recent collaboration with the Yocto Project is focused on integrating the

5 See spdx.windriver.com and http://spdx.windriver.com/pkg_upload.aspx
6 http://linuxconcloudopenna2013.sched.org/event/2faecbb5c51ea6089cdc5eb5159bc154#.UfAYgGQ6U4Q
7 See http://wiki.spdx.org/view/Business_Team/Adoption
8 http://www.chathamhouse.org/about-us/chathamhouserule

production of SPDX documents into the Yocto build process. The joint project utilizes the FOSSology SPDX plug-in[9] developed at the University of Nebraska Omaha[10] to identify licenses in Yocto project packages, prepare package and file level license information, and produce and archive SPDX documents. In addition, discussions with the Apache Software Foundation and OpenMAMA both offer potential upstream projects where SPDX could impact broader adoption of the standard.

The SPDX License List

Perhaps the best starting point for adoption is the SPDX License List, which is a standardized index of over 200 of the most common open source licenses.[11] Every license on the list contains a short identifier (e.g., Apache-2.0), a long name (Apache License 2.0), a url to the license text, and the official header for labelling source code files, if the license designates one. In 2011, the Open Source Initiative (OSI) announced that it was adopting and standardizing on the SPDX short names, which was a big step in helping the industry move toward using a consistent set of names for open source licenses. As of DEP5, Debian supports the SPDX short identifiers as does OpenSUSE.[12] The SPDX legal team continues work to ensure the SPDX License List includes licenses found on other community lists, such as FSF and Fedora.

For tool providers this will make detection of open source licensing much more reliable, leading to more accurate generation of SPDX data files. As of version 2.1.1, FOSSology, the open source license scanner, adopted the SPDX License List short identifiers.[13] Likewise, Ninka supports SPDX identifiers.[14] Commercial tools from Black Duck Software and NexB also use the SPDX License List to reference licenses. Known companies using the SPDX License List include Texas Instruments, Siemens, Micro Focus, and Wind River.

Besides the obvious advantage of having a reliable and common way to accurately report a given open source license, the SPDX License List also has the potential to be used as a license declaration.[15] The SPDX License List short identifiers provide an easy and concise way to identify the license for a particular file in the source code.[16] Already, Composer, a dependency manager for PHP, and npm, a package manager for node, have adopted or encourage the use of the SPDX License List short identifiers.[17] U-Boot, a popular open source boot loader for embedded devices, is using SPDX short identifiers as its standard for specifying licensing in files.[18] This enables unambiguous license information in a single line and eases automated processing. This kind of adoption by open source projects greatly simplifies the creation of SPDX documents.

Tooling

While the aforementioned survey pointed to a number of factors that are important to broad

9 http://ocrl.unomaha.edu/organizational-participation-in-open-communities/tooling/
10 http://www.ist.unomaha.edu/
11 http://spdx.org/licenses/
12 http://dep.debian.net/deps/dep5/ and http://en.opensuse.org/openSUSE:Packaging_guidelines; also see: http://www.linuxfoundation.org/news-media/announcements/2011/08/widespread-industry-support-spdx-10
13 http://www.fossology.org/projects/fossology/wiki/Release_Notes#220-Released-June-28-2013 and http://lwn.net/Articles/556850/
14 http://www.linuxfoundation.org/news-media/announcements/2012/08/supporting-comments-spdx-11
15 See http://wiki.spdx.org/view/Technical_Team/SPDX_Meta_Tags for a working draft proposal.
16 Indeed, the SPDX work group members are not the only ones who think so, as evidenced by this post: http://hakre.wordpress.com/2012/07/25/using-the-spdx-license-list-for-tagging-and-linking/
17 http://getcomposer.org/doc/04-schema.md#license; https://github.com/isaacs/npm/pull/3673
18 http://spdx.org/news/2013-10-22/spdx-releases-version-1.2-of-the-specification

adoption of such a standard, tooling for producing SPDX files was considered very important by the most survey participants. In the past six months there have been some great advances along this dimension.

The SPDX group hosts a handful of open source tools for validating, reading, and translating SPDX documents. Now several FOSS and commercial tools have added the ability to produce SPDX documents.

At the 2013 Linux Collaboration Summit, the SPDX work group hosted a "bake off" or interoperability testing session to compare the output of several tools as well as some manually generated SPDX files. The testing analysed output from two open source tools (FOSSology, hosted by the University of Nebraska Omaha and Ninka, from University of Victoria) and one commercial tool (Black Duck). SourceAuditor has driven development of the SPDX open source tools, and Wind River also shared results from their internal processes and tooling. The extensive analysis uncovered the need for further clarity in the specification in order to ensure more consistency among differently-generated SPDX documents. This sharing represents significant progress against what is considered one of the biggest impediments to adoption. As these tools advance, it will become increasingly practical for organizations to use SPDX to exchange software BoMs information.

The Future

As with any open source project, the future will emerge from the activities of all the companies and individuals involved. But there are some clear directions for the project. As tools implementing the specification have become a reality, the group has been able to begin a cycle of testing the tools and at the same time, essentially testing the specification. Comparing the output of a variety of tools has enabled the group to identify some limitations and ambiguities. The work group recently released version 1.2 of the specification, which addresses these issues.

Beyond that, there are two areas where SPDX needs to be enhanced: hierarchy and signing. In regards to hierarchy, the current specification provides a fairly "flat" structure for licensing and copyright information with package and file level views. In other words, there is no explicit way to identify files for one package (and associated licenses) contained within other packages. Based on internal and outside input and due to license compliance requirements that are dependent on how software interacts, the work group has identified a requirement for accommodating the hierarchical nature of software. Because applications are made up of components, which can in turn be made up of other components, this suggests the desirability of a similar structure for SPDX documents to be able to describe the contents of those packages, and for SPDX documents to comprise other SPDX documents of lower level components.

The idea of signing is to allow creators of SPDX documents to associate their name with the work as long as the document isn't modified. This provides the ability for a SPDX document recipient to make a judgement call as to the reliability of the information provided therein. It is related to hierarchy in that in a hierarchy signing should be maintained by a branch such that if pieces get combined or modified it remains clear who did what.

Beyond the technical evolution of the specification, the SPDX legal team continues to evolve the license list and process around it, with the latest developments being around guidelines for matching to license text. Such guidelines will help ensure consistent matching among tools and SPDX document creators so that when any SPDX document identifies a license using a SPDX

License List short identifier, it can be relied upon to be consistent with the identification of that same license in other SPDX documents. There already exists a glossary of synonyms, for example to allow matching the American "license" to the English "licence," and some other guidelines about handling spaces, punctuation, and copyright notices. Recent work focuses on handling variable text like the copyright holder names in the BSD licenses, as well as the overall implementation for the license matching guidelines.

Conclusion

There is clearly a need for a standard format for exchanging software bill of material information. SPDX is viable today for all open source projects and several early adopter companies. The specification will continue to improve and evolve, especially as more users and potential users from corporations to community groups become involved in shaping the standard.

About the authors

Phil Odence *is the Vice President of Corporate and Business Development at Black Duck Software where he is responsible for all corporate and business development activities. A frequent speaker at open source industry events, Phil chairs the Linux Foundation's Software Package Data Exchange working group and participates on the GENIVI marketing team. Phil has over 20 years of software industry experience. He earned an AB in Engineering Science and a MS in System Simulation from the Thayer School of Engineering at Dartmouth College.*

Scott Lamons *works in HP's Open Source Program office which is responsible for the companies' open source policy, review process, and compliance related activities. He has been a member of HP's Open Source Review Board (OSRB) since 2005. Over this time he has reviewed over 3000 open source proposals from teams across HP and has been involved in delivering company-wide training and consulting on a variety of open source activities. He also works closely with HP's vendors and partners in the open source community and currently co-lead's the SPDX business team.*

Jilayne Lovejoy *participates in open source industry groups including co-leading the SPDX legal team. Jilayne is also a frequent speaker and writer on topics related to open source licensing and compliance. Previously, she was the corporate counsel at OpenLogic where she helped ensure that OpenLogic's scanning and compliance software met the needs of users and assisted customers with understanding open source license compliance and policy considerations to reduce barriers to open source software adoption. She earned her BA from the University of Colorado, Boulder and her JD from the Sturm College of Law at the University of Denver.*