How to manage FOSS compliance information in an ecosystem

Nov 17th, 2017
Lei Maohui, Fujitsu
leimaohui@cn.fujitsu.com
Fujitsu’s contributions to Yocto community

- Data comes from yocto (2016-10-31 ~ 2017-11-01)

**Contributions in yocto**

<table>
<thead>
<tr>
<th>Layers</th>
<th>Changesets</th>
</tr>
</thead>
<tbody>
<tr>
<td>poky</td>
<td>122</td>
</tr>
<tr>
<td>oe-core</td>
<td>122</td>
</tr>
<tr>
<td>meta-oe</td>
<td>97</td>
</tr>
<tr>
<td>bitbake</td>
<td>0</td>
</tr>
</tbody>
</table>
Fujitsu’s contributions to Yocto community

- Data comes from yocto (2016-10-31 ~ 2017-11-01)

### Top changeset contributors by employer

<table>
<thead>
<tr>
<th>No.</th>
<th>Employer</th>
<th>Changesets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intel</td>
<td>8454</td>
</tr>
<tr>
<td>2</td>
<td>Wind River</td>
<td>1616</td>
</tr>
<tr>
<td>3</td>
<td>Fujitsu</td>
<td>343</td>
</tr>
<tr>
<td>4</td>
<td>Axis Communications</td>
<td>181</td>
</tr>
<tr>
<td>5</td>
<td>simens</td>
<td>130</td>
</tr>
</tbody>
</table>

### Developers with the most changesets

#### poky

<table>
<thead>
<tr>
<th>No.</th>
<th>Our Developer</th>
<th>Changesets</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>Huang Qiyu</td>
<td>81 (0.7%)</td>
</tr>
<tr>
<td>38</td>
<td>fan.xin</td>
<td>28 (0.2%)</td>
</tr>
</tbody>
</table>

#### oe-core

<table>
<thead>
<tr>
<th>No.</th>
<th>Our Developer</th>
<th>Changesets</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Huang Qiyu</td>
<td>81 (1.1%)</td>
</tr>
<tr>
<td>29</td>
<td>fan.xin</td>
<td>28 (0.4%)</td>
</tr>
</tbody>
</table>

#### Meta-oe

<table>
<thead>
<tr>
<th>No.</th>
<th>Our Developer</th>
<th>Changesets</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Huang Qiyu</td>
<td>71 (1.7%)</td>
</tr>
<tr>
<td>28</td>
<td>fan.xin</td>
<td>12 (0.3%)</td>
</tr>
</tbody>
</table>
Agenda

Introduction of SPDX

• Background of SPDX
• What is SPDX
• Who are working for SPDX
• The status of SPDX specification

meta-spdxscanner in Yocto

• What is Yocto
• SPDX in Yocto
• meta-spdxscanner
• Features in meta-spdxscanner
• How to use meta-spdxscanner

Manage SPDX files

• SPDX file works in OpenChain
• Manage SPDX files by dnf in your ecosystem
• Future work
Introduction of SPDX

- Background of SPDX
- What is SPDX
- Who are working for SPDX
- The status of SPDX specification
- Issues we met
Background of SPDX

FOSS

GPLv2
GPLv3
ZLIB
BSD
Apache

MIT

License: MIT
Copyright ……

……
What is SPDX (1/2)

What is SPDX

• The full name of SPDX is Software Package Data Exchange, which is a standard format for communicating the components, licenses and copyrights associated with a software package.

Vision of SPDX

• achieve license compliance with minimal cost across the supply chain.

SPDX will be a good solution, if a SPDX implementation can generate SPDX file including license information automatically.

Obtain details from
• https://spdx.org/tools
What is SPDX (2/2)

**Important or useful tags**
- SPDXVersion
- DataLicense
- Creator
- PackageName
- PackageOriginator
- PackageVersion
- PackageHomePage
- PackageLicenseDeclared

**A sample of SPDX file**

```
SPDXVersion: SPDX-2.0
DataLicense: CC0-1.0
PackageName: Foo
PackageOriginator: David A. Wheeler
PackageHomePage: https://github.com/david-a-wheeler/spdx-tutorial/
PackageLicenseDeclared: MIT
```
## Who are working for SPDX

### Technical Team
- **Primary responsibility**
  - Drafts the specification
  - Develops documentation templates, samples and tools.
- **Delivered**
  - SPDX Spec (2.1, 2.0, 1.2, 1.1, 1.0)
  - Tool (fossology)
  - Spreadsheet Template
- **Recent**
  - New spec
  - New license list

### Legal Team
- **Primary responsibility**
  - Supports and provides recommendations to the SPDX working groups regarding licensing issues.
  - Maintains the [SPDX License List](http://spdx.org/licenses)
  - Promotes the SPDX specification to the legal community at-large
- **Delivered**
  - License Expression Syntax
  - License Inclusion Guidelines (Background)
  - Dealing with Public Domain within SPDX Files
- **Recent**
  - Joint Call with Tech Team
  - License List

### Outreach Team
- **Primary responsibility**
  - Launch activities for new versions of the SPDX specification.
  - Outreach
  - Participation in events;
  - The SPDX website
- **Delivered**
  - Launch for 1.0 and 1.1
  - Process for Adding to License List (Draft)
  - SPDX Vision & Mission Discussion Document
  - SPDX Vision & Mission Statements (Final Draft)
- **Recent**
  - Working on tool to generate test files for scanners

---

*Obtain details from*

[http://spdx.org/participate](http://spdx.org/participate)
[http://wiki.spdx.org/view/General_Meeting/Minutes](http://wiki.spdx.org/view/General_Meeting/Minutes)
The status of SPDX Specification (1/2)

**History**

- **Origin**
  - The “SPDX” name was adopted

- **SPDX v1.0**
  - Specification 1.0 released

- **SPDX v1.1**
  - Specification 1.1 released

- **SPDX v1.2**
  - Specification 1.2 released

- **SPDX v2.0**
  - Specification 2.0 released

- **SPDX v2.1**
  - Specification 2.1 released

**Features in SPDX**

- **SPDX v1 File**
  - Creation Information
  - Package Information
  - Other Licensing Information
  - File Information
  - Review Information

- **SPDX v2.0 File**
  - Document Creation Information
  - Package Information
  - File Information
  - Other Licensing Information
  - Relationships
  - Annotations

- **SPDX v2.1 Document contains:**
  - Document Creation Information
  - Package Information
  - File Information
  - Snippet Information
  - Other Licensing Information
  - Relationships
  - Annotations
Kernel v4.14 (2017-11-17) added one-liners come from SPDX

```plaintext
# SPDX-License-Identifier: GPL-2.0
```

The status of SPDX Specification (2/2)
We met such an issue in checking CVE.

<table>
<thead>
<tr>
<th>CVE-ID</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Severity Rating</strong></td>
<td><strong>Fix Information</strong></td>
</tr>
<tr>
<td><strong>Vulnerable Software Versions</strong></td>
<td><strong>SCAP Mappings</strong></td>
</tr>
</tbody>
</table>

**Description**
Race condition in the rmtree and remove_tree functions in the [File-Path module before 2.13](https://metacpan.org/pod/File-Path#module) for Perl allows attackers to set the mode on arbitrary files via vectors involving directory-permission loosening logic.

**References**
Note: References are provided for the convenience of the reader to help distinguish between vulnerabilities. The list is not intended to be complete.

- CONFIRM: http://cpansearch.perl.org/src/JKEENAN/File-Path-2.13/Changes
- CONFIRM: https://rt.cpan.org/Ticket/Display.html?id=121951

But it is difficult to judge the version of perl SPDX file.

```
$ grep File-Path perl-5.24.1.spdx
FileName: ./cpan/File-Path/lib/File/Path.pm
FileName: ./cpan/File-Path/t/FilePathTest.pm
FileName: ./cpan/File-Path/t/Path.t
FileName: ./cpan/File-Path/t/Path_root.t
FileName: ./cpan/File-Path/t/Path_win32.t
FileName: ./cpan/File-Path/t/taint.t
```
We met another issue in checking CVE.

<table>
<thead>
<tr>
<th>CVE-ID</th>
<th>Learn more at National Vulnerability Database (NVD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVE-2017-11164</td>
<td>- Severity Rating - Fix Information - Vulnerable Software Versions - SCAP Mappings</td>
</tr>
</tbody>
</table>

**Description**
In PCRE 8.41, the OP_KETRMAX feature in the match function in `pcre_exec.c` allows stack exhaustion (uncontrolled recursion) when processing a crafted regular expression.

**References**
Note: References are provided for the convenience of the reader to help distinguish between vulnerabilities. The list is not intended to be complete.

- **MISC:** http://openwall.com/lists/oss-security/2017/07/11/3
- **BID:** 99575
- **URL:** http://www.securityfocus.com/bid/99575

**Assigning CNA**
MITRE Corporation

**Date Entry Created**
20170710

---

$ grep pcre glib-2.0-2.50.3.spdx
Relationship: SPDXRef-DOCUMENT DESCRIBES SPDXRef-file-pcre_h-a607-c18ce65f
Relationship: SPDXRef-DOCUMENT DESCRIBES SPDXRef-file-pcre_byte_order_c-8b9c-fa7dc829
Relationship: SPDXRef-DOCUMENT DESCRIBES SPDXRef-file-pcre_chartables_c-3397-ca3605c6
......
meta-spdxscanner in Yocto

- What is Yocto
- SPDX in Yocto
  - Status of spdx.bbclass
  - Our team contribution for Yocto-SPDX
- meta-spdxscanner
- Features in meta-spdxscanner
- How to use meta-spdxscanner
What is Yocto

The Yocto Project is an open source collaboration project that help you create custom Linux-based systems for embedded products.

https://www.yoctoproject.org/
SPDX in Yocto (1/2)

Yocto

Poky (Core Layer)
- meta
- meta-poky
- meta-yocto
- meta-yocto-bsp
- ...

Other Layers
- meta-openembedded
- meta-cgl
- meta-virtualization
- ...

spdx.bbclass

Bitbake taskflow
- do_fetch
- do_unpack
- ...
- do_spdx
- ...

Copyright 2017 FUJITSU COMPUTER TECHNOLOGIES LIMITED
SPDX in Yocto (2/2)

- Yocto+SPDX is not very compliant with SPDX Specification.
- Can’t get package information
- Only support python2
- Only support fossology2.
- Hard to use, you have to set up the fossology2 server.
Our team contribution for Ycoto-SPDX

2015
Submitted a patch to make spdx.bbclass support SPDX 1.2

2016
Submitted a patch to make spdx.bbclass support SPDX 2.0

2017/03
Create a new layer meta-spdxscanner based on python2

2017/11
Make meta-spdxscanner support lid based on python3

Base on spdx.bbclass
meta-spdxscanner
meta-spdxscanner (1/3)

meta-spdxscanner

• Git Repository: https://github.com/dl9pf/meta-spdxscanner

• Our contribution to make Yocto+SPDX support SPDX2.0

• Project Activity: Maintained by our team.

<table>
<thead>
<tr>
<th>Layer name</th>
<th>Description</th>
<th>Type</th>
<th>Repository</th>
</tr>
</thead>
<tbody>
<tr>
<td>meta-spdxscanner</td>
<td>SPDX support</td>
<td>Distribution</td>
<td><a href="https://github.com/dl9pf/meta-spdxscanner">https://github.com/dl9pf/meta-spdxscanner</a></td>
</tr>
</tbody>
</table>
meta-spdxscanner (3/3)

- FOSS

- Patches come from 3rd party

Yocto

- meta-spdxscanner
- meta
- meta-oe
- meta-……

SPDX files
Features in meta-spdxscanner

- **Supports dosocsv2**
- **Supports LiD**
- **Can work with Yocto 2.3**
- **More relationship information**
- **Support SPDX 2.0**

**Meta-spdxscanner**

- **class**
  - lid-scan.bbclass
  - dosocs.bbclass

- **recipes**
  - python3-dosocs2
  - fossology-nomos
  - python-lid
  - perl
  - glib
  - ......

- **bbappend**
Supports Dosocsv2 and LiD

What is DoSOCSv2

- dosocsv2 is a command-line tool for managing SPDX 2.0 documents and data. ([Website](#))

- By default, DoSOCSv2 uses nomos comes from fossology as it’s license scanner.

What is LiD

- Qualcomm OSTG (Open Source Technology Group) LiD (License Identifier) tool scans source code and identifies the license and the license text region using a standard set of license templates from sources like SPDX. ([Website](#))
Dosocsv2 vs LiD (1/3)

File format
- File created by Dosocsv2 is compliance with SPDX2.0
- File created by LiD isn’t compliance with SPDX

```bash
$ dosocs2 oneshot cpio-2.11
dosocs2: cpio-2.11: package_id: 1
dosocs2: running nomos on package 1
cccccpio-2.11: document_id: 1
```

**SPDXVersion: SPDX-2.0**
- DataLicense: CC0-1.0
- DocumentNamespace: sqlite:///home/leimh/.config/dosocs2/dosocs2.sqlite3/cpio-2.11-f6eb-4fa85311
- DocumentName: cpio-2.11
- SPDXID: SPDXRef-DOCUMENT
- DocumentComment: <text></text>

```bash
$ license-identifier -l /home/test/zlib-1.2.11 | less
=== Found 0 results for '/home/test/zlib-1.2.11/CMakeLists.txt':
=== Found 0 results for '/home/test/zlib-1.2.11/ChangeLog':
=== Found 3 results for '/home/test/zlib-1.2.11/FAQ':
Summary of the analysis
Name of the input file: /home/test/zlib-1.2.11/FAQ
Matched license type is Vim
Score for the match is 0.0513
Rank for the match is ScoreOutOfRange
License text begins at line 343.
License text ends at line 364.
Start byte offset for the license text is 15451.
End byte offset for the license text is 16418.
The found license text has the score of 0.00108
The following text is found to be license text
------BEGIN------
41. I'm having a problem with the zip functions in zlib, can you help?
There are no zip functions in zlib. You are probably using minizip by
Giles Vollant, which is found in the contrib directory of zlib. It is not
part of zlib. In fact none of the stuff in contrib is part of zlib. The files
in there are not supported by the zlib authors. You need to contact the
authors of the respective contribution for help.

42. The match.asm code in contrib is under the GNU General Public License. Since it's part of zlib, doesn't that mean that all of zlib falls under the GNU GPL?
------
```
Dosocsv2 vs LiD (2/3)

License Text Region Identification

- Nomos (the scanner used by default in Dosocsv2) finds snippets
- LiD finds the whole license text

$ FileName: ./contrib/iostream2/zstream.h
SPDXID: SPDXRef-file-zstream_h-d034-fcdf1afd
FileType: SOURCE
FileChecksum: SHA256:
d0343e0c57ff58008b6f29643d289c72713aa2d653fe3dcd2e939fc77e7e20b6
LicenseConcluded: NOASSERTION
LicenseInfoInFile: LicenseRef-MIT-style
LicenseComments: <text></text>
……..

LicenseID: LicenseRef-MIT-style
LicenseName: MIT-style
ExtractedText: <text>Permission to use, copy, modify, distribute and sell this software and its documentation for any purpose is hereby granted without fee, provided that the above copyright notice appear in all copies</text>
LicenseCrossReference: 
LicenseComment: <text>found by nomos</text>

$ license-identifier -l /home/test/zlib-1.2.11 | less
Name of the input file: /yocto/work001/build/poky/build-poky-dnf/tmp/work/i586-poky-linux/zlib/1.2.11-r0/zlib-1.2.11/contrib/iostream2/zstream.h
Matched license type is NTP
Score for the match is 0.269
Rank for the match is 3
License text beings at line 2.
License text ends at line 15.
Start byte offset for the license text is 6.
End byte offset for the license text is 642.
The found license text has the score of 0.027
The following text is found to be license text

-----BEGIN-----
* Copyright (c) 1997
* Christian Michelsen Research AS
* Advanced Computing
* Fantoftvegen 38, 5036 BERGEN, Norway
* http://www.cmr.no
* 
* Permission to use, copy, modify, distribute and sell this software
* and its documentation for any purpose is hereby granted without fee,
* provided that the above copyright notice appear in all copies and
* that both that copyright notice and this permission notice appear
* in supporting documentation. Christian Michelsen Research AS makes no
* representations about the suitability of this software for any
* purpose. It is provided "as is" without express or implied warranty.

-----END-----
## Dosocsv2 vs LiD (3/3)

<table>
<thead>
<tr>
<th>Item</th>
<th>DoSOCsv2</th>
<th>LiD</th>
</tr>
</thead>
<tbody>
<tr>
<td>License</td>
<td>GPLv2</td>
<td>BSD-3-Clause</td>
</tr>
<tr>
<td>Can create SPDX files</td>
<td>SPDX 2.0</td>
<td>-</td>
</tr>
<tr>
<td>Include whole license text</td>
<td>-</td>
<td>√</td>
</tr>
<tr>
<td>Performance (e.g. zlib-1.2.11)</td>
<td>1min(1st) 5s(2nd)</td>
<td>21min</td>
</tr>
<tr>
<td>scanner</td>
<td>Fossology(nomos)</td>
<td>LiD</td>
</tr>
<tr>
<td>Project Activity (<a href="http://www.openhub.net">http://www.openhub.net</a>)</td>
<td>Very Low</td>
<td>Low</td>
</tr>
<tr>
<td>GUI</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Command Line</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>
Can work with Yocto 2.3

Build system

~ Yocto 2.2

python-lxml python-lexml
python-cryptography
python-…

Python2

make tar unzip

wget …

Meta-spdxscanner

python-dosocs2 python-magic
python-dosocs2

Build system

Yocto 2.3 ~

python3-lxml python3-lexml
python3-cryptography
python3-…

Python3

make tar unzip

wget …

Meta-spdxscanner

python3-dosocs2 python3-magic
python3-dosocs2
python3-…
More relationship information (1/2)

More useful relationship informations are add into meta-spdxscanner.

You can get informations such as:

- Issue-1 can be resolved

```
$ grep File-Path perl-5.24.1.spdx
Relationship: perl CONTAINS File-Path-2.12
FileName: ./cpan/File-Path/lib/File/Path.pm
FileName: ./cpan/File-Path/t/FilePathTest.pm
FileName: ./cpan/File-Path/t/Path.t
FileName: ./cpan/File-Path/t/Path_root.t
FileName: ./cpan/File-Path/t/Path_win32.t
FileName: ./cpan/File-Path/t/taint.t
```

- Issue-2 can be resolved

```
$ grep pcre glib-2.0-2.50.3.spdx
Relationship: glib STATIC_LINK system-pcre
Relationship: SPDXRef-DOCUMENT DESCRIBES
SPDXRef-file-pcre_h-a607-c18ce65f
Relationship: SPDXRef-DOCUMENT DESCRIBES
SPDXRef-file-pcre_byte_order_c-8b9c-fa7dc829
Relationship: SPDXRef-DOCUMENT DESCRIBES
SPDXRef-file-pcre_chartables_c-3397-ca3605c6
......
```
How to use meta-spdxscanner

- Generate SPDX File from Yocto building.

1. get meta-spdxscanner
   - $ git clone https://github.com/dl9pf/meta-spdxscanner

2. Add meta-spdx and it’s depences
   - $ cd [yocto_build_dir]
   - $ tail –n 3 conf/conf/bblayers.conf
     - ./yocto/community/meta-spdxscanner
     - ./yocto/community/meta-openembedded/meta-oe
     - ./yocto/community/meta-openembedded/meta=python

3. Enable dosocs.bbclass
   - $ cd [yocto_build_dir]
   - $ tail –n 4 conf/local.conf
     - INHERIT += "dosocs"
     - ARCHIVER_MODE[src] = "patched"
     - SPDX_DEPLOY_DIR = "/yocto/build/poky/build-poky/spdx-out"

4. Start building
   - $ cd [yocto_build_dir]
   - $ bitbake core-image-minimal

SPDX Files
Manage SPDX files

- Manage SPDX files by dnf in your ecosystem
- SPDX file works in OpenChain
  - What is OpenChain
  - Make SPDX files work in OpenChain
- Future work
What is DNF

- **DNF** is a software package manager that installs, updates, and removes packages on **RPM**-based Linux distributions. It automatically computes dependencies and determines the actions required to install packages. DNF also makes it easier to maintain groups of machines, eliminating the need to manually update each one using **rpm**. Introduced in Fedora 18, it has been the default package manager since Fedora 22.

Why Yum was forked into DNF

- An undocumented API—this meant more work for developers. In order for developers to do what they needed, it was often necessary to browse through the Yum code base just to be able to write a call. This meant development was very slow.
- Python 3—Fedora was about to make the shift to Python 3 and Yum wouldn’t survive this change, whereas DNF can run using either Python 2 or 3.
- Broken dependency solving algorithm—this has been an Achilles heel of the Fedora package manager for a long time. DNF uses a state-of-the-art satisfiability (SAT)-based dependency solver. This is the same type of dependency solver used in SUSE’s and openSUSE’s Zypper.
Accompanied with the package files and SRPM packages, SPDX files are created to manage license information.
Manage SPDX files by dnf in your ecosystem (1/2)

There is a Demo!

You can reference to:
https://github.com/ubinux/dnf/tree/dnf-yocto-dev
OpenChain (1/2)

What is OpenChain

- The **OpenChain** Project helps to identify and share the core components of a high quality Free and Open Source Software (FOSS) compliance program. OpenChain builds trust in Open Source by making things simpler, more efficient and more consistent. It is the industry-standard for managing Open Source compliance across the supply chain. ([Website](#))

- OpenChain aims to help companies avoid potential pitfalls in the compliance process:
  1. Intellectual Property (IP) pitfalls
  2. License Compliance pitfalls
  3. Compliance Process pitfalls
Make SPDX files work in OpenChain

an example of an enterprise process comes from OpenChain

https://wiki.linuxfoundation.org/_media/openchain/openchain-curriculum-for-1-1.pdf
Make SPDX files work in OpenChain

an example of an enterprise process comes from OpenChain

Generate SPDX files by Source Code Scanner.

Discover and Resolve issues by using SPDX files.

Review and approve compliance record of FOSS software components

Compile notices for publication

Post publication verification

Review SPDX files and Approvals FOSS software Components.

Outgoing Software

Notices & Attributions

Written Offer

SPDX files

Identify FOSS components for review

Scan or audit source code — and — Confirm origin and license of source code

Resolve any audit issues in line with company FOSS policies

Record approved software/version in inventory per product and per release

Verify source packages for distribution — and — Verify appropriate notices are provided

Archive Source Packages and SPDX files.

Example of Compliance Management End-to-End Process

https://wiki.linuxfoundation.org/_media/openchain/openchain-curriculum-for-1-1.pdf
Future work

- Make dnf-host support UI.
- Submit dnf patch to Yocto.
- Add more SPDX sanners into meta-spdxscanner.
- Go on adding useful relationship informations into meta-spdxscanner.
- Release meta-spdxscanner following the step of Yocto.
Any Questions?
shaping tomorrow with you