

Creating curated data -  
transparency is key!

- Curation database – Why and How
- Structure and provided artifacts
- How we provide transparency
  - Why we use FOSSology
- The curation database in action – a showcase
- Potential usages
- Final remarks

- Our goal is to lower the required effort as much as possible for all who want to make use of OSS in a license compliant way
- One of the tasks in OSS compliance work is the analysis of OSS packages. We believe that it does not make any sense that everyone doing checks of packages again and again. Which currently happens thousand fold in many organizations
- A database with curated license and copyright information is the fundament to integrate the license compliance process in the CI/CD pipelines and to automate license compliance process



Picture by Geralt / pixabay.com / Pixabay License

- To achieve this we develop, share and improve the artifacts commonly used to fulfill the requirements of the different Free and Open Source Software licenses by applying the Open Source Software development principles. I.e. we do it via an Open Source project
- We provide transparency on how the curations are made
- We provide curated package analysis in different formats



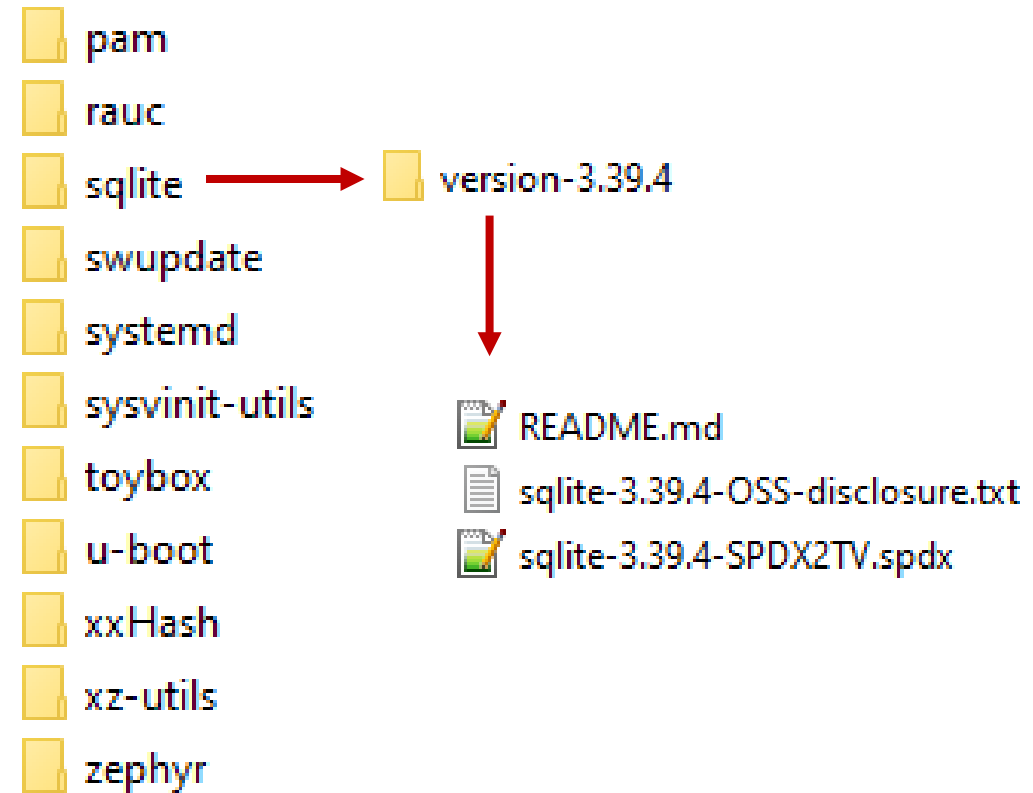
Picture by MIH83 / pixabay.com / Pixabay License

- Curation database – Why and How
- **Structure and provided artifacts**
- How we provide transparency
  - Why we use FOSSology
- The curation database in action – a showcase
- Potential usages
- Final remarks

# Structure and provided artifacts



- It is available as GitHub repo:  
<https://github.com/Open-Source-Compliance/package-analysis>
- License: CC0-1.0
- Simple and intuitive structure
- Per package version one directory, with
  - README
  - Packagename-version-OSS-disclosure.txt
  - Packagename-version-SPDX2TV.spdx



- The README.md file contains:
  - the URL where the package was downloaded from
  - the list of reviewers, who already had a look at the files and disclosed their name or GitHub id

```
## Download Location  
  
https://github.com/coreutils/coreutils/archive/refs/tags/v9.1.tar.gz  
  
## Reviewers  
  
The information was reviewed by:  
  
* Oliver Fendt
```

# Structure and provided artifacts



- The OSS-disclosure file is a ready to use file
- The OSS-disclosure file contains
  - all concluded licenses
  - an aggregated list of all identified (and added) copyright statements
  - all acknowledgments, which are required by the concluded licenses of the entire package
- It can be used to:
  - generate the OSS-disclosure document of a product
  - to decide whether the package is from a license point of view suited to be integrated in a product because it provides a complete overview about the license situation of the entire package

```
=====
openssl-OpenSSL_1_1_1s
-----
```

```
LICENSES
-----
```

```
OpenSSL
```

```
OpenSSL License
```

```
Copyright (c) 1998-2008 The OpenSSL Project. All rights reserved.
```

```
Redistribution and use in source and binary forms, with or without
modification, are permitted provided that the following conditions
are met:
```

```
1. Redistributions of source code must retain the above copyright
```

```
=====
ACKNOWLEDGEMENTS
-----
```

```
OpenSSL
```

```
This product includes software developed by the OpenSSL Project for
use in the OpenSSL Toolkit (http://www.openssl.org/)
```

```
This product includes cryptographic software written by Eric Young
(eay@cryptsoft.com).
```

```
This product includes software written by Tim Hudson (tjh@cryptsoft.com).
```

```
-----
Copyright notices
```

```
Copyright 1995-2021 The OpenSSL Project Authors. All Rights Reserved
```



# Structure and provided artifacts



- Similar to the OSS-disclosure file the SPDX tag-value file is ready to use
- The SPDX2TV file contains per file
  - meta information
  - all copyright statements
  - All concluded licenses
  - the explanation of “un-obvious” license conclusions to provide transparency
- They are human and machine readable, compliance officers can review the license conclusions per file
- They can be parsed by a machine and integrated in the build process in a way that only the licenses and copyright statements of those files are considered which will end up in the built artifact

```
PackageName: nghttp2-1.51.0.tar.gz
PackageFileName: nghttp2-1.51.0.tar.gz
SPDXID: SPDXRef-upload48
PackageDownloadLocation: NOASSERTION
PackageVerificationCode: 3726177083d753a05aac392f5803ee80a5eaa099
PackageChecksum: SHA1: 80838967c5646f81a01b4c0eabfb190111e6e8bd
PackageChecksum: SHA256: dfcf41e0b093765a79c9f1fc0ba6dc3d524555e9
PackageChecksum: MD5: 74d4d49e2c507cea8de79cc0f523e0b0
PackageLicenseConcluded: NOASSERTION
PackageLicenseDeclared: NOASSERTION
PackageLicenseComments: <text> licenseInfoInFile determined by
Scanners:
- nomos ("4.1.0.95".82b3b2)
- monk ("4.1.0.95".82b3b2)
- oio ("4.1.0.95".82b3b2) </text>
PackageLicenseInfoFromFiles: NOASSERTION
PackageCopyrightText: NOASSERTION
```

# Structure and provided artifacts



```
##File
FileName: nghttp2-1.51.0.tar.gz/nghttp2-1.51.0.tar/nghttp2-1.51.0/doc/_themes/sphinx_rtd_theme/search.html
SPDXID: SPDXRef-item192427
FileChecksum: SHA1: 9e05a0f08e1d2cd27bbe45c51fa8699e8f7df938
FileChecksum: SHA256: e66060ecbb8a70cdf1e7759bd46c86bcdf6bce720dd31633a9a144657c5697c2
FileChecksum: MD5: 6c8d4254a5b9822a5550ce1682bc94d9
LicenseConcluded: LicenseRef-BSD-2-Clause-08605c6785f75f9a31cf9d0d3df25bc2
LicenseComments: <text>The information in the file is:
:license: BSD, see https://github.com/sphinx-doc/sphinx/blob/master/LICENSE for details.

checking the given link revealed that the BSD-2-Clause license applies (which we copied as license text).
Thus we concluded BSD-2-Clause as license of this file
The information was retrieved on 12th of Nov 2022 </text>
LicenseInfoInFile: LicenseRef-BSD
FileCopyrightText: <text> copyright: Copyright 2007-2013 by the Sphinx team, see AUTHORS. </text>
```

- Curation database – Why and How
- Structure and provided artifacts
- **How we provide transparency**
  - Why we use FOSSology
- The curation database in action – a showcase
- Potential usages
- Final remarks

# Transparency is key



We provide as much transparency concerning the curations as possible, to achieve this:

- we provided a detailed description how curations are done in several scenarios, see <https://github.com/Open-Source-Compliance/package-analysis/README.md>
- Additionally we provide an explanation via LicenseComments when curations are not self explaining. These explanations usually contain:
  - The license information in the file
  - Description which steps have been undertaken to determine the “real” license
  - In case of an internet search the date when the information was retrieved

This information is available in the SPDX tag-value files

## License identification and conclusion

As already explained in the overall process description a licensing expert will review the scanner findings. During the review depending on the scanner matches, the corresponding text sections and the context in the files the following tasks are carried out:

- confirm scanner findings either file by file or via bulk statement
- correct scanner findings either file by file or via bulk statement

The following subsections provide more information about the tasks carried out:

### Correcting scanner findings

There might be cases where the scanner matches some license information in a file but this information is not the license of the file. For example

```
"DT binding documents should be licensed (GPL-2.0-only OR BSD-2-Clause)\n" . $herecurr && $fix) {$fixed[$fixlinenr] =~ s/SPDX-License-Identifier: .*/SPDX-License-Identifier: (GPL-2.0-only OR BSD-2-Clause)/;
```

- Curation database – Why and How
- Structure and provided artifacts
- How we provide transparency
  - **Why we use FOSSology**
- The curation database in action – a showcase
- Potential usages
- Final remarks

# Transparency is key

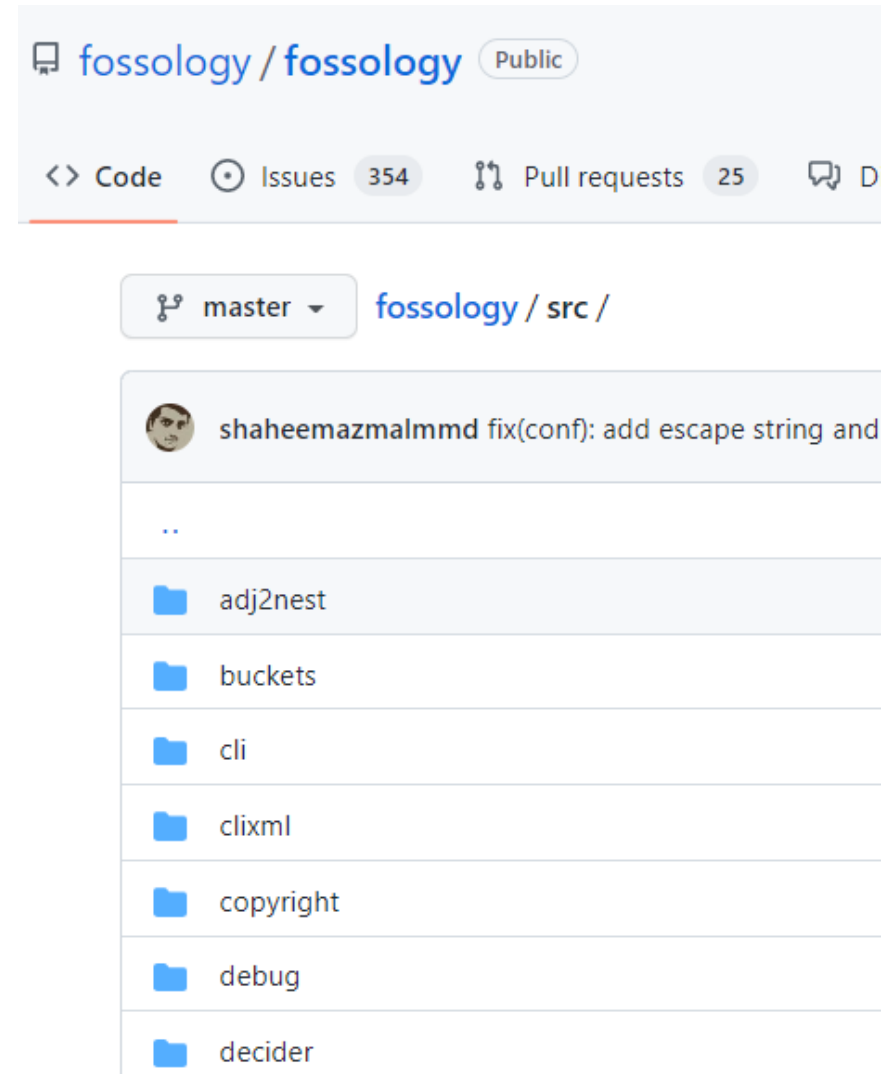
- The OSS package analysis and curation is done on file level and for every file contained in the package
- The analysis and curation is performed via the GPL-2.0 licensed tool FOSSology (<https://www.fossology.org/>)
- For license and copyright statement identification FOSSology provides different "agents" the user is able to select, which "agents" shall run, currently the following "agents" are available for license identification:
  - Nomos
  - Monk
  - Ojo
  - Scancode
- Each agent was build with a different main focus and we think that running them combined produces the best output. Which agents were run for a concrete package analysis is available in the SPDX2 tag-value file.



# Why do we use FOSSology



- FOSSology itself provides transparency since its source code is freely available
- It provides a GUI which allows a person to review the detected license information and in case it is necessary to correct it
- It provides functionality to detect deviations of license texts stored in the FOSSology database
- It offers the possibility to comment on license decisions
- It allows to add new found licenses to the FOSSology database
- It can export the license analysis in different formats
- It even allows to import existing license analysis files and apply the decisions available in the license analysis file



# Why do we use FOSSology



```
/*
 * Ceph - scalable distributed file system
 *
 * Copyright (C) 2004-2010 Sage Weil <sage@newdream.net>
 *
 * This is free software; you can redistribute it and/or
 * modify it under the terms of the GNU Lesser General Public
 * License version 2.1, as published by the Free Software
 * Foundation. See file COPYING.
 */

#ifndef CEPH_RBD_TYPES_H
#define CEPH_RBD_TYPES_H

#include <linux/types.h>

/* For format version 2, rbd image 'foo' consists of objects
 * rbd_id.foo - id of image
 * rbd_header.<id> - image metadata
 * rbd_object_map.<id> - optional image object map
 * rbd_data.<id>.0000000000000000
 * rbd_data.<id>.00000000000000001
 * ... - data
 * Clients do not access header data directly in rbd format 2.
 */

#define RBD_HEADER_PREFIX "rbd_header."
#define RBD_OBJECT_MAP_PREFIX "rbd_object_map."
#define RBD_ID_PREFIX "rhd id."
```

Go through all files with licenses and no clearing result

Clearing decision scope

Apply decision to all future occurrences of this file

Clearing decision type

- No license known
- To be discussed
- Irrelevant
- Identified
- Do not use
- Non functional

Action	License	Source	License Text	Acknowledgement	Comment
	LGPL-2.1	nomos: #1	Click to add	Click to add	Click to add

Showing 1 to 1 of 1 entries

User Decision ... Bulk Recognition ... Clearing History ...



# License identification and review

fossology Version: [unknown], Branch: [master], Commit: [#b04657] 2022

Folder: Software Repository/  
pax-utils-custom.zip/pax-utils-custom/pax-utils-1.3.3/BSD-License

Software Heritage | License Browser | File Browser • Conf |

Cleared: 0/41

Hide Legend

```
# Redistribution and use in source and binary forms, with or without
modification, are permitted provided that the following conditions are
met:

* Redistributions of source code must retain the above copyright
notice, this list of conditions and the following disclaimer.
* Redistributions in binary form must reproduce the above
copyright notice, this list of conditions and the following disclaimer
in the documentation and/or other materials provided with the
distribution.
* Neither the name of Google Inc. nor the names of its
contributors may be used to endorse or promote products derived from
this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS
"AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT
LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR
A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT
OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL,
SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT
LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE,
DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY
THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT
(INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE
OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
```

fossology.osuosl.org/repo/?mod=popup-license&rf=336

Software License  
Copyright (c) 2010, Google Inc. All rights reserved.





Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

Neither the name of Google nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Action	License	Source	License Text	Acknowledgment
 	WebM	monk: #1 (94 %)	Click to add	Click to add
 	BSD-3-Clause	nomos: #1	Click to add	Click to add

Showing 1 to 2 of 2 entries

# Curations

```
/*
 * Copyright (c) 2009, 2022 IBM Corp.
 *
 * All rights reserved. This program and the accompanying materials
 * are made available under the terms of the Eclipse Public License v2.0
 * and Eclipse Distribution License v1.0 which accompany this distribution.
 *
 * The Eclipse Public License is available at
 * https://www.eclipse.org/legal/epl-2.0/
 * and the Eclipse Distribution License is available at
 * http://www.eclipse.org/org/documents/edl-v10.php.
 *
 * Contributors:
 * Ian Craggs - initial API and implementation and/or initial documentation
 * Ian Craggs - add SSL support
 */
/**
 * @file
 * \brief functions which apply to client structures
 */

#include "Clients.h"

#include <string.h>
#include <stdio.h>

/**
 * List callback function for comparing clients by clientid
 * @param a first integer value
 * @param b second integer value
 */
```

**Legend:**  
license relevant text

Go through all files with licenses and no clearing result ⓘ

## Clearing decision scope

Apply decision to all future occurrences of this file ⓘ

## Clearing decision type

- No license known ⓘ
- To be discussed ⓘ
- Irrelevant ⓘ
- Identified ⓘ
- Do not use ⓘ
- Non functional ⓘ

Action ⓘ ▲	License ⓘ	Source ⓘ	License Text ⓘ	Acknowledgement ⓘ	Comment ⓘ
✗ ★	EPL-2.0	scancode: #1 (95 %) nomos: #1 Bulk: #173421	Click to add	Click to add	Click to add
✗ ★	EDL-1.0	nomos: #1 Bulk: #173419	Eclipse Distrib...	Click to add	Click to add
✗ ★	Dual-license	Bulk: #173425	This program an...	To the extend f...	The LICENSE fil..
✚ ★	BSD-3-Clause	scancode: #1 (95 %)	-	-	-

Showing 1 to 4 of 4 entries

# License comments

Acknowledgement ⓘ	Comment ⓘ
Click to add	Click to add
Click to add	Click to add
To the extend f...	The LICENSE fil...

Please enter the text phrase:

The LICENSE file in the root directory contains the following information:

Eclipse Public License - v 2.0

This program and the accompanying materials are made available under the terms of the Eclipse Public License v2.0 and Eclipse Distribution License v1.0 which accompany this distribution.

The Eclipse Public License is available at <https://www.eclipse.org/legal/epl-2.0/> and the Eclipse Distribution License is available at <http://www.eclipse.org/org/documents/edl-v10.php>.

For an explanation of what dual-licensing means to you, see: <https://www.eclipse.org/legal/eplfaq.php#DUALLIC>

Visiting the link showed the information below

<https://www.eclipse.org/legal/eplfaq.php#DUALLIC>  
For Eclipse projects which are dual-licensed, your file headers state that the code is being made available under two licenses. For example: "This program and the accompanying materials are made available under the terms of the Eclipse Public License v1.0 and Eclipse Distribution License v. 1.0 which accompanies this distribution." What is meant by the use of the conjunction "and"?  
The code is being made available under both of the licenses. The consumer of the code can select which license terms they wish to use, modify and/or further distribute the code under.

Thus the user is free to choose the License, it is a classical dual license case.

The information was retrieved on 1st of Dec 2022

## SPDX-Tag-Value file

```
##File
FileName: paho.mqtt.c-1.3.11.tar.gz/paho.mqtt.c-1.3.11.tar/paho.mqtt.c-1.
SPDXID: SPDXRef-item177361
FileChecksum: SHA1: 9b4eb0be9fa09135d21ca4e702e2d2bd26a0e939
FileChecksum: SHA256: d717a76c97c9e3dfe0f4488f2c2caab76af0f686dc1f1d5d5a8
FileChecksum: MD5: 0e18e7eb1341ad376d5b0fe69884c3d5
LicenseConcluded: EPL-2.0 OR LicenseRef-EDL-1.0-2e85a4c8a3142ff9a5a7f9332
LicenseComments: <text>...The LICENSE file in the root directory contains
Eclipse Public License - v 2.0
This program and the accompanying materials
are made available under the terms of the Eclipse Public License v2.0
and Eclipse Distribution License v1.0 which accompany this distribution.
The Eclipse Public License is available at
https://www.eclipse.org/legal/epl-2.0/
and the Eclipse Distribution License is available at
http://www.eclipse.org/org/documents/edl-v10.php.
For an explanation of what dual-licensing means to you, see:
https://www.eclipse.org/legal/eplfaq.php#DUALLIC
Visiting the link showed the information below
```

- Curation database – Why and How
- Structure and provided artifacts
- How we provide transparency
  - Why we use FOSSology
- **The curation database in action – a showcase**
- Potential usages
- Final remarks

- The product “Killer App” integrates some OSS packages
- Killer App shall be distributed
- The development team maintains the SBOM of Killer App
- A company wide template for products is provided by the companies’ legal department
- The product owner and the development team want to integrate the production of the OSS compliance artifacts for Killer App in the build pipeline

## Killer App SBOM

```
rauc 1.8  
nghttp2 1.50.0  
gnutls 3.7.8  
sqlite 3.39.4
```

# Description of the automated procedure

- Download the zipped repo content
- Unzip the content
- Read list of used packages and search for package name and version
- Copy found OSS-disclosure documents to a temp directory
- Fetch the OSS-disclosure template – add product/delivery specific data
- Combine:
  - Product specific OSS-disclosure template
  - List of used packages
  - OSS-disclosure documents of all used packages
- Store the generated OSS-disclosure document
- (Clean up)

## LICENSE INFORMATION – FREE AND OPEN SOURCE SOFTWARE

### 1. General Information

This product contains third party Free and Open Source Software which is provided under a number of different licenses (hereinafter referred to as „FOSS“). The respective license texts are listed below, and you can obtain rights and licenses directly from the right holders to the extent specified therein. The FOSS licenses prevail all other license conditions and contractual agreements with [COMPANY] with regard to the corresponding FOSS software components contained in the product.

### 2. Source Code Offer

This product contains software components that are licensed by the copyright holders as Free Software or Open Source software under the GNU General Public License, version 2 and/or 3, and/or GNU Lesser General Public License, version 2.1 and/or 3.0. Anyone can obtain the source code for these software components from us on a data carrier (CD-ROM, DVD or USB memory stick). This offer is valid within three years after the most recent conveyance of the object code by us, and valid for as long as we offer spare parts or customer support for the respective product. Please send your request to the following email address [email address] or via regular mail to the following address:

[Company name]

[Department]

[Address]

The curation database in action



**It's showtime**



- Curation database – Why and How
- Structure and provided artifacts
- How we provide transparency
  - Why we use FOSSology
- The curation database in action – a showcase
- **Potential usages**
- **Final remarks**



# Potential usage of the curation database

- compare the provided information against the results of the inhouse established process
- use it as base for the decision whether a package is suited to be integrated in a product from a license point of view
- generate several OSS-disclosure documents of a product using
  - SPDX2TV
  - OSS-disclosure
- use it as reference for inbound OSS
- use it to generate the source code bundle
- use it as base for other package versions
- and many more



Picture by Geralt / pixabay.com / Pixabay License

# Finally

- Evaluate it, use it, improve it, share it
- We welcome contributions:
  - Package analysis
  - Tools, e.g. for automation
  - Bug fixes
  - Bug reports
  - Other material required in the OSS compliance processes
- If you like it – “star” the repo



Picture by MetsikGarden / pixabay.com / Pixabay License

**Let's reduce license compliance efforts to the minimum**

# Thank you for listening

Please contact [oliver.fendt@fossea.de](mailto:oliver.fendt@fossea.de)  
if you need any support