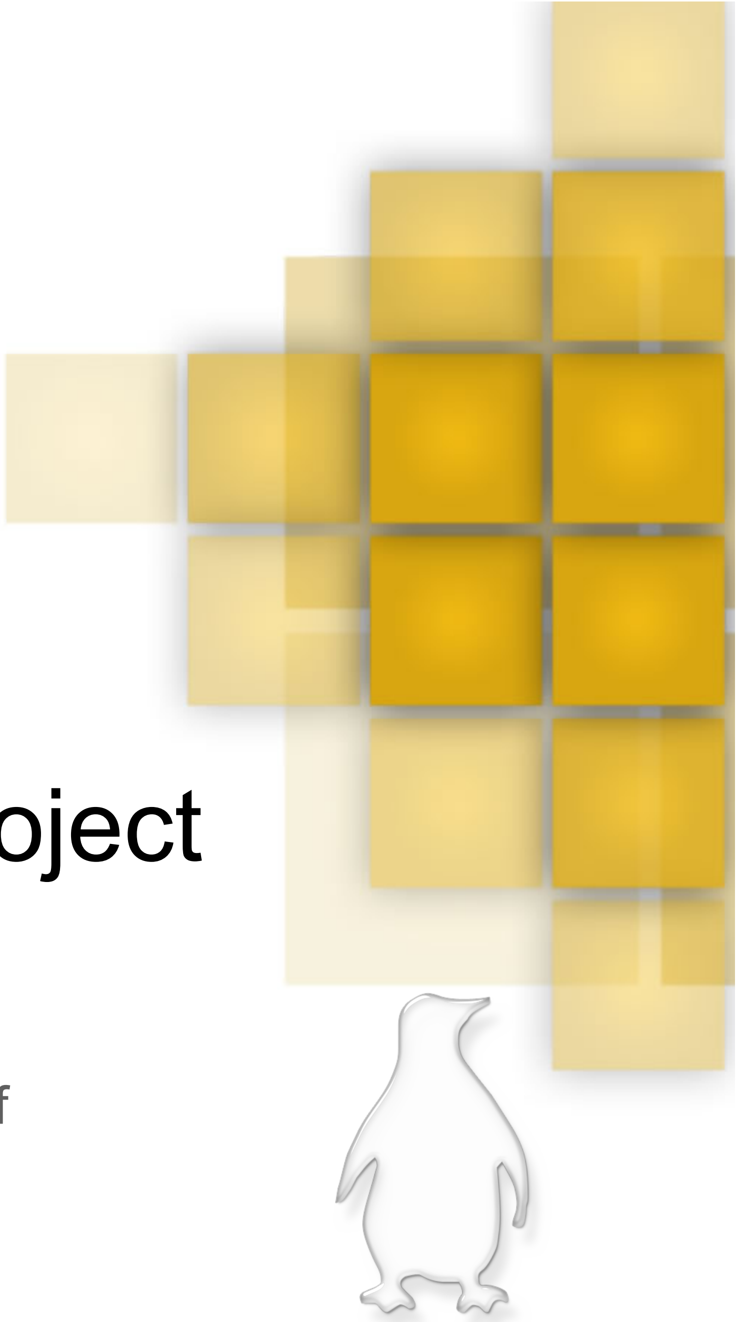
SPDX and the Yocto Project

Embedded Linux Conference – Europe
Edinburgh International Conference Centre
Oct 25, 2013

Mark Hatle, Senior Member of Technical Staff
Wind River



DISCLAIMER:

All opinions in this presentation are strictly those of its authors and do not represent the opinion, policy or position of any organization with which the authors are currently or has previously been associated.

This presentation is for educational purposes only. Consult your legal counsel for advise or guidance with regard to your specific situation.

Linux Foundation Projects

- The Yocto Project
  - [yoc•to] – smallest unit of measurement equal to 10^{-24}
  - A development environment to custom build a Linux distribution
  - Hardware independent

- The Software Package Data Exchange (SPDX)
  - [SPDX] – Software Package Data Exchange
  - A specification for communicating the components, licenses and copyrights associated with a software package
  - The PDF of license information sharing

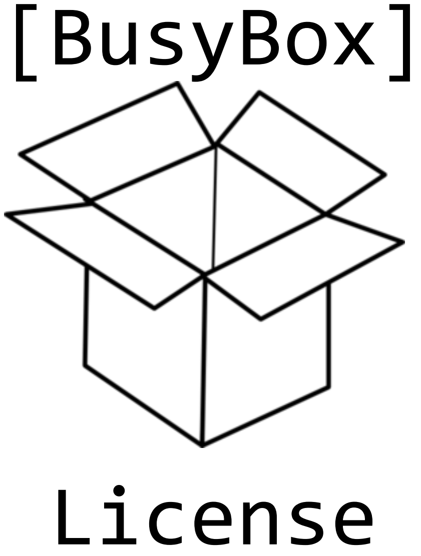
Yocto Project, Software Package Data Exchange and SPDX are registered trademarks of the Linux Foundation

✓ The Problem

The Problem

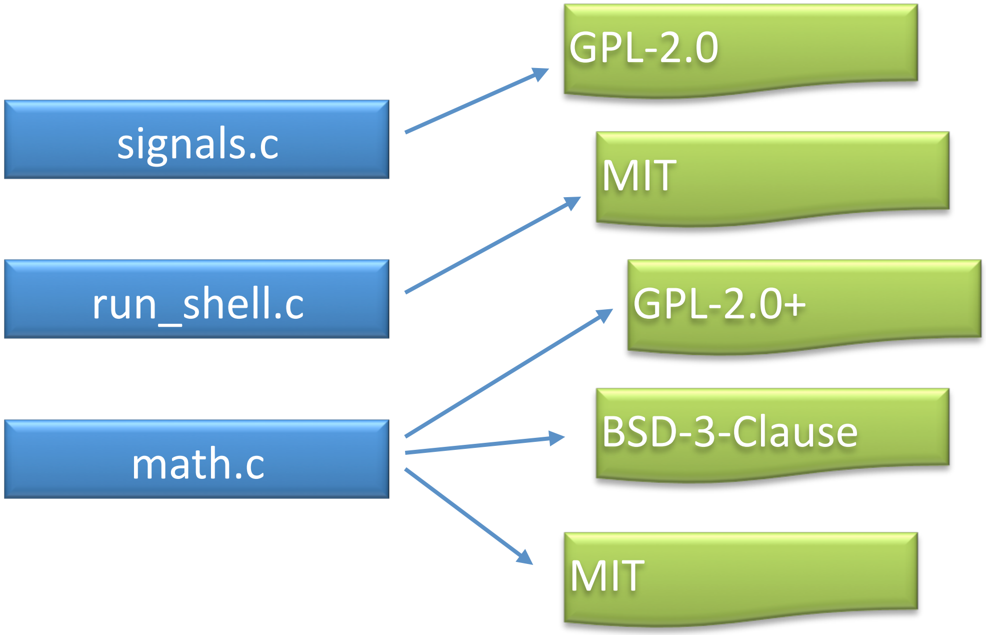
- What is the software license(s) of the software that makes up your product?
  - Open Source
  - Proprietary
  - Homegrown
- What are your obligations?

An Example: BusyBox

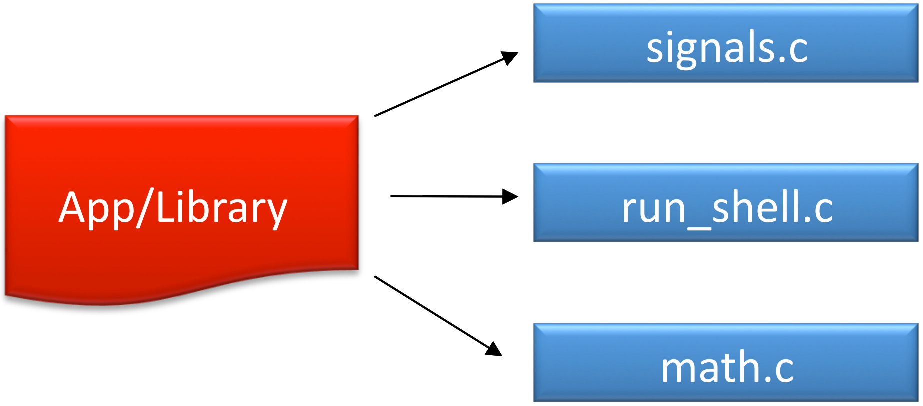


- GPL-2.0?
- Busybox is particularly interesting because it consists of many files from other projects (under different licenses)

File/License Relation



File/License Relation



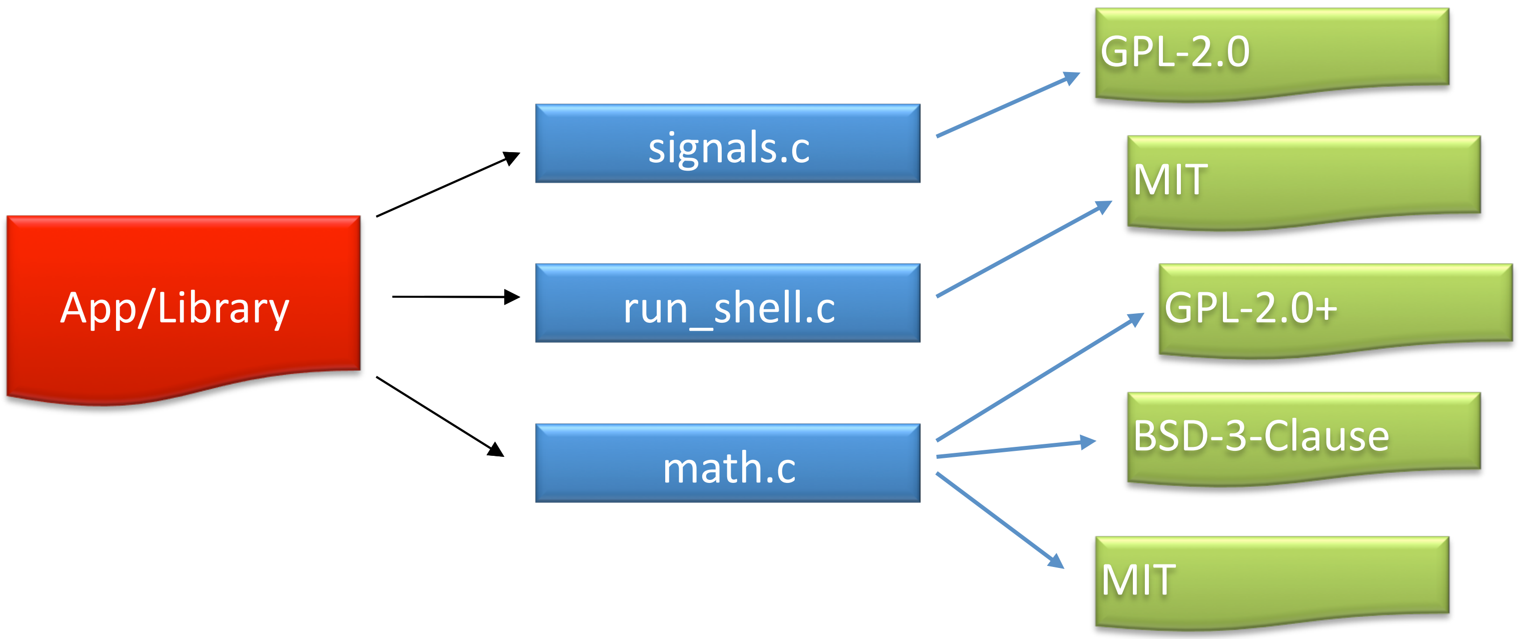
App/Library

signals.c

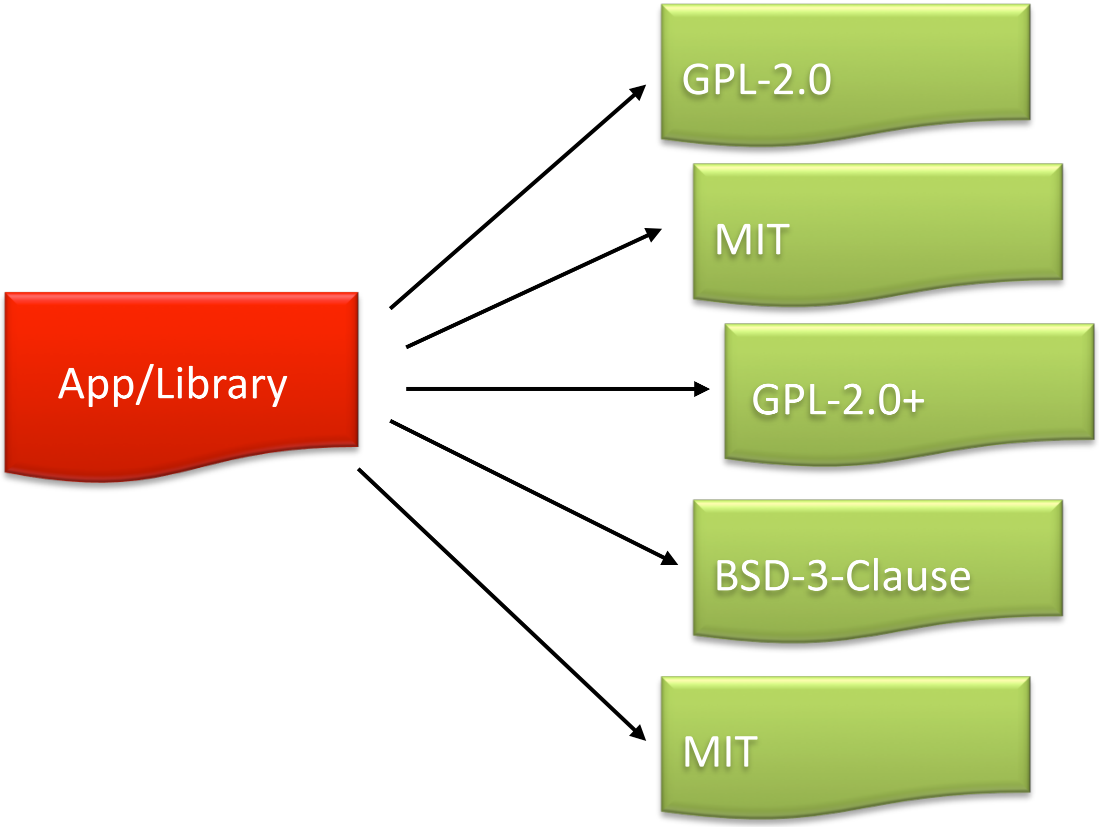
run_shell.c

math.c

File/License Relation

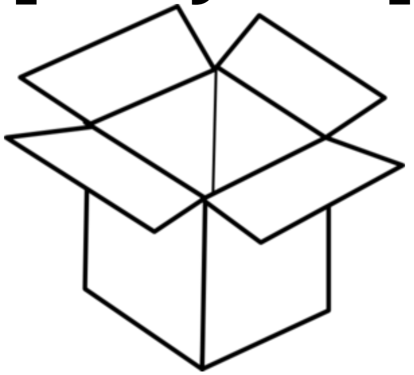


File/License Relation



An Example: BusyBox

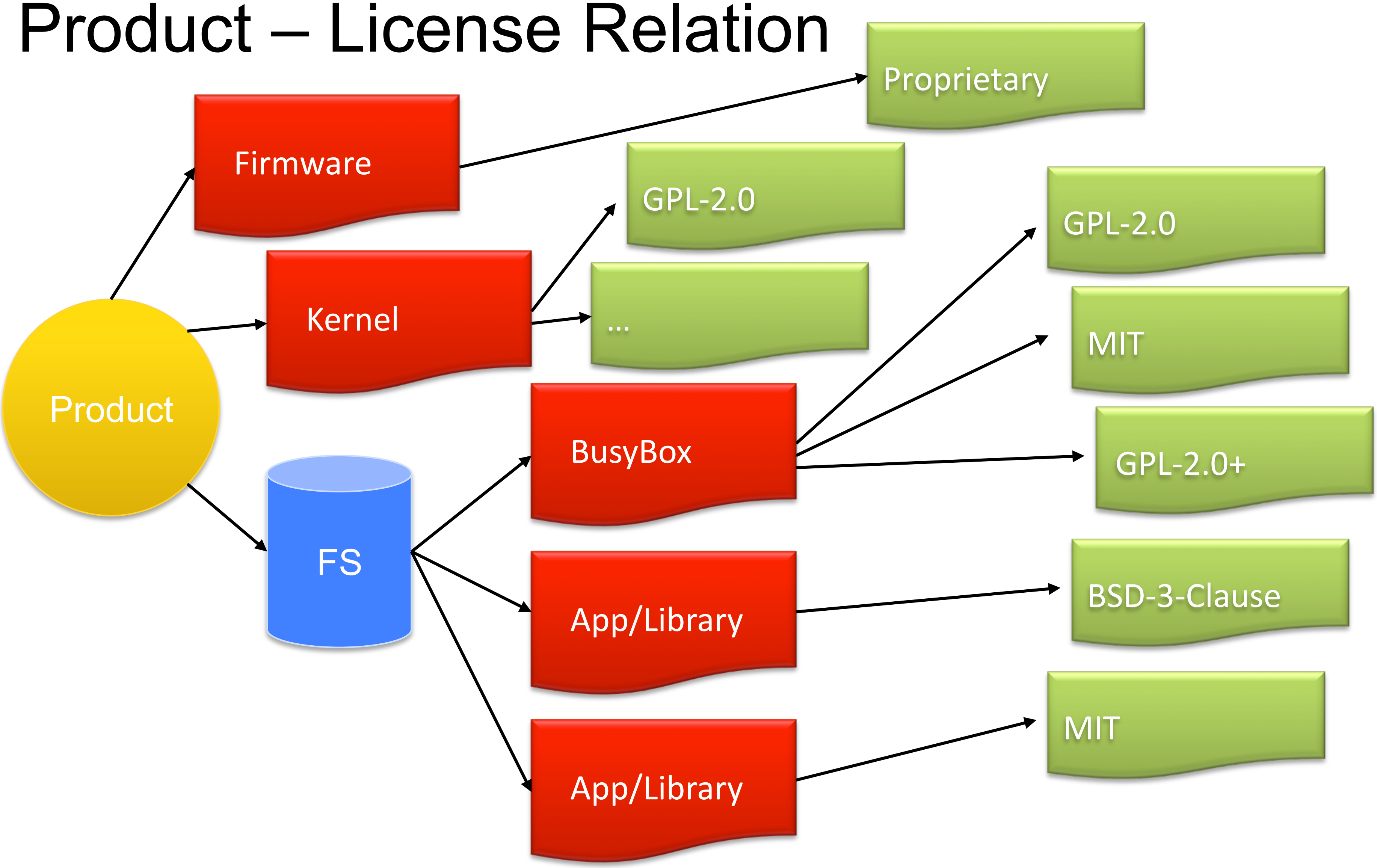
[BusyBox]



License

- Info from files: (GPL-1.0+$_{58}$ and GPL-2.0$_{136}$ and GPL-2.0+$_{281}$ and LGPL-2.1+$_{19}$ and BSD-3-Clause$_{20}$ and MIT and LicenseRef-1$_{6}$ and LicenseRef-2 and LicenseRef-3 and LicenseRef-4 and LicenseRef-5 and LicenseRef-5 and PublicDomain$_{6}$ and NOASSERTION$_{27}$ and NONE$_{26}$)
- Declared by maintainer: GPL-2.0
- Concluded by reviewer: GPL-2.0

Product – License Relation



✓ SPDX Overview

SPDX File Contents

- Specification Information: Version, License, Comment
- Creation Information: Creator, Date, Comment
- Package Information: Name, Version, URL, Summary, FileName, Supplier, Checksum, Description, …
- File Information: For each file – Type, License, Checksum, License, CopyrightText, FileName
- License Information: For each license – Text, Name, ID, CrossReference

Package Information

```
PackageName: zlib
PackageVersion: 1.2.8
PackageDownloadLocation: http://www.zlib.net/zlib-1.2.8.tar.xz
PackageSummary: <text>NOASSERTION</text>
PackageFileName: zlib.tar.gz
PackageSupplier: Person:NOASSERTION
PackageOriginator: Person:NOASSERTION
PackageChecksum: SHA1: 11f624a495a33fbbdba7b28028f06fcf6b5ba3d8
PackageVerificationCode: da39a3ee5e6b4b0d3255bfef95601890afd80709
PackageDescription: <text>zlib version 1.2.8</text>
PackageCopyrightText: <text>NOASSERTION</text>
PackageLicenseDeclared: (LicenseRef-0 AND LicenseRef-1)
PackageLicenseConcluded: NOASSERTION
PackageLicenseInfoFromFiles: LicenseRef-0
PackageLicenseInfoFromFiles: LicenseRef-1
```

File Information

```
FileType: SOURCE
LicenseInfoInFile: LicenseRef-0
FileChecksum: SHA1: 55d01b1ae60a09743e786523cbf1cd65e8577f50
LicenseConcluded: NOASSERTION
FileCopyrightText: <text>copyright (c) 1995-2013 mark adler
 * for conditions of distribution and use, see copyright notice in
   zlib.h
copyright[] =
   " inflate9 1.2.8 copyright 1995-2013 mark adler ";</text>
FileName: contrib/infback9/inftree9.c
```

License Information

```
ExtractedText: <text>Please see online publication for the full
   text of this license</text>

LicenseName: Zlib-possibility

LicenseID: LicenseRef-0

LicenseCrossReference:


ExtractedText: <text>GPL is referenced without a version number.
   Please look up GPL in the License Admin to view the different
   versions.</text>

LicenseName: GPL

LicenseID: LicenseRef-1

LicenseCrossReference:
```

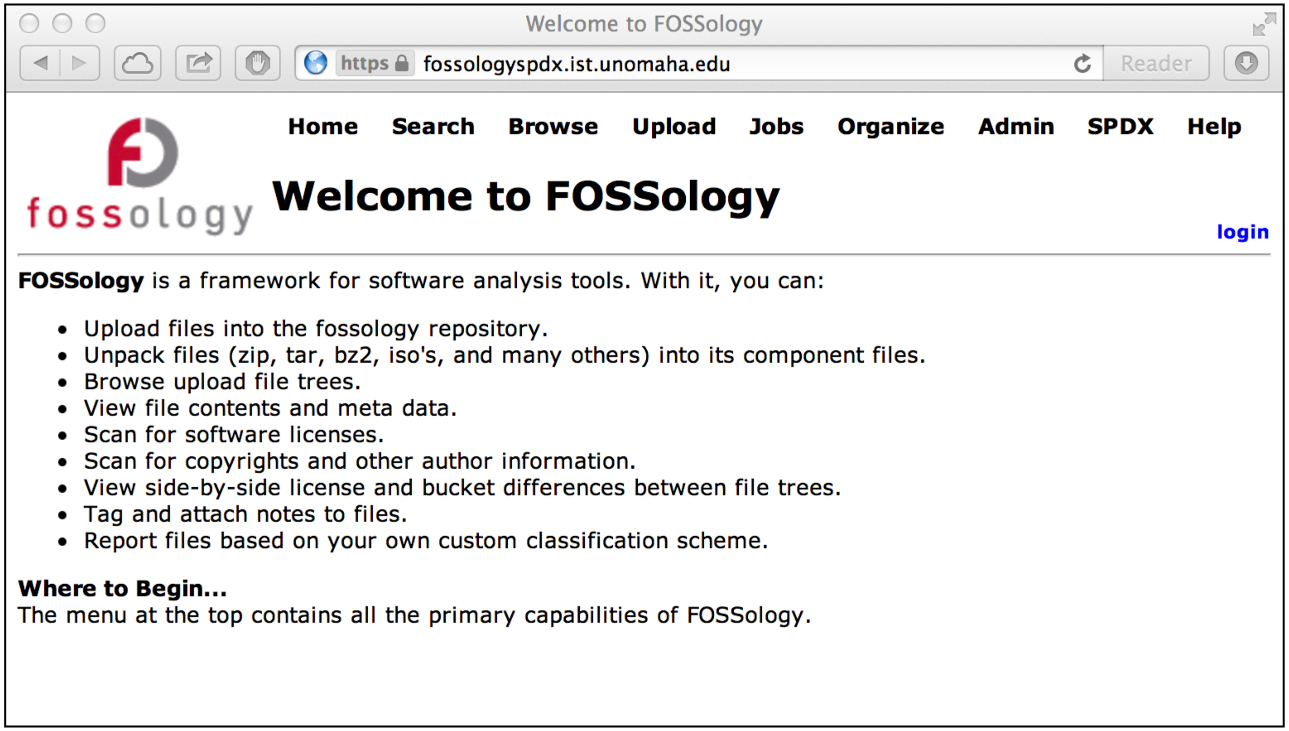
✓ Generating SPDX

SPDX Data Creation

- Good – Machine Generated, license info generated only using computer automation.
- Better – Human Reviewed, license info generated using computer automation, enhanced by human review.
- Best – Human Created/Interpreted, license info generated by human intelligence: code review, inspection, and interpretation.

Machine Generated

- FOSSology – Free data analysis tools. Predominant use of FOSSology is scanning files for licenses and copyrights.



Welcome to FOSSology

FOSSology is a framework for software analysis tools. With it, you can:

- Upload files into the fossology repository.
- Unpack files (zip, tar, bz2, iso's, and many others) into its component files.
- Browse upload file trees.
- View file contents and meta data.
- Scan for software licenses.
- Scan for copyrights and other author information.
- View side-by-side license and bucket differences between file trees.
- Tag and attach notes to files.
- Report files based on your own custom classification scheme.

Where to Begin...
The menu at the top contains all the primary capabilities of FOSSology.

FOSSologySPDX

- Real-time license scanning for packages using FOSSOlogy agent to return file level information like sha1, license, copyright, etc. (SPDX file level spec)
- Project of UNIVERSITY OF Nebraska Omaha
- License scanning result output in
  - JSON format
  - Plain text format

More information could be found here:

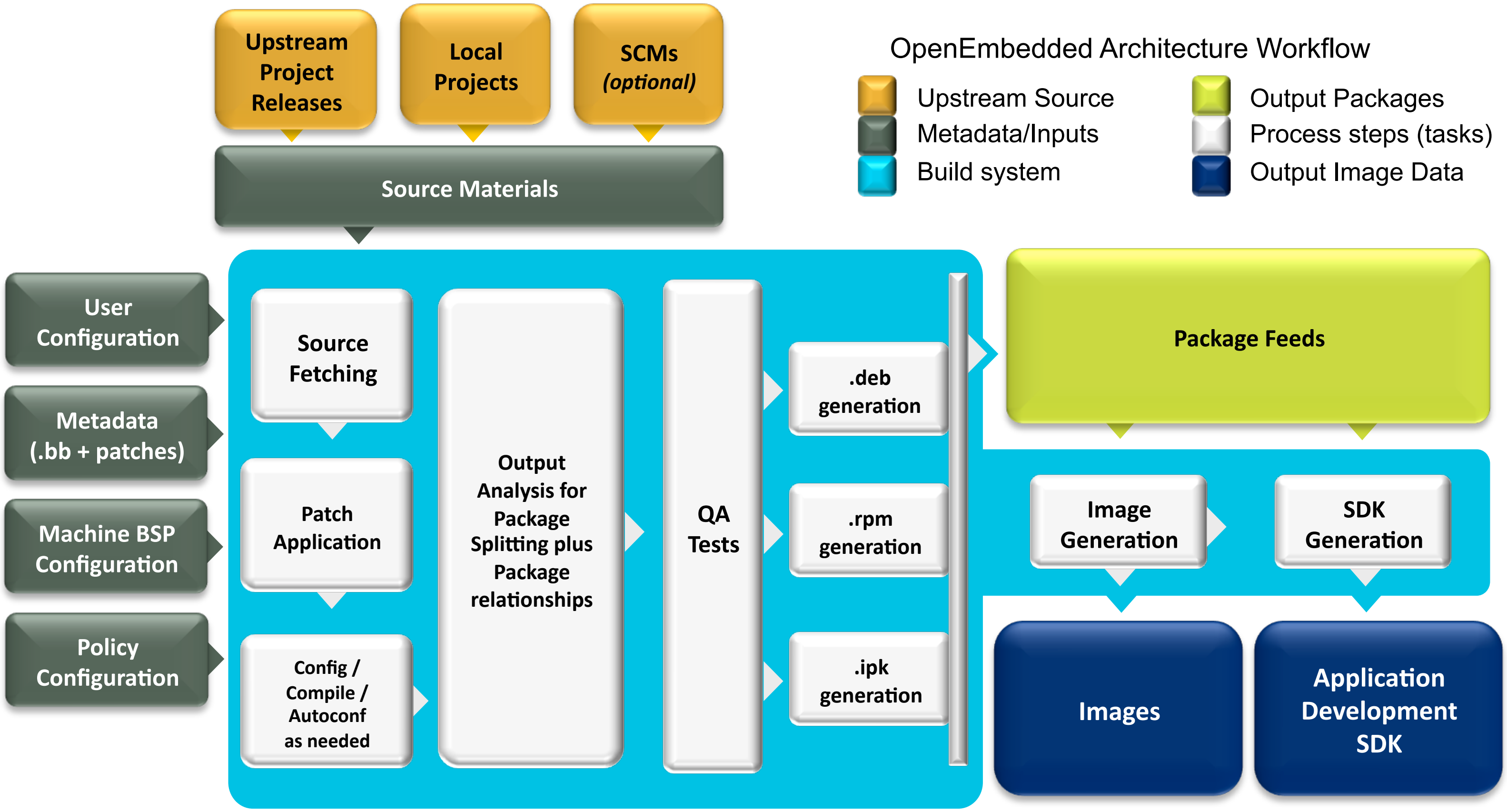
https://github.com/spdx-tools/fossology-spdx/wiki/Fossology-spdx-web-api

JSON format output

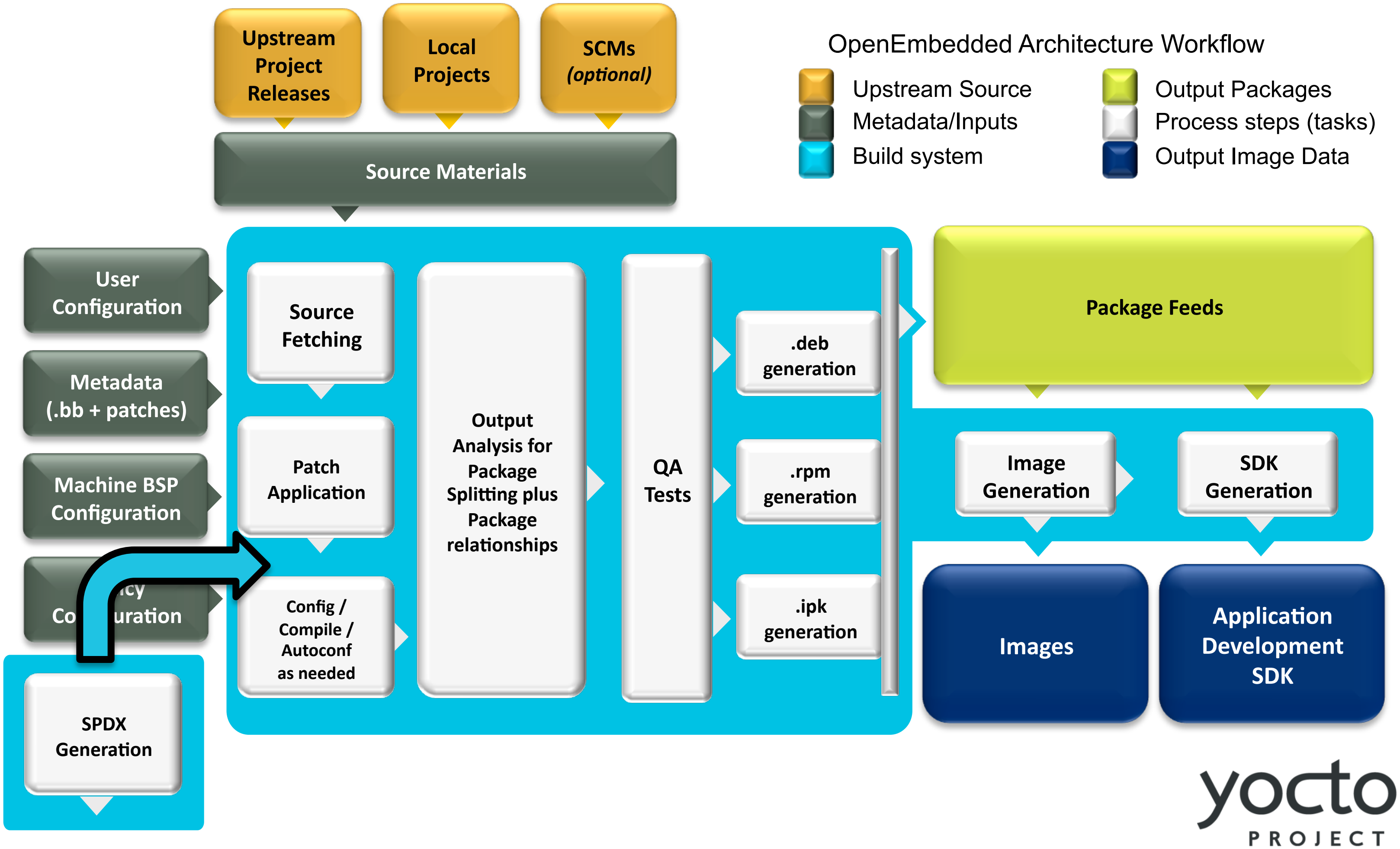
```
{
   "file_level_info":[
      {
         "FileName":"stamp-vti",
         "FileType":"SOURCE",
         "FileChecksum":"8e5113f6f47ce34e0437c2105441dbb70f01491a",
         "FileChecksumAlgorithm":"SHA1",
         "LicenseConcluded":"NOASSERTION",
         "LicenseInfoInFile":"No_license_found",
         "FileCopyrightText":"<text>NOASSERTION<\/text>"
      },
...
   ],
   "extracted_license_info":[
      {
         "LicenseName":"FSF",
         "ExtractedText":"<text>Copyright (C) 2003, 2006-2007 Free Software Foundation, Inc.\r
\nThis file is free software; the Free Software Foundation\r\ngives unlimited permission to copy
and\/or distribute it,\r\nwith or without modifications, as long as this notice is preserved.<\/
text>",
         "LicenseCrossReference":""
      },
```

✓ Yocto Project

Build System Workflow



Build System Workflow



Upstream Project Releases

Local Projects

SCMs *(optional)*

Source Materials

User Configuration

Metadata (.bb + patches)

Machine BSP Configuration

cy

Co uration

SPDX Generation

Source Fetching

Patch Application

Config / Compile / Autoconf as needed

Output Analysis for Package Splitting plus Package relationships

QA Tests

.deb generation

.rpm generation

.ipk generation

Package Feeds

Image Generation

SDK Generation

Images

Application Development SDK

OpenEmbedded Architecture Workflow

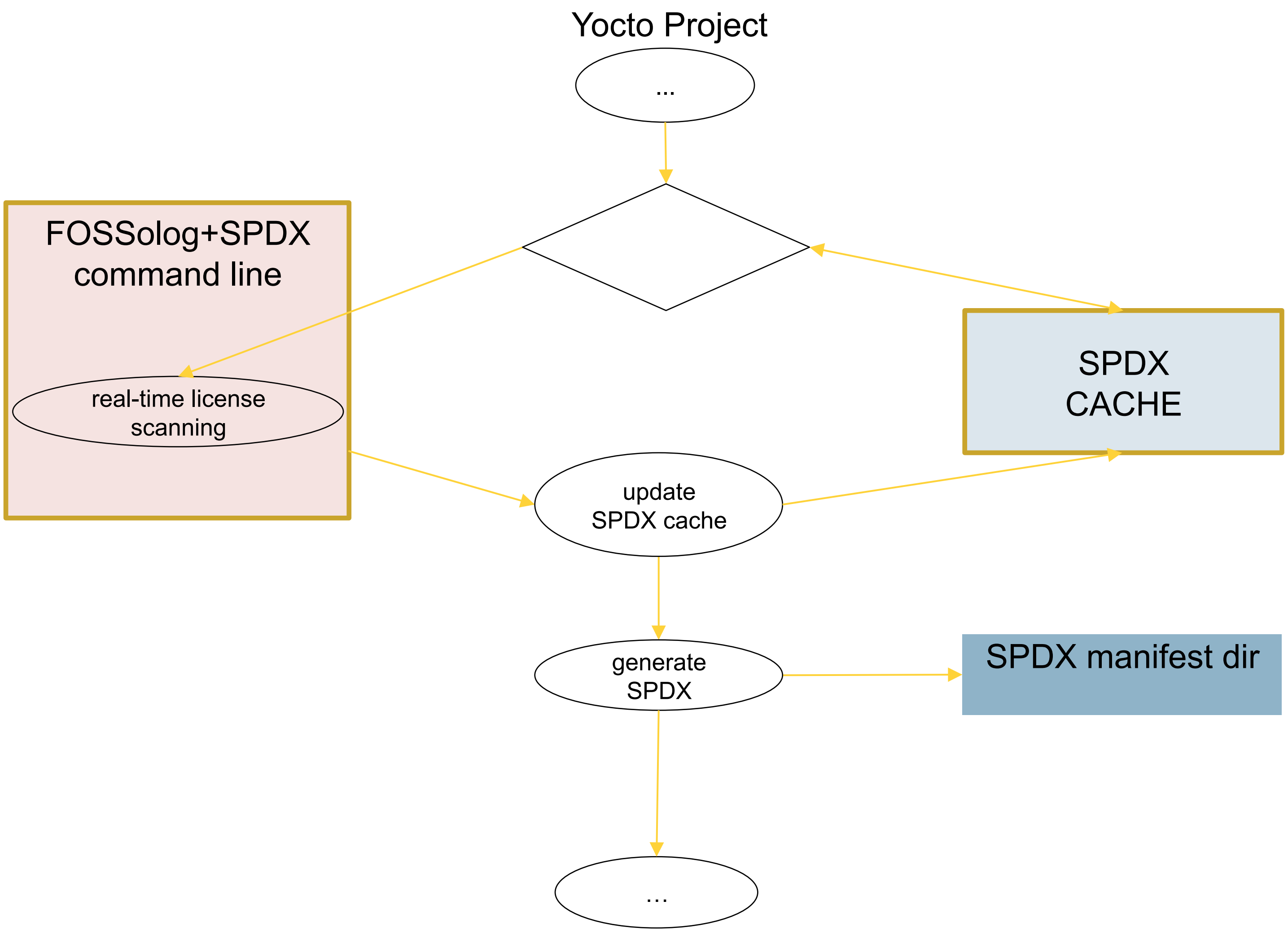
🟧	Upstream Source	🟩	Output Packages
⬛	Metadata/Inputs	⬜	Process steps (tasks)
🟦	Build system	🟦	Output Image Data

do_spdx

A task plugged in to the Yocto Project build to generate SPDX

- clean up old log files
- create SPDX temp work folder
- get SPDX information from cache, if it exists
- if the SPDX cache does not exist; tar sources and send them to be scanned
- get file level information and extracted license information from FOSSologySPDX Real time Scanning
- generate SPDX file to SPDX manifest directory
- write SPDX (JSON) to cache
- clean up the temp work dir

do_spdx


Yocto Project
...
FOSSolog+SPDX command line
real-time license scanning
SPDX CACHE
update SPDX cache
generate SPDX
SPDX manifest dir
…

SPDX Manifest

```
SPDXVersion: SPDX-1.1
DataLicense: CC0-1.0
DocumentComment: <text>SPDX for acl version 2.2.51</text>

## Creation Information
Creator: fossology-spdx
Created: 2013-10-14T17:53:00
CreatorComment: <text>UNO</text>

## Package Information
PackageName: acl
PackageVersion: 2.2.51
PackageDownloadLocation: http://download.savannah.gnu.org/releases/acl/acl-2.2.51.src.tar.gz
PackageSummary: <text>NOASSERTION</text>
...
ackageDescription: <text>acl version 2.2.51</text>
PackageCopyrightText: <text>NOASSERTION</text>

PackageLicenseDeclared: (GPL-2.0 AND LGPL-2.0 AND LGPL-2.1 AND LicenseRef-0 AND LicenseRef-1 AND LicenseRef-10 AND
LicenseRef-2 AND LicenseRef-3 AND LicenseRef-4 AND LicenseRef-5 AND LicenseRef-6 AND LicenseRef-7 AND ...
PackageLicenseConcluded: NOASSERTION

## File Information
FileType: SOURCE
LicenseInfoInFile: LicenseRef-0
FileChecksum: SHA1: 9784178d4cbea71e6a5876253dec85f36356b9e6
LicenseConcluded: NOASSERTION
FileCopyrightText: <text>copyright (c) 1999, 2000
  andreas gruenbacher, <a.gruenbacher@bestbits.at>
a.gruenbacher@bestbits.at</text>
FileName: libacl/acl_get_fd.c
```

do_spdx

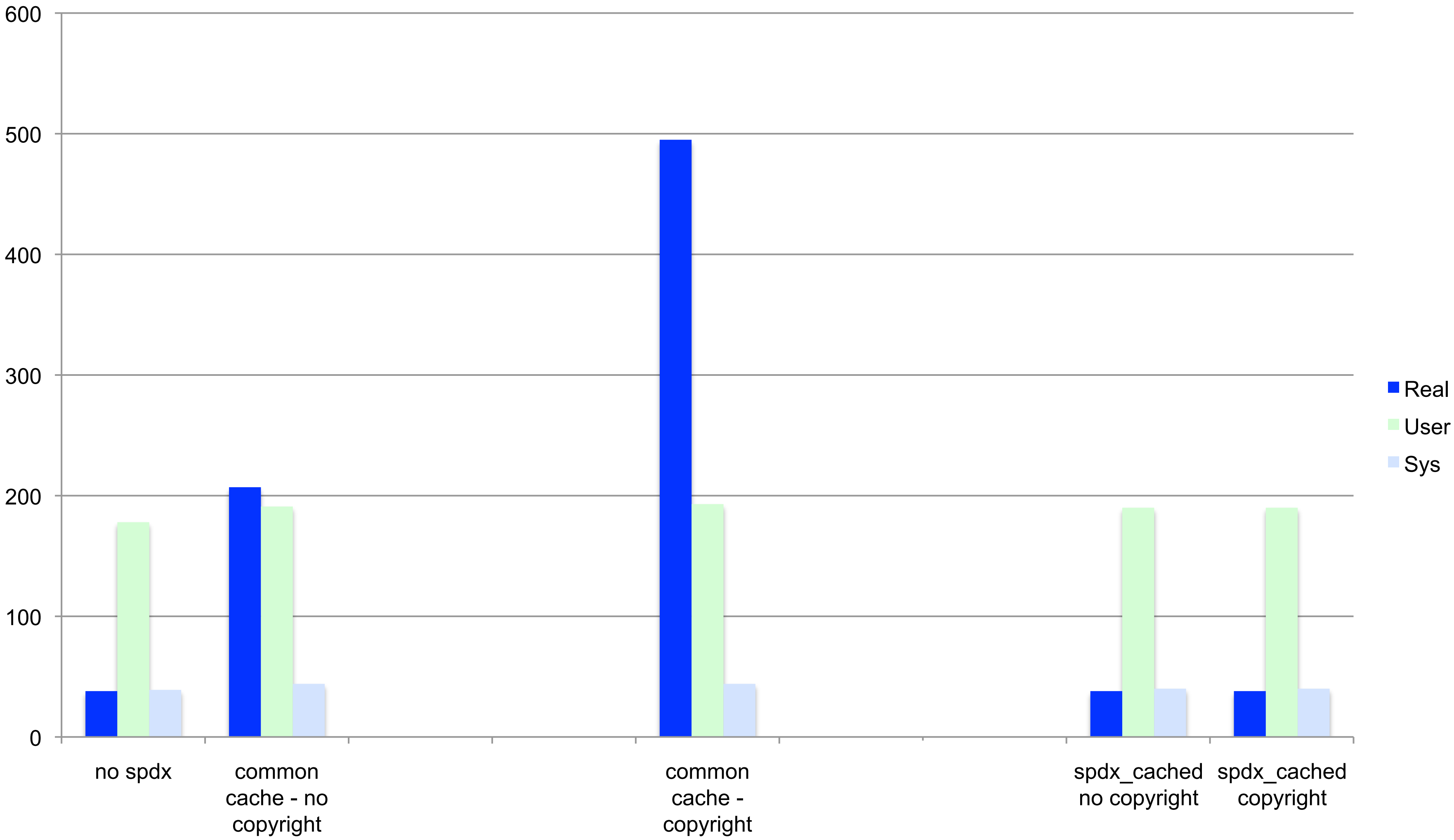
- In the Yocto Project 1.5!
  - Prototype/Beta quality
- To enable:
  - Setup a FOSSology server add the FOSSologySPDX module
    - Be sure to set apache/postgres/php timeout, memory sizes and other configurations according to docs
  - add “spdx” to the USER_CLASSES in your local.conf
  - Configure SPDX class, look at meta/conf/license.conf for details

Performance

- Test Configuration:
  - FOSSology Machine
    - Intel® Xeon® X5450 @ 3.00GHz (8-Cores)
    - 48 GB of RAM
    - 6 TB SAS RAID
    - CentOS 6.4
  - Build Machine
    - Intel® Xeon® X5560 @ 2.80GHz (8-Cores / 16-Threads)
    - 12 GB of RAM
    - 2 TB SATA RAID
    - Fedora 13
    - Building ‘core-image-minimal’

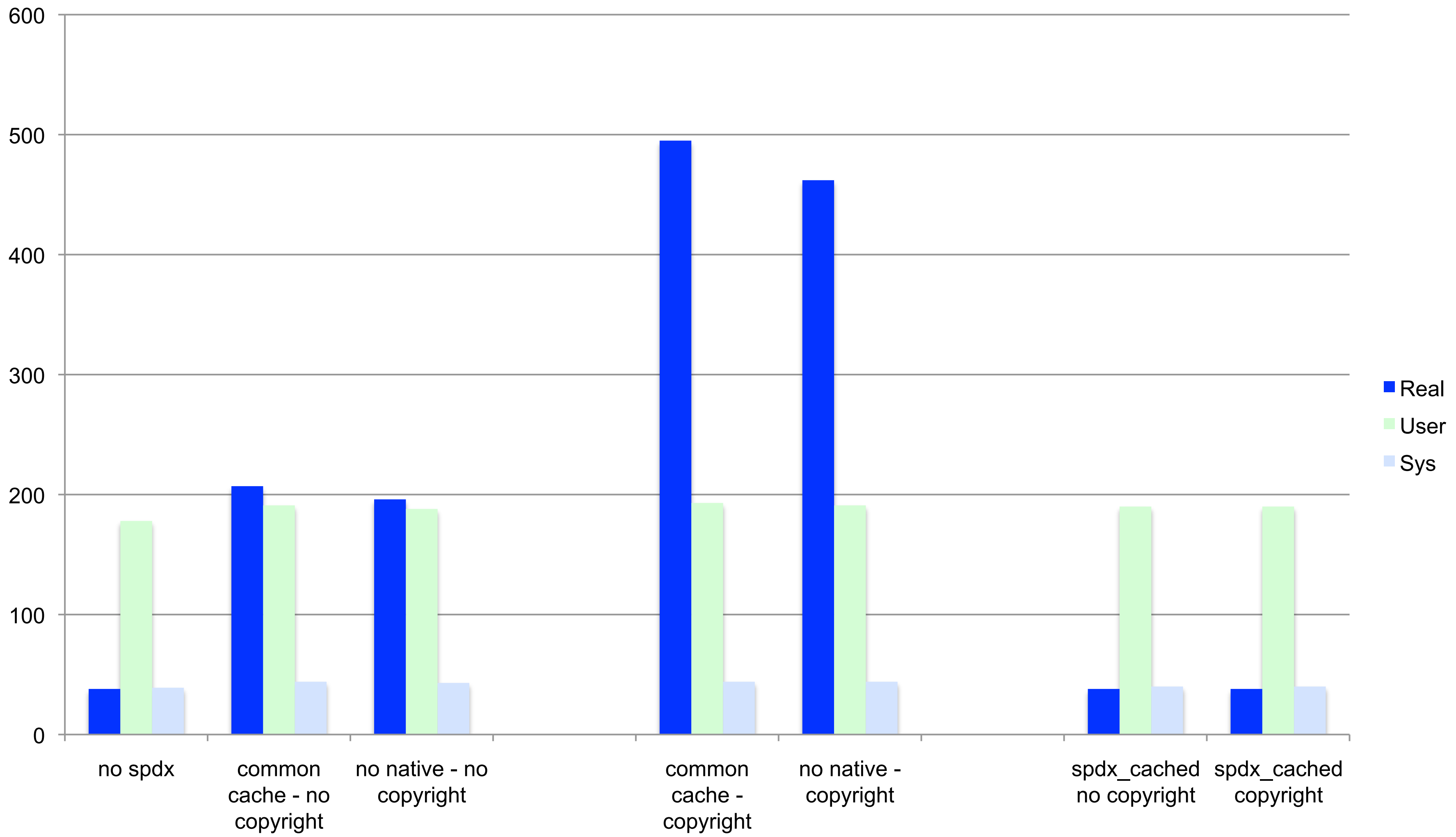
Performance

- Changes from YP 1.5
  - Additional license information processing
  - JSON output vs tagged
  - General fixes
  - Use ‘BPN’ for caching – avoid multilib issues, encourages internal reuse of results
  - Lock the spdx cache file access

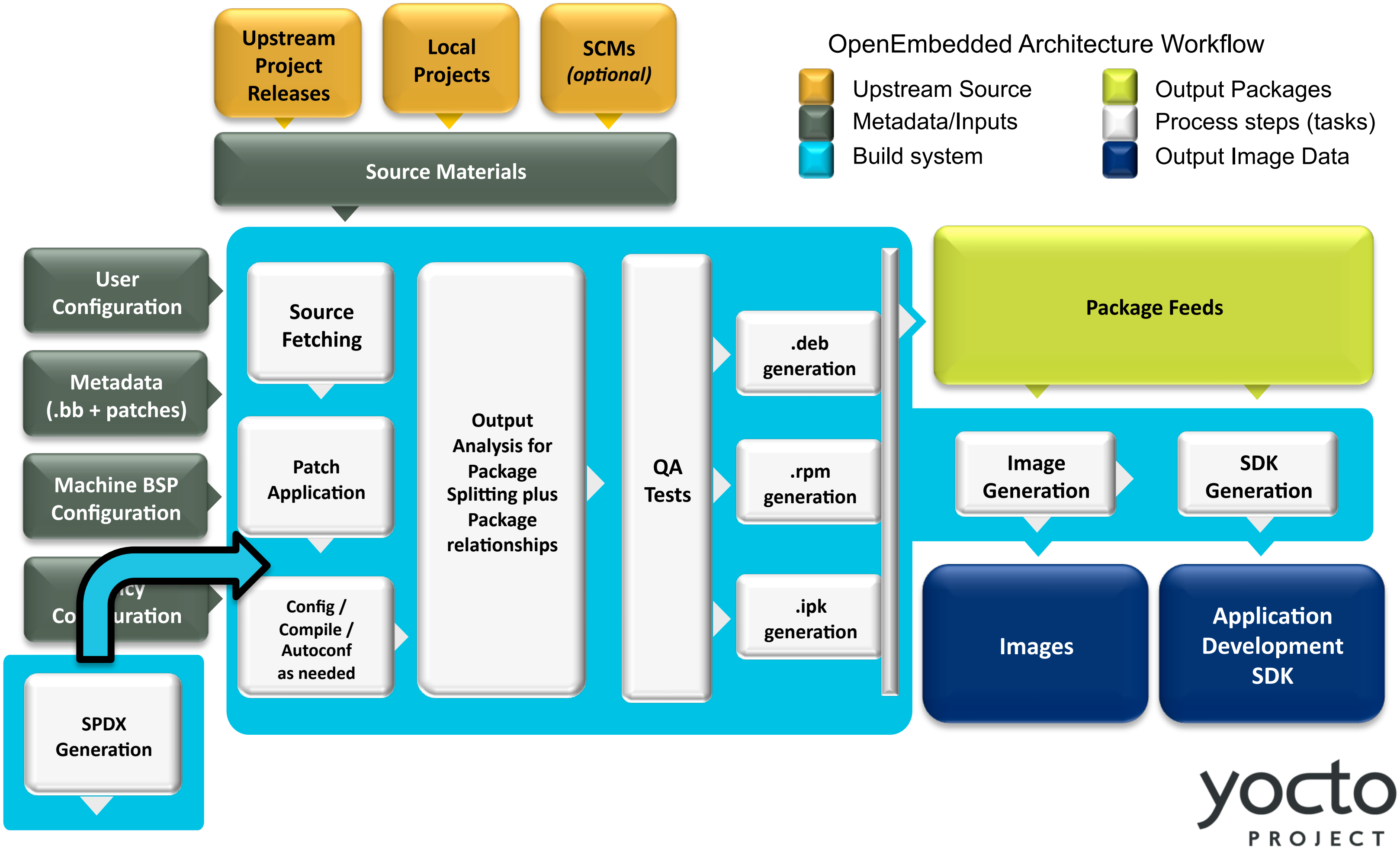

600
500
400
300
200
100
0
no spdx
common cache - no copyright
common cache - copyright
spdx_cached no copyright
spdx_cached copyright
Real
User
Sys

Performance

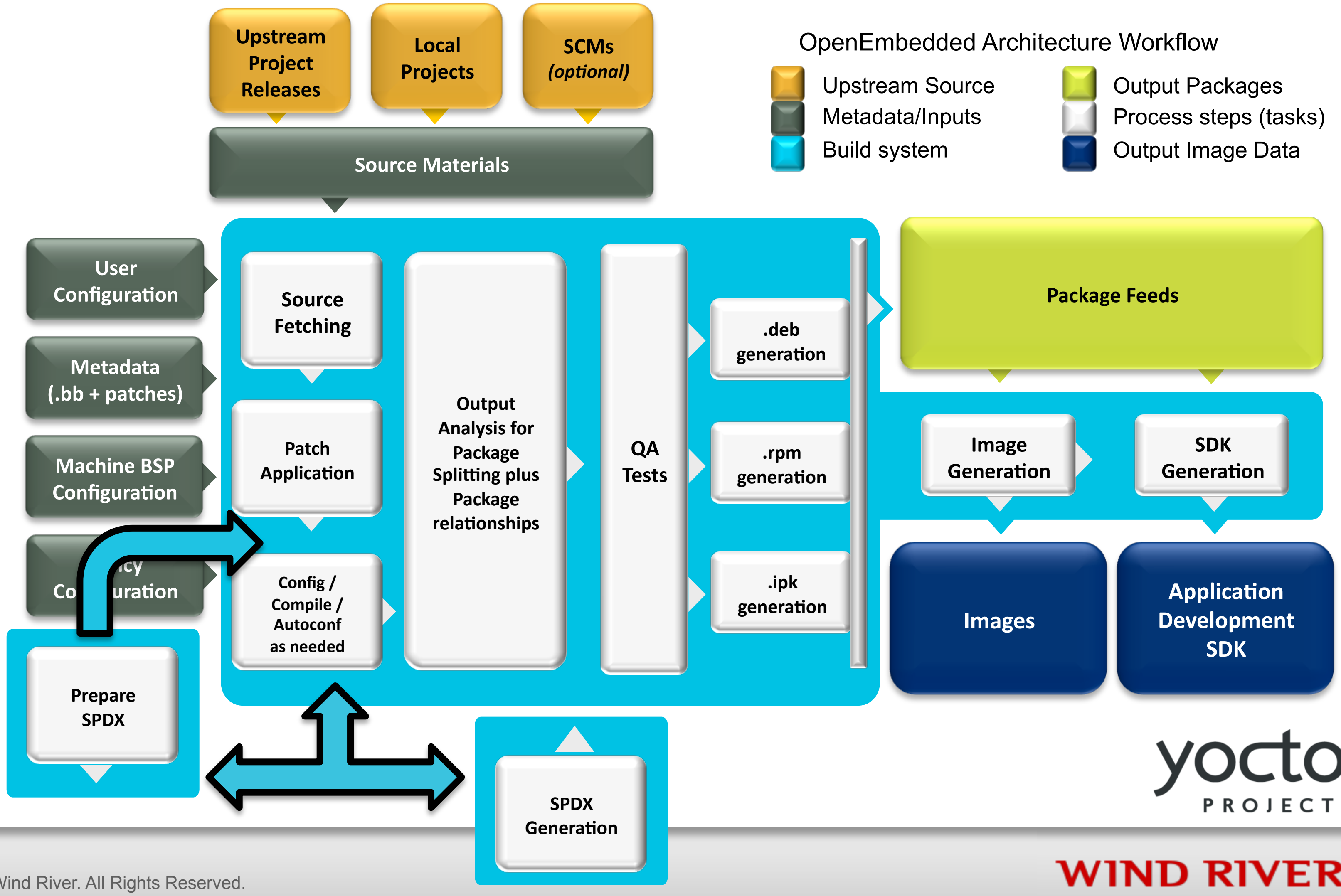
- Additional Experimental changes
  - Remove ‘native’ packages from calculations
    - Never shipped to end customers


600
500
400
300
200
100
0
no spdx
common cache - no copyright
no native - no copyright
common cache - copyright
no native - copyright
spdx_cached no copyright
spdx_cached copyright
Real
User
Sys

Build System Workflow



Build System Workflow

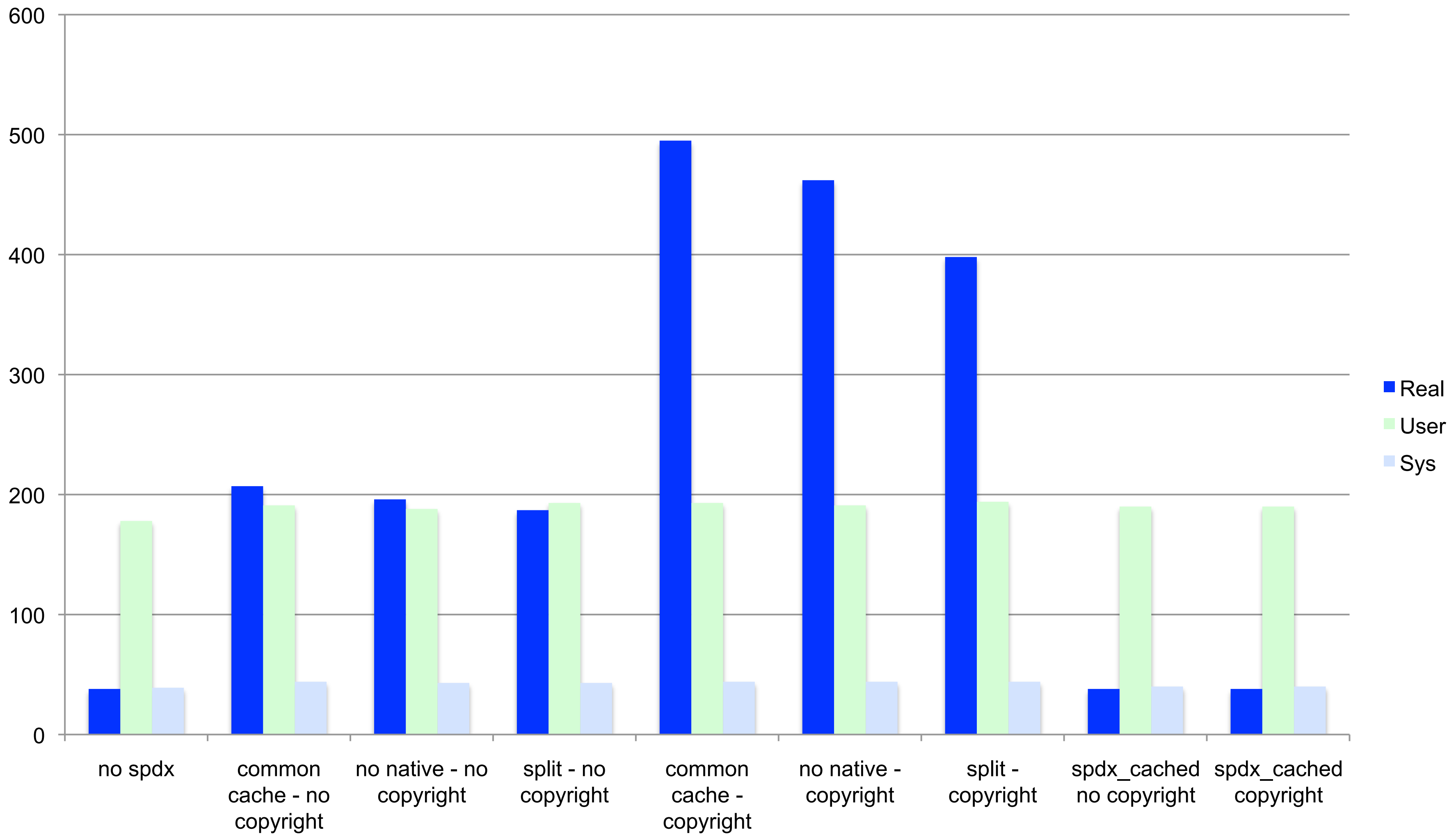


yocto PROJECT

WIND RIVER

Performance

- Additional Experimental changes
  - Split do_spdx into two section, unblock the build process


600
500
400
300
200
100
0
no spdx
common cache - no copyright
no native - no copyright
split - no copyright
common cache - copyright
no native - copyright
split - copyright
spdx_cached no copyright
spdx_cached copyright
Real
User
Sys

Findings…

- FOSSology
  - Performance
    - Much of the FOSSology processing is single threaded per package
    - Causes long connection times, which if closed cause retry’s or hangs
- Yocto Project
  - Recipes without upstream sources
    - Empty Archives
    - WORKDIR = S, odd contents sent for processing
  - sstate-cache

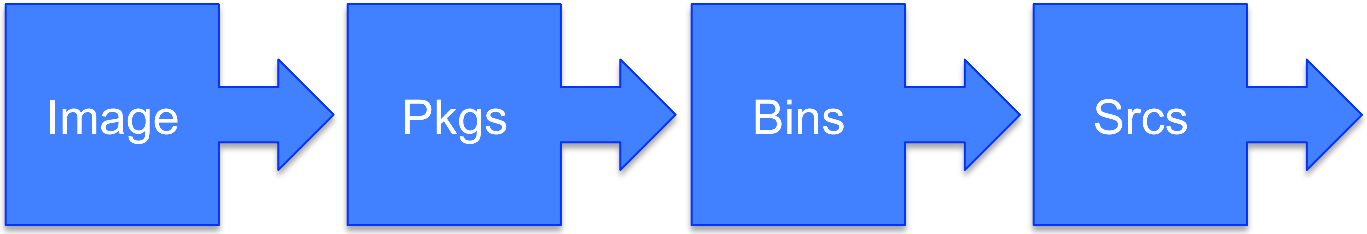
✓ Future Work

Future Work – FOSSologySPDX

- Generalize service beyond SPDX+Yocto
  - Build generic API/service
- Integrate service as part of other build processes
- Automatic SPDX upload/population build process
  - Dashboard
- Performance Improvements:
  - Global SPDX package cache
  - Multiprocessing? (FOSSology limitations)

Yocto Project Vision – End to End

- Trace the files installed into the image back to their respective packages
  - Allows traceability back to SPDX files
  - Allows traceability back to the original source code



Future Work – Yocto Project

- sstate-cache integration
  - Integrate json cache, generated SPDX files and other related components into sstate-cache
- Human Reviewed/Human Generated SPDX support
- Binary to Source correlation
  - Determine binary licenses (package based)
  - Determine image licenses (image manifest)
- Tools for working with SPDX files
  - Editing tools for modifying json/SPDX files
  - Generating Legal Manifest
  - Generating Notices file

WIND RIVER