SPDX 2.0

what, why, how & specifics

Software Package Data Exchange®

- **Standard:**
  - A standard format for communicating the components, licenses and copyrights associated with a software package.
  - Key pillar in Linux Foundation's Open Compliance Program
- **SPDX Group:**
  - Workgroup of Linux Foundation
  - Participation from over 20 organizations including software, systems and tool vendors, consultants and foundations
- **Vision:**
  - To help reduce redundant work in determining software license information and facilitate compliance

Specification Elements

- **File Format:**
  - For license and copyright information to accompany packages
  - Can reflect package hierarchy through relationships
  - Guiding Principle: Focus on capturing facts; avoid interpretations.
- **License List:**
  - A standardized short form to refer to common licenses
  - Handles common license exceptions
- **Benefits**
  - Allows easy exchange of license information between companies reducing burden on both suppliers and consumers
  - Avoids due diligence redundancy where the same source code package is analyzed multiple times by different receivers
  - Provides a unified method for exchanging license information

SPDX® 2.0 - what's new?

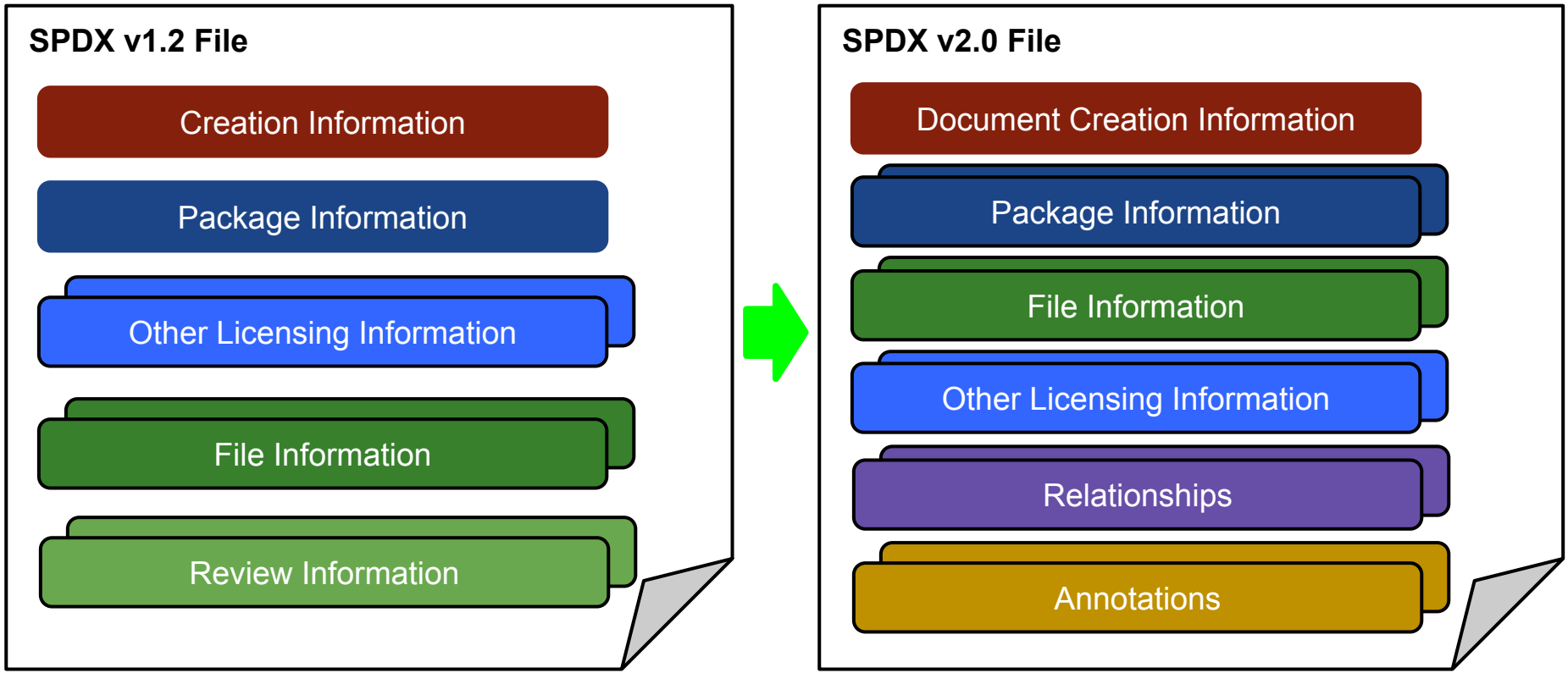
- Multiple packages can now be described in a single SPDX document.
- Relationships between packages, files, and external SPDX documents, can now be described.
- Annotations can be provided on any specific element in an SPDX document.
- Additional file types & checksum algorithms are now supported.
- A new License expression syntax has been introduced with improved license matching guidelines.
- License exceptions are separate section in license list.

SPDX® 2.0 - what's changed from 1.2?

- Review Information Section **replaced** by Annotations.
  - now able to provide specific information on file, package or document level.

- Document and Creation Information Sections **merged** into a single section.
  - all fields from 1.2 remain, just regrouped, and some additional ones added.

The SPDX Document



License List

License Identifier	Recognized Exceptions	Full name of License
AFL-3.0		Academic Free License 3.0
AGPL-3.0		(GNU) Affero General Public License v3
APL		Adaptive Public License
ASL-2.0		Apache License, 2.0
APSL-2.0		Apple Public Source License 2.0
Artistic-2.0		Artistic license 2.0
AAL		Attribution Assurance License
BSD-4-Clause		BSD 4-clause "Original" or "Old" License
BSD-3-Clause		BSD 3-clause "New" or "Revised" License
BSD-2-Clause		BSD 2-clause "Simplified" or "FreeBSD" License
BSL-1.0		Boost Software License 1.0
CATOSL-1.1		Computer Associates Trusted Open Source License 1.1
CC-BY-1.0		Creative Commons Attribution 1.0
CC-BY-NC-1.0		Creative Commons Attribution Non Commercial 1.0
CC-BY-ND-1.0		Creative Commons Attribution No Derivatives 1.0
CC-BY-SA-1.0		Creative Commons Attribution Share Alike 1.0
CC-BY-NC-ND-1.0		Creative Commons Attribution Non Commercial No Derivatives 1.0
CC-BY-NC-SA-1.0		Creative Commons Attribution Non Commercial Share Alike 1.0
CC-BY-2.0		Creative Commons Attribution 2.0
CC-BY-NC-2.0		Creative Commons Attribution Non Commercial 2.0
CC-BY-ND-2.0		Creative Commons Attribution No Derivatives 2.0
CC-BY-SA-2.0		Creative Commons Attribution Share Alike 2.0
CC-BY-NC-ND-2.0		Creative Commons Attribution Non Commercial No Derivatives 2.0
CC-BY-NC-SA-2.0		Creative Commons Attribution Non Commercial Share Alike 2.0

open source initiative

- ~300 Licenses
  - Short IDs for easy reference
  - Exact text of licenses
  - Available on SPDX® website – URLs won't change
- License Matching Guidelines
  - For matching licenses against those included on the SPDX License List
- License Templates
  - Denote license text which is optional or replaceable per the license matching guidelines
- Separate Exceptions List
  - Common modifications to some licenses
  - Simple expression language for expressing

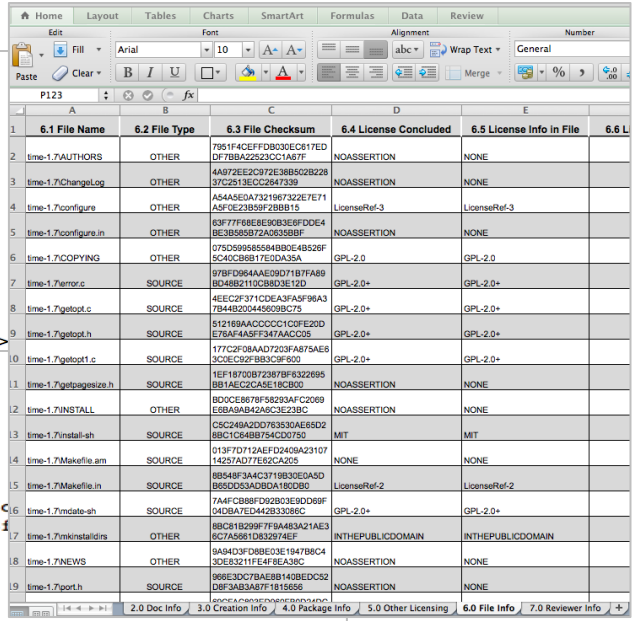
SPDX® 2.0 - what's the same as 1.2?

- **Most of it!**
  - Approx 90% of the fields are basically the same as in 1.2 (42/46).
  - The 4 depreciated have been replaced with more generalized support.
- Linkage to External Licenses
  - more licenses added to recognized license list.
- Same basic file formats supported
  - Tag:Value
  - RDF/XML
  - translation to spreadsheets

Support Forms: RDF & Tag:Value Spreadsheet thru translation

```
– <rdf:Description rdf:nodeID="A0">
    <rdfs:comment/>
  – <copyrightText>
      * <I>Copyright</I> (C) 1999 by Randolph Chung &lt;tausq@debian.org&gt;
    </copyrightText>
    <licenseComments/>
    <licenseInfoInFile rdf:resource="http://spdx.org/licenses/GPL-2.0+"/>
    <licenseConcluded rdf:resource="http://spdx.org/rdf/terms#noassertion"/>
    <fileType rdf:resource="http://spdx.org/rdf/terms#fileType_source"/>
    <checksum rdf:nodeID="A1"/>
    <fileName>networking/hostname.c</fileName>
    <rdf:type rdf:resource="http://spdx.org/rdf/terms#File"/>
  </rdf:Description>
– <rdf:Description rdf:nodeID="A2">
    <checksumValue>dc90a437e03f31ab04e7059d8da5f88b28cde77d</checksumValue>
    <algorithm rd
    <rdf:type rdf:
  </rdf:Descriptio
– <rdf:Descriptio
    <rdfs:comme
    <copyrightTe
    <licenseComm
```

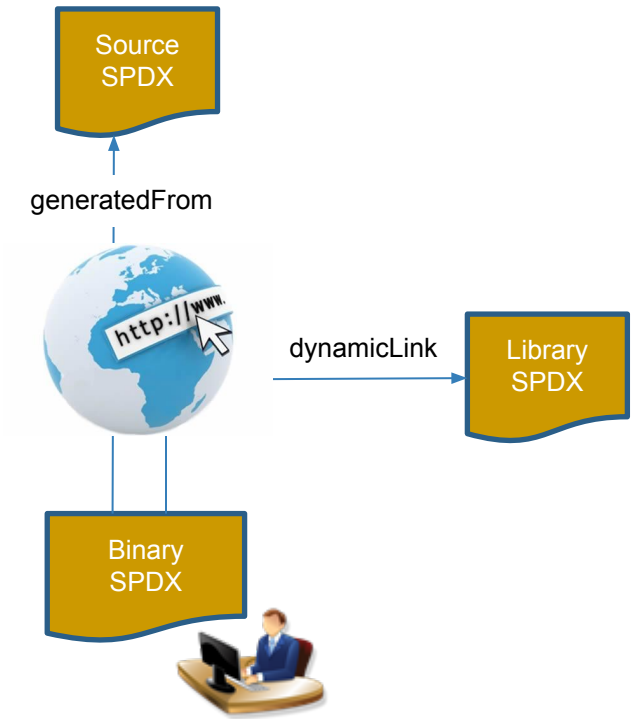
```
##-------------------------
## Package Information
##-------------------------

PackageName: time-1.7.tar.gz
PackageFileName: time-1.7.tar.gz
PackageDownloadLocation: NOASSERTION
PackageVerificationCode: dd5cf0b17bfef4284c6
PackageChecksum: SHA1: dde0c28c74269607369331
PackageLicenseConcluded: GPL-2.0+
PackageLicenseDeclared: GPL-2.0+
PackageLicenseInfoFromFiles: GPL-2.0
PackageLicenseInfoFromFiles: GPL-2.0+
PackageLicenseInfoFromFiles: MIT
PackageLicenseInfoFromFiles: LicenseRef-1
PackageLicenseInfoFromFiles: LicenseRef-2
PackageLicenseInfoFromFiles: LicenseRef-3
PackageCopyrightText: NOASSERTION

##-------------------------
```



	A	B	C	D	E
1	6.1 File Name	6.2 File Type	6.3 File Checksum	6.4 License Concluded	6.5 License Info in File
2	time-1.7\AUTHORS	OTHER	7951F4CEFFDB030EC817ED DF7BBA22523CC1A67F	NOASSERTION	NONE
3	time-1.7\ChangeLog	OTHER	4A972EE2C972E38B502B228 37C2513ECC2847339	NOASSERTION	NONE
4	time-1.7\configure	OTHER	A54A5E0A7321967322E7E71 A5F0E23B59F2BBB15	LicenseRef-3	LicenseRef-3
5	time-1.7\configure.in	OTHER	63F77F68E8E90B3E6FDDE4 BE3B585B72A0635BBF	NOASSERTION	NONE
6	time-1.7\COPYING	OTHER	075D599585584BB0E4B526F 5C40CB6B17E0DA35A	GPL-2.0	GPL-2.0
7	time-1.7\error.c	SOURCE	97BFD964AAE09D71B7FA89 BD48B2110CB8D3E12D	GPL-2.0+	GPL-2.0+
8	time-1.7\getopt.c	SOURCE	4EEC2F371CDEA3FA5F96A3 7B44B2004456609B275	GPL-2.0+	GPL-2.0+
9	time-1.7\getopt.h	SOURCE	512189AACCCC1C9FE20D E76AF4A5FF347AACC05	GPL-2.0+	GPL-2.0+
10	time-1.7\getopt1.c	SOURCE	177C2F08AAD7203FA875AE6 3C0EC92FBB3C9F600	GPL-2.0+	GPL-2.0+
11	time-1.7\getpagesize.h	SOURCE	1EF18700B72387BF6322695 BB1AEC2CA5E18CB00	NOASSERTION	NONE
12	time-1.7\INSTALL	OTHER	BD0CE8678F58293AFC2069 E6BA9AB42A6C3E23BC	NOASSERTION	NONE
13	time-1.7\install-sh	SOURCE	C5C249A2DD763530AE65D2 8BC1C64BB754CD750	MIT	MIT
14	time-1.7\Makefile.am	SOURCE	013F7D712AEFD2409A23107 14257AD77E62CA205	NONE	NONE
15	time-1.7\Makefile.in	SOURCE	8B548F3A4C3719B30E0A5D B65DD53ADBDA180DB0	LicenseRef-2	LicenseRef-2
16	time-1.7\mdate-sh	SOURCE	7A4FCB88FD92B03E9DD69F 04DBA7ED442B33086C	GPL-2.0+	GPL-2.0+
17	time-1.7\mkinstalldirs	OTHER	8BC81B299F7F9A483A21AE3 6C7A5661D832974EF	INTHEPUBLICDOMAIN	INTHEPUBLICDOMAIN
18	time-1.7\NEWS	OTHER	9A94D3FD8BE03E1947B8C4 3DE83211FE4F8EA38C	NOASSERTION	NONE
19	time-1.7\port.h	SOURCE	966E3DC7BAE8B140BEDC52 D8F3AB3A87F1815656	NOASSERTION	NONE

Key Use Cases to Support

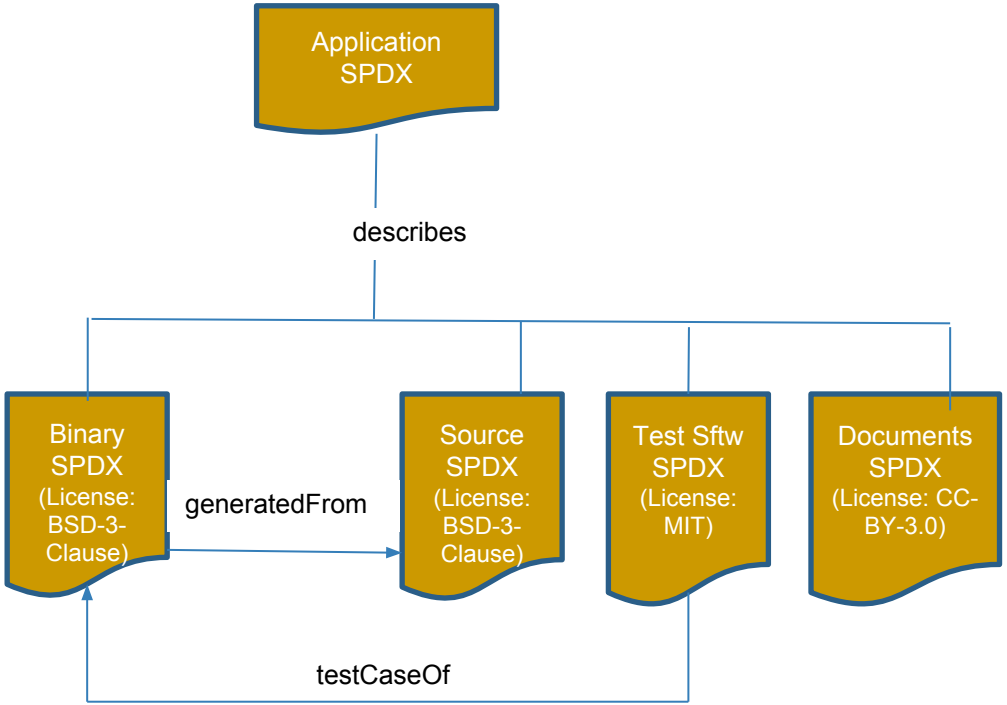
Relationship Use Cases



Binary only delivery

- SPDX for the binary points to SPDX doc for the code used to build it (generatedFrom)
- SPDX for the binary points to SPDX doc for a library it links with at run time (dynamicLink)

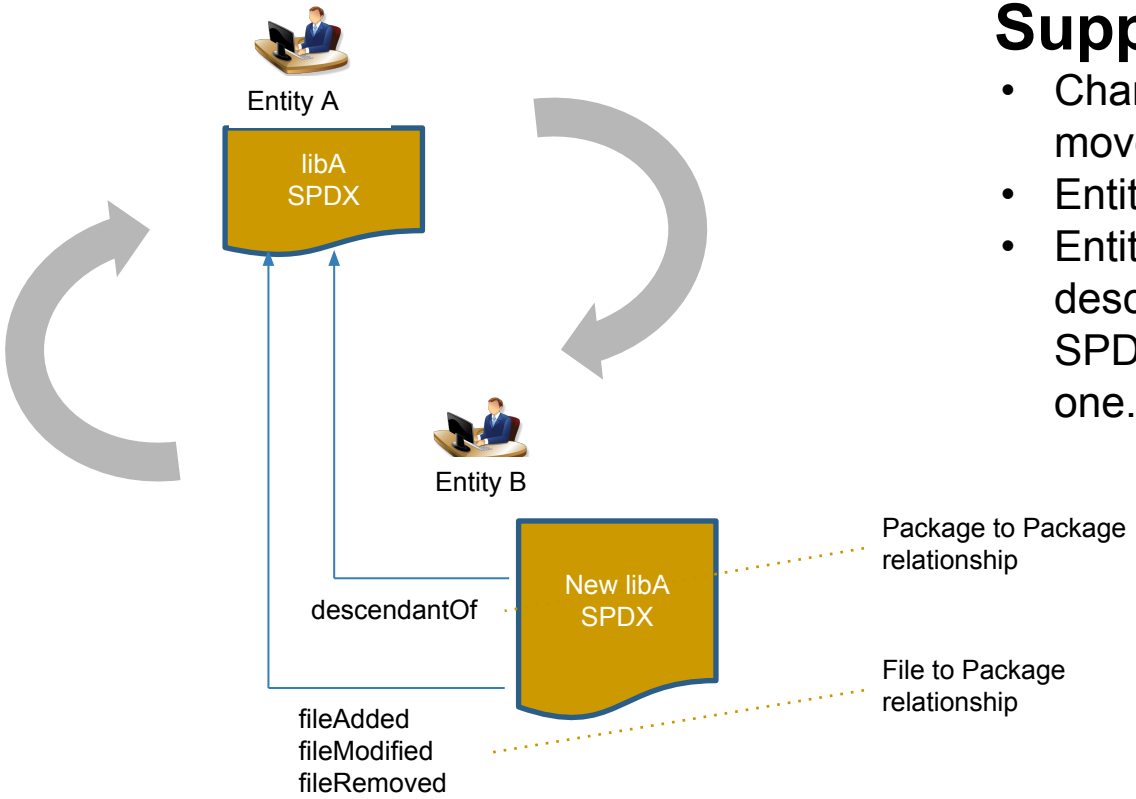
Relationship Use Cases



More precise description of the “bits”

- Instead of a single SPDX file with “the kitchen sink” or multiple ones that must have a document to say what they are, we can now be more precise and have the docs refer to themselves

Relationship Use Cases



Supply Chains

- Changes can be tracked as software moves through a supply chain
- Entity A gives a library, libA, to Entity B
- Entity B makes changes to libA and describes those changes with a new SPDX doc that refers to the original one.

Reworking the Underlying Model

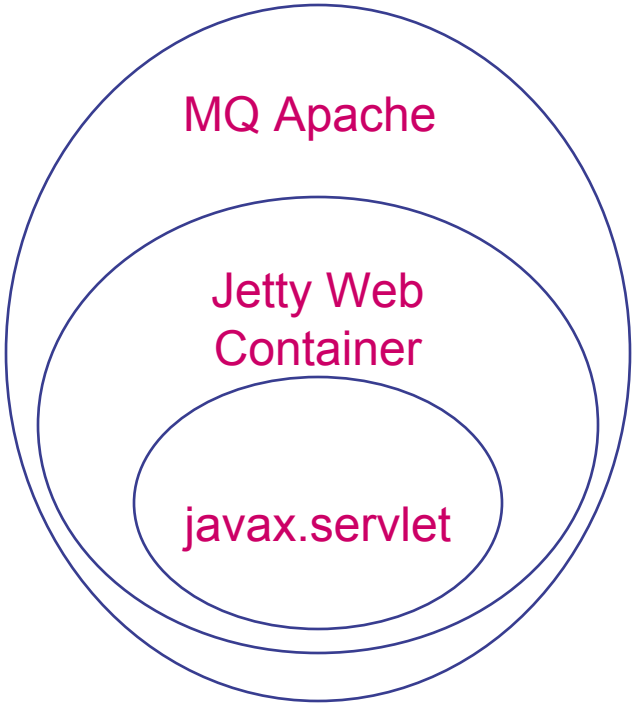
SPDX® 2.0 - Model Overview

- Result of merging two model proposal
- Designed to support all of the proposed use cases for 2.0 (and then some)
- Contains several new “abstractions” to allow for future extensions

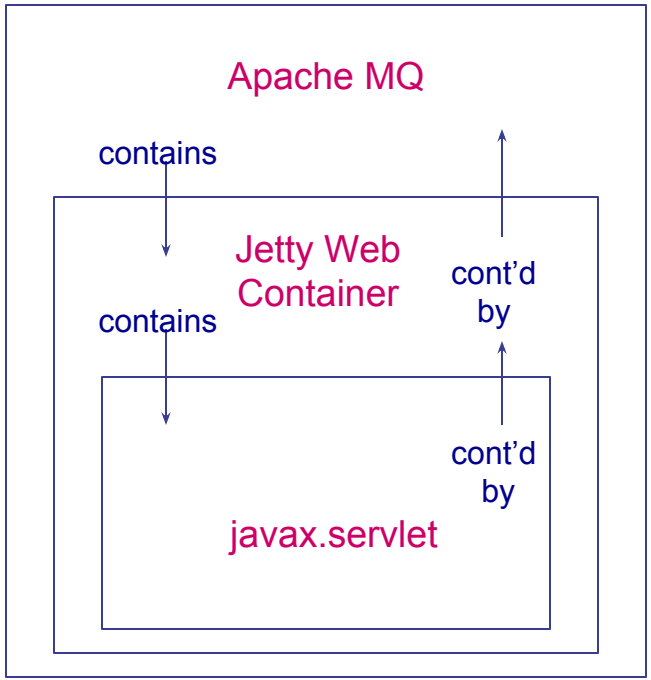
- Available in the spec and at http://wiki.spdx.org/view/Technical_Team/Model_2_0

SPDX handles Package Relationships

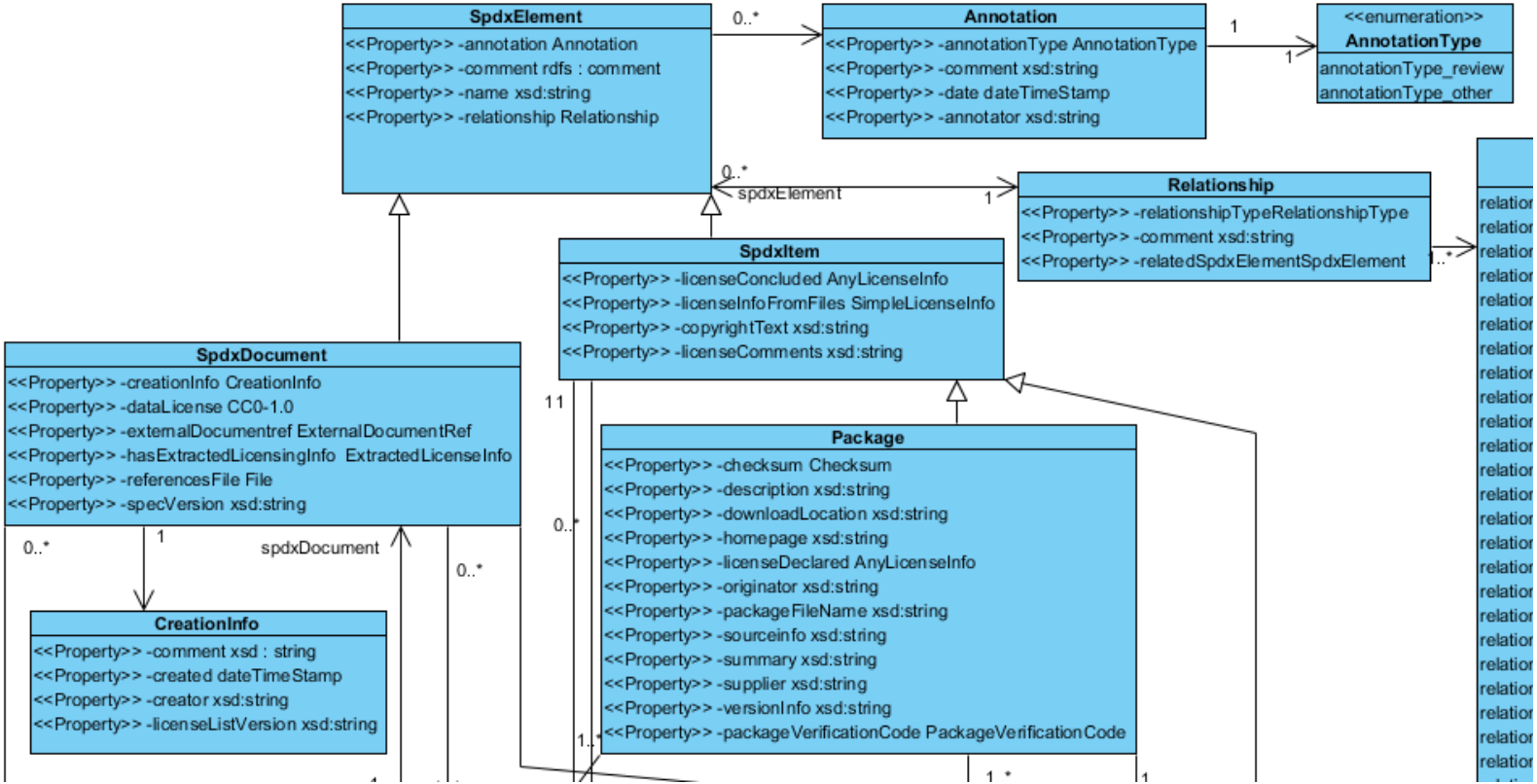
Package



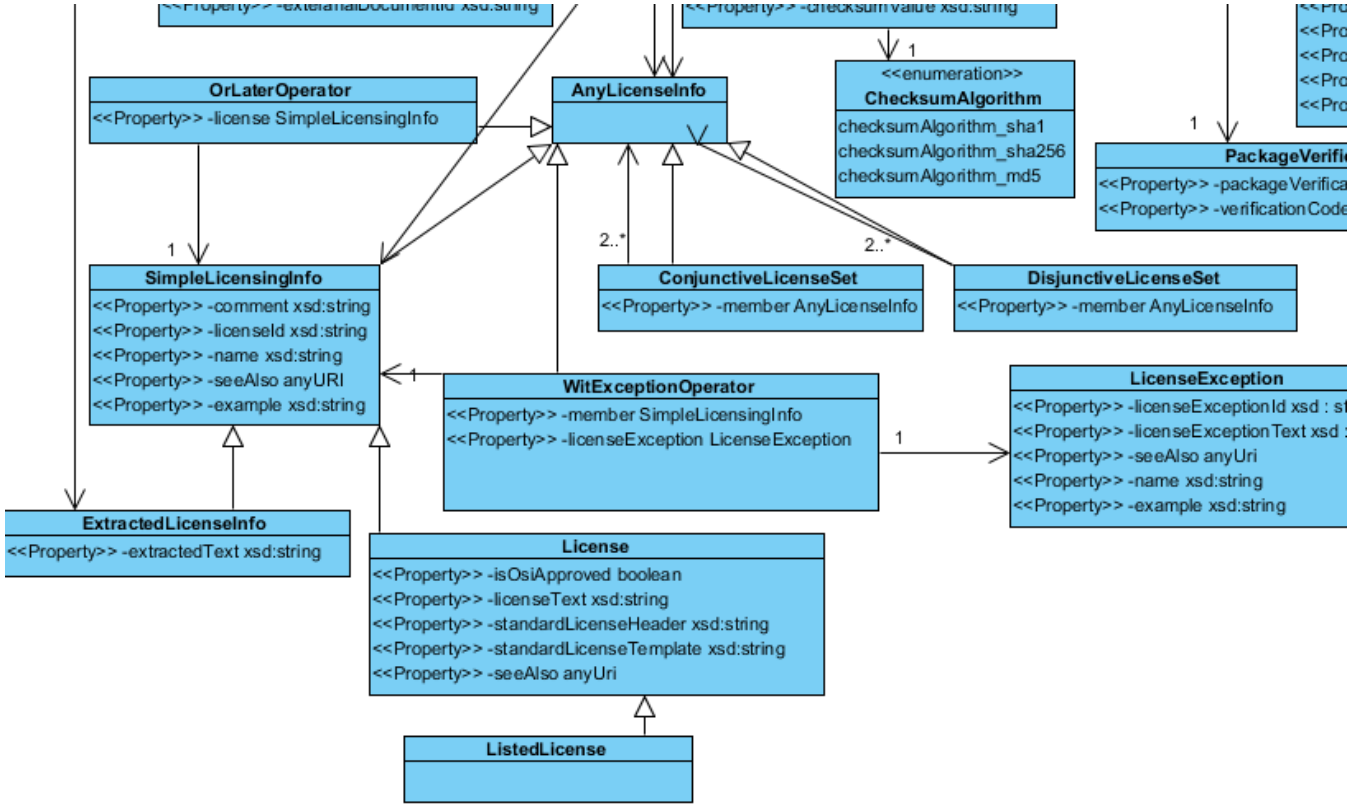
SPDX Doc



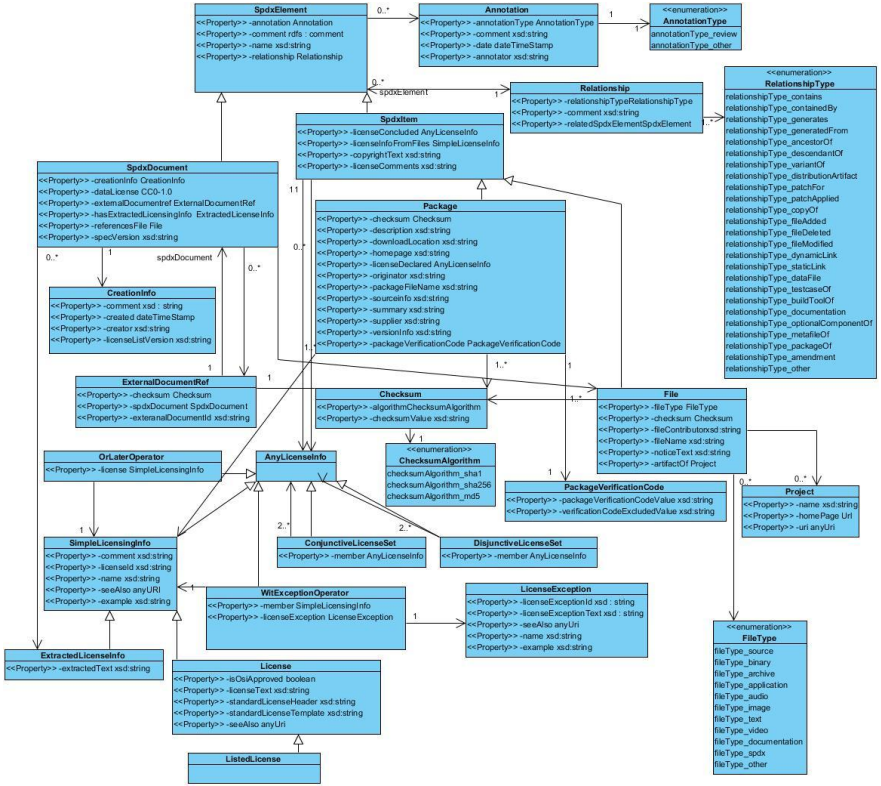
SPDX Elements - the fundementals



Licenses

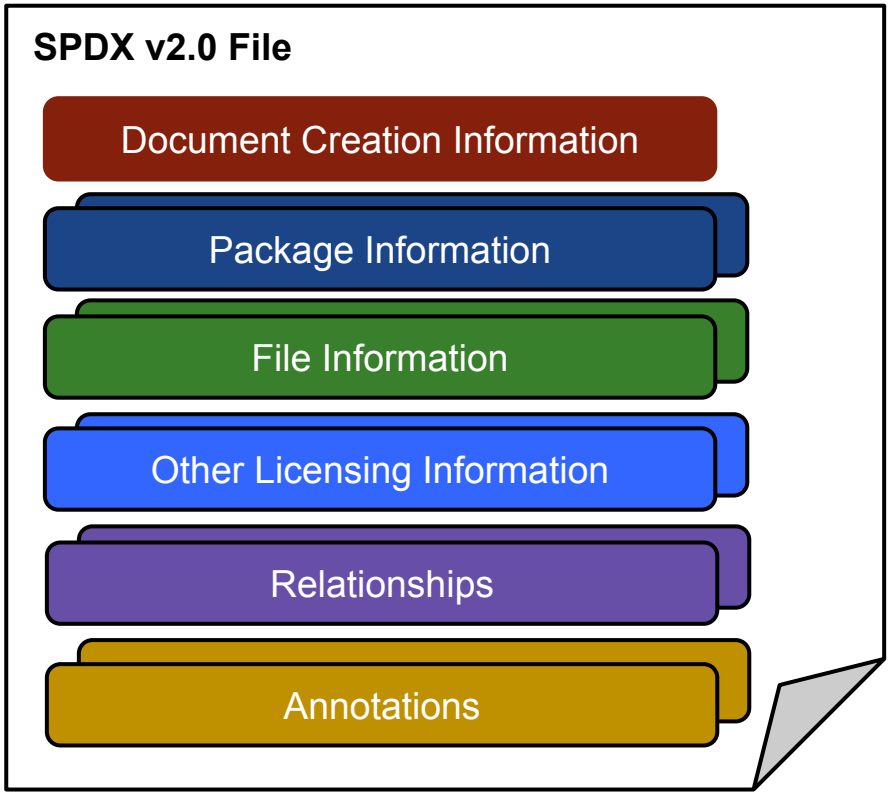

OrLaterOperator
<<Property>> -license SimpleLicensingInfo
AnyLicenseInfo
<<enumeration>>
ChecksumAlgorithm
checksumAlgorithm_sha1
checksumAlgorithm_sha256
checksumAlgorithm_md5
SimpleLicensingInfo
<<Property>> -comment xsd:string
<<Property>> -licenseId xsd:string
<<Property>> -name xsd:string
<<Property>> -seeAlso anyURI
<<Property>> -example xsd:string
ConjunctiveLicenseSet
<<Property>> -member AnyLicenseInfo
DisjunctiveLicenseSet
<<Property>> -member AnyLicenseInfo
WitExceptionOperator
<<Property>> -member SimpleLicensingInfo
<<Property>> -licenseException LicenseException
LicenseException
<<Property>> -seeAlso anyUri
<<Property>> -name xsd:string
<<Property>> -example xsd:string
ExtractedLicenseInfo
<<Property>> -extractedText xsd:string
License
<<Property>> -isOsiApproved boolean
<<Property>> -licenseText xsd:string
<<Property>> -standardLicenseHeader xsd:string
<<Property>> -standardLicenseTemplate xsd:string
<<Property>> -seeAlso anyUri
ListedLicense
1
2..*
2..*

The Big (and complex) Picture



How the model got implemented

SPDX® 2.0 - Section Overview



Document and Creation Information

- SPDX Version (used in creation of SPDX file)
- Licensing of meta data
- SPDX Identifier for the document itself
- Name of this Document
- SPDX Document Namespace (URI)
- External SPDX Doc References
- License List Version
- Creator (how was the file created)
  - Manual review (who, when)
  - Tool (id, version, when)
- When was it created
- Comments on creator and document itself

Package Information

- Identification
  - Formal Name of Package (Full name given by originator and version information)
  - SPDX Identifier (unique ID for referencing from elsewhere)
  - Package File Name (Name package obtained under (.tar, .rpm, etc.))
  - Package Supplier and Originator
  - Package Download Location (download URL)
  - Package Verification Code and Checksum (SHA1, MD5, SHA256)
  - Package Homepage and Source Information
- Licensing for Package
  - Declared License- License(s) that has/have been asserted for the package
  - Concluded License- License that Creator has concluded
  - List of file licenses
  - Comments Field (for example, to explain conclusion)
- Copyright Text
- Description of Package (summary and detailed options) and comments about the package

File Information

- Identification
  - File Name
  - SPDX Identifier (for referencing from elsewhere)
  - File Type (source, binary, archive,application,audio,image,text,video,documentation,spdx)
  - Artifact of Project Name, Homepage & URI (project it came from)
  - File Checksum (SHA1, MD5, SHA256)
- Licensing for File
  - Concluded License (license determined by SPDX file creator)
  - License Information in File
  - Comments on License
- Copyright Text
- File Notices
- File Contributor
- ~~File Dependencies~~
- File Comments

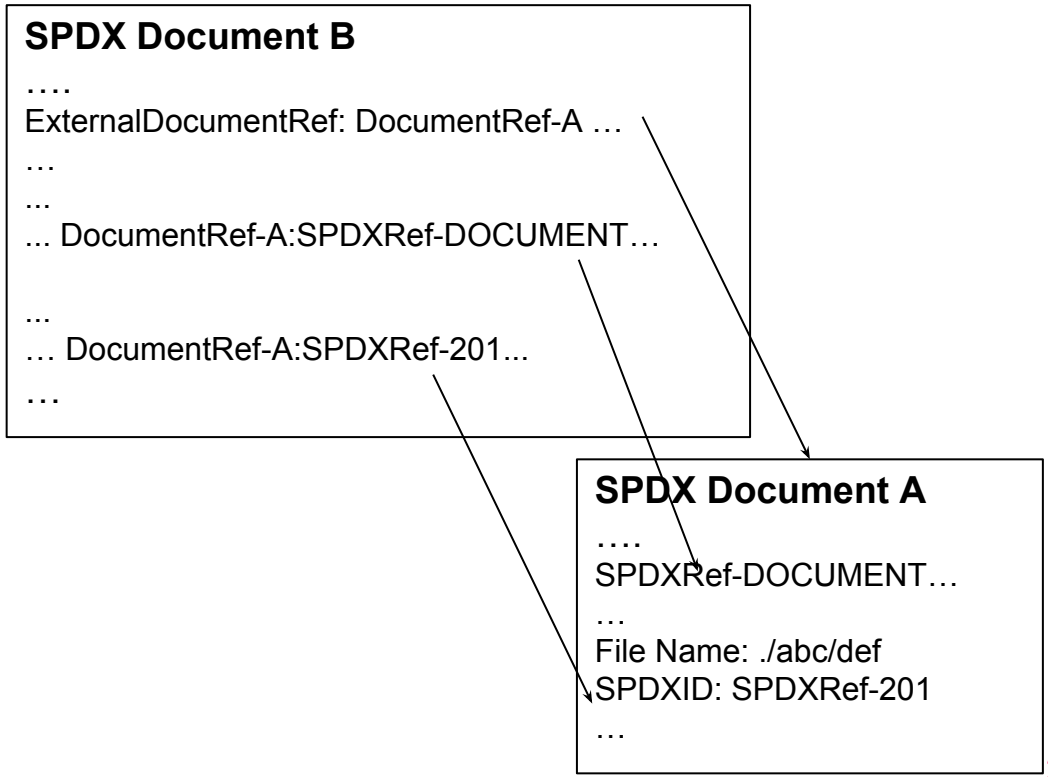
Other Licensing Information

- Identifier Assigned (unique short form to this document)
- Extracted Text
- Name of License
- Cross References
- Comments

- NOTES:
  - Provides a way to identify licenses not on the SPDX License List.
  - Aim for ~90% coverage with standard short forms NOT exhaustive
  - Although there are a lot of licenses “in the wild,” a smaller number covers most projects:
    - Black Duck identifies >2000 licenses in use , however ~20 licenses responsible for nearly all licensed open source projects
      - http://www.blackducksoftware.com/oss/licenses#top20
    - and, OSI currently recognizes 69 licenses as “open source”
      - http://www.opensource.org/licenses

Relationships between Elements

- Each SPDX Document has a unique identifier
- Elements within a document have an identifier unique to the SPDX document (e.g. Document itself, Package, File & License)
- Elements in external documents are referenced using the external document ID followed by the local unique reference.



Review Information

- ~~Reviewer~~
- ~~Review Date~~
- ~~Review Comment~~

REPLACED BY Annotations

Annotations

- Annotation allows for comments on **any** SPDX element
- Annotations can provide a changelog for any changes made to specific SPDX elements
- Annotations contain:
  - annotator (the person, company, or tool which provided the annotation)
  - date the annotation made
  - type of annotation (review or other)
  - SPDX identifier reference (element the annotation refers to)
  - comments

RDF Overview

- What?
  - Resource Description Framework - standard for encoding data for the Semantic Web
- Why?
  - Precise
  - Widely adopted
  - Web based standard
  - Support for “reasoning”

SPDX 2.0 RDF Changes

- Additional classes and properties to match the SPDX 2.0 model

- Use of the SPDX document namespace to uniquely identify all SPDX elements in the document
  - All documents will have a unique URI for a namespace
  - All elements will have a URI with the namespace + #ElmentID

RDF and Tag:Value

- Tools to translate both ways
- Common names for “most” of the properties
  - Exceptions for enumeration values which must be unique in RDF (e.g. annotationType_review = REVIEW)
- Document Namespace tag key to URI

Notes for Implementers

- Careful of the Infinite Recursion of Relationships
- External Document References key to building URI's for external documents
- Leverage existing implementations (git.linuxfoundation.org)
- The RDF schema can be found at http://spdx.org/rdf/ontology/spdx-2-0-rev-11/

QUESTIONS?

Thank you!

Backup Slides

Resources for SPDX®

- Open Source Tools (hosted on SPDX Git Repo)
  - Viewer
  - Spreadsheet to RDF/Tag Value xlator
  - RDF/Tag Value to Spreadsheet xlator
  - License file generator (from Spreadsheet)
  - Spreadsheet template
  - FOSSology via University of Nebraska Omaha
- Commercial Tools
  - Scanning tools to provide SPDX® support
- http://spdx.org/

Overview

Kate - SPDX 2.0 new features

Jack - New use cases enabled by the features

Kate - Specification overview

Gary - Model overview

Kate - tag/value overview

Gary - RDF overview

Jack - Resources

Getting involved…

- See:
  - http://www.spdx.org
  - Mailing lists, meetings, wiki
- Contact:
  - Phil Odence (Chair) - podence@blackducksoftware.com
  - Kate Stewart (Tech Team Chair) - stewart@linux.com
  - Gary O'Neal (Tools Lead) - gary@sourceauditor.com
  - Jilayne Lovejoy (Legal Team Co-Chair) - opensource@jilayne.com
  - Paul Maddick, (Legal Team Co-Chair) - paul.madick@hp.com
  - Jack Manbeck (Business Team Co-Chair) - j-manbeck2@ti.com
  - Mikael Söderberg (Business Team Co-Chair) - mikael.soderberg@pelagicore.com

Contribution

- Runs like an open source project without centralized constitution or bylaws
- Intellectual property contributed by participants members is covered under the Creative Commons license (CC-BY-3.0)
- Tools developed by the work group are licensed under the Apache 2.0 license (Apache-2.0) maintained in a Git repository
- http://spdx.org

Workgroup Operation

- Structure
  - General Meeting and mailing list
  - Teams with separate meetings and lists
    - Technical
    - Business
    - Legal
- Very inclusive process
  - Self-subscription for interested participants
  - Those willing to “do” can influence direction
  - Mail-list, WIKI, phone calls, BOFs…
  - Face to face meetings at Linux Foundation and other events