Open Compliance Summit

15 November 2017

THE LINUX FOUNDATION

A note on Chatham House Rules

"When a meeting, or part thereof, is held under the Chatham House Rule, participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed."

See more at: http://www.chathamhouse.org/about/chatham-house-rule#sthash.q6ybxH03.dpuf


As systems require data to learn and evolve, no one organization can build, maintain and source all data required.


Data communities are forming
• AI and ML use cases
• Autonomous systems
• Connected civil infrastructure

Data is not the same as source code

- In the US and elsewhere, data itself is generally not protectable IP (see Feist Publications, Inc., v. Rural Telephone Service Co.)[1]
- Only the creative expression of the data is protectable by copyright; Facts are not
- Some data provider organizations are trying any means available to lock down access to data, sometimes with direct or ambiguous terms around usage rights
  - "*Intellectual Property Rights means the rights in and to patents, trademarks, service marks, trade and service names, copyrights, database rights and design rights, rights in know-how, moral rights, trade secrets and all rights or forms of protection of a similar nature or having similar or equivalent effect which may subsist anywhere in the world now existing or hereafter arising*."

[1] Available at: http://caselaw.findlaw.com/us-supreme-court/499/340.html

The CDLA license agreements enable sharing data openly, embodying best practices learned over decades sharing source code.

Community Data License Agreement

- On October 23, we announced Version 1.0 of the Community Data License Agreements
- There are two CDLA license agreements:
  - "Sharing" – based on a form of copyleft, designed to encourage recipients to participate in reciprocal sharing of data
  - "Permissive" – an approach similar to permissive open source licenses (e.g. Apache, BSD or MIT) where recipients are not required to share any changes

If there is a sharing obligation (e.g. copyleft), where does it begin and end?

- Includes:
  - Modifications to data received
  - Additions to data received
- Excludes:
  - The results of any analysis
    - Results may be included voluntarily
    - Contributions will be limited if results have to be shared
    - Similar to internal use exclusion in GPL

But what about Personally Identifiable Information?

- Each Data Provider represents that Publication of the Data that it Publishes does not violate any privacy or confidentiality obligation undertaken by that Data Provider.
- If You choose to Publish Data that You have Received under the Agreement, You are not asked to make a representation that no other Data Provider has included Data that is subject to a privacy or confidentiality obligation that was undertaken by that Data Provider.
- Does that mean that You can pass along Data when You know that someone else has inserted personal or confidential information into that Data?
  - No. Each Data Provider represents that the Data Provider has exercised reasonable care to assure that the Data it Publishes was obtained from others with the right to Publish the Data under this Agreement.
  - Furthermore, although the Agreement may contain no requirement to make representations on behalf of other Data Providers, You are still required to comply with all applicable laws in Publishing and Using Data Received under the Agreement.

The CDLA in use – Cisco's Network Anomaly Telemetry data

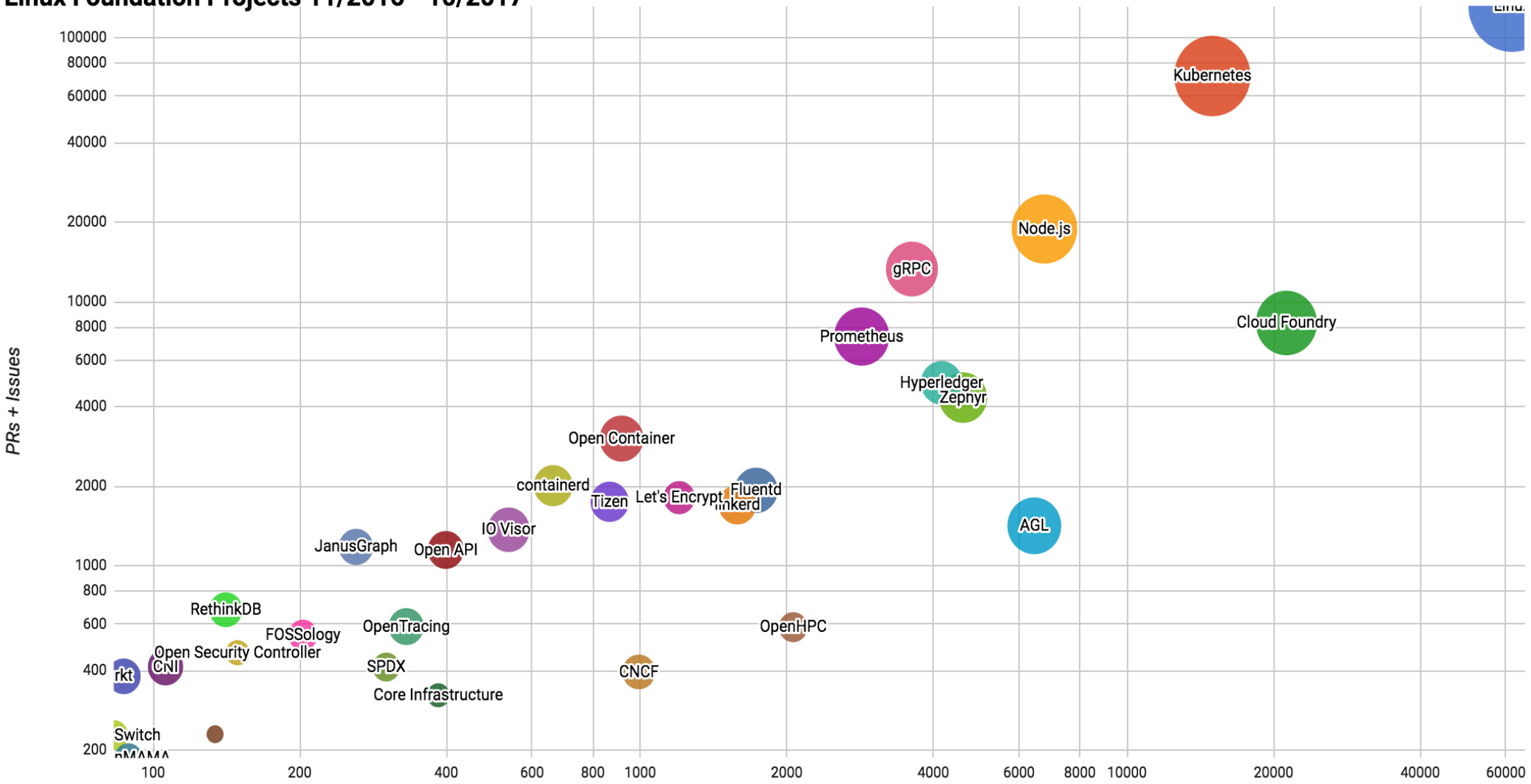
- https://github.com/cisco-ie/telemetry
- The data sets are based around network anomalies, e.g. port flaps, bgp issues, optic failures, etc.
- The purpose is to allow the development of models to identify the unique signatures of the events as close to the actual time of the event versus identifying it minutes after.
- Cisco is working with a number of universities who are adding the data sets in to their data science and research coursework at both undergrad and graduate levels.
- This level and type of network anomaly data sets have not been available for the data science and machine learning communities let alone the majority of companies to use in developing automation and remediation of network events.
- Liked that CDLA is a data-specific agreement as opposed to being just a copyright license, which doesn't really fit data well

cdla.io

Open Source License Compliance in 2016

- Basic tenants of compliance best practices are well established and are helping those who've automated also deal with security risk mitigation
- Permissive licenses tend to pose few issues – copyright notices are fairly well established and often automated b/c they are predictable
- Copyleft licenses are typically the areas of concern as they may trigger additional work
- Most compliance challenges we hear about tend to arise from an issue in processes related to product development or sourcing from suppliers
- Tooling and supply chain standards are the answer to enable repeatable, dependable license compliance workflows – automation brings down the cost of compliance and reduces opportunity for errors

Linux Foundation Projects 11/2016 - 10/2017



Trends we see in projects

- Exponential Scale: more vertical industries involved, broader use cases, more developers, more companies
- Every form factor of technology uses open source: sensors, wearables, embedded, client, server, cloud, mainframe and HPC
- More "new to open source" companies than ever
- More "new" companies, means a greater demand for education, help, guidance and collaboration we all provided each other years and in some cases decades ago
- Weak compliance does not get better as open source adoption scales out exponentially

Tools and standards continue to evolve and provide opportunities to automate

- SPDX: specification, license list and short code identifiers to document the license for files in a codebase and easily exchange the license information with others through a machine readable specification
- OpenChain: building a standard for use by companies with their supply chain to ensure training, auditability and conformance of processes with best practices for compliance with open source software license requirements.
- FOSSology: an open source license scanning tool to identify licenses and potential compliance issues, useful to generate SPDX files and maintain OpenChain conformant processes.
- Cregit: an open source contribution analysis tool that identifies contributions down to the smallest parse-able unit of code ("token").
- New tools on the horizon: Intel's software artifact blockchain, Grafeas, Quartermaster build integration, etc.

Cregit evolved this year adding email2git which shows LKML patch discussion history

```
if (args.addr == 0)
        pr_err("symbol '%s' not found in symbol table\n", name);
else if (args.count > 1 && sympos == 0) {
        pr_err("unresolvable ambiguity for symbol '%s' in object '%s'\n",
               name, objname);
} else if (sympos != args.count && sympos > 0) {
        pr_err("symbol position %lu for symbol '%s' in object '%s' not found\n",
               sympos, name, objname ? objname : "vmlinux");
```

b2b018ef48675a9a524fa9791ea7d67fdac405f7

View commit on github | View file on LXR

Patch (beta)

LKML

Sent: 11/13/2015, 1:59:50 AM

Patch Discussion / Previous Attempts

#	Patch	Time Sent
1	LKML	11/17/2015, 2:03:05 AM
2	LKML	12/2/2015, 11:40:54 AM
3	LKML	11/14/2015, 1:23:20 AM

	Commits	CommitProp
2%	1	2
2%	1	2
8%	1	2
3%	1	2
5%	1	2
0%	5	10

This year we also added a new eBook to our compliance education library

- Compliance Basics for Developers (e-Learning)
  - https://training.linuxfoundation.org/linux-courses/open-source-compliance-courses/compliance-basics-for-developers
- Open Source Compliance in the Enterprise (eBook) by Ibrahim Haddad
  - http://go.linuxfoundation.org/open-source-compliance-ebook
- **New!** Practical GPL Compliance (eBook) by Armijn Hemel and Shane Coughlan
  - https://www.linuxfoundation.org/publications/practical-gpl-compliance-download-this-free-guide-today/
  - Offers practical tips and guidance for developers and compliance engineers
- **New!** Open Source Guides for the Enterprise
  - https://www.linuxfoundation.org/resources/open-source-guides/
  - Includes guidance for compliance and open source program offices

In 2017, the Linux kernel community released its Kernel Enforcement Statement

Linux Kernel Enforcement Statement

As developers of the Linux kernel, we have a keen interest in how our software is used and how the license for our software is enforced. Compliance with the reciprocal sharing obligations of GPL-2.0 is critical to the long-term sustainability of our software and community.

Although there is a right to enforce the separate copyright interests in the contributions made to our community, we share an interest in ensuring that individual enforcement actions are conducted in a manner that benefits our community and do not have an unintended negative impact on the health and growth of our software ecosystem. In order to deter unhelpful enforcement actions, we agree that it is in the best interests of our development community to undertake the following commitment to users of the Linux kernel on behalf of ourselves and any successors to our copyright interests:

> Notwithstanding the termination provisions of the GPL-2.0, we agree that it is in the best interests of our development community to adopt the following provisions of GPL-3.0 as additional permissions under our license with respect to any non-defensive assertion of rights under the license.

https://www.kernel.org/doc/html/latest/process/kernel-enforcement-statement.html

In 2017, the Linux kernel community released its Kernel Enforcement Statement

> However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.
>
> Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.
>
> Our intent in providing these assurances is to encourage more use of the software. We want companies and individuals to use, modify and distribute this software. We want to work with users in an open and transparent way to eliminate any uncertainty about our expectations regarding compliance or enforcement that might limit adoption of our software. We view legal action as a last resort, to be initiated only when other community efforts have failed to resolve the problem.
>
> Finally, once a non-compliance issue is resolved, we hope the user will feel welcome to join us in our efforts on this project. Working together, we will be stronger.

https://www.kernel.org/doc/html/latest/process/kernel-enforcement-statement.html

Open Source License Compliance as we head into 2018

- Beginnings of ecosystem scale out
  - Moving beyond the typical companies you see in open source projects
  - Supply chains can touch thousands of companies
- Model projects, developer engagement in licensing
  - Adopting best practices as a community, in security and compliance (e.g. CII Badge, SPDX tagging)
  - Linux, Node.js, and critical projects continue to adopt SPDX identifiers, expose more developers to using them
  - New projects starting with best practices (e.g. Hyperledger)
- Community response to McHardy

Linux kernel example

8 lines (5 sloc) | 155 Bytes

```
1  /* SPDX-License-Identifier: GPL-2.0 */
2  #ifndef _LIVEPATCH_CORE_H
3  #define _LIVEPATCH_CORE_H
```

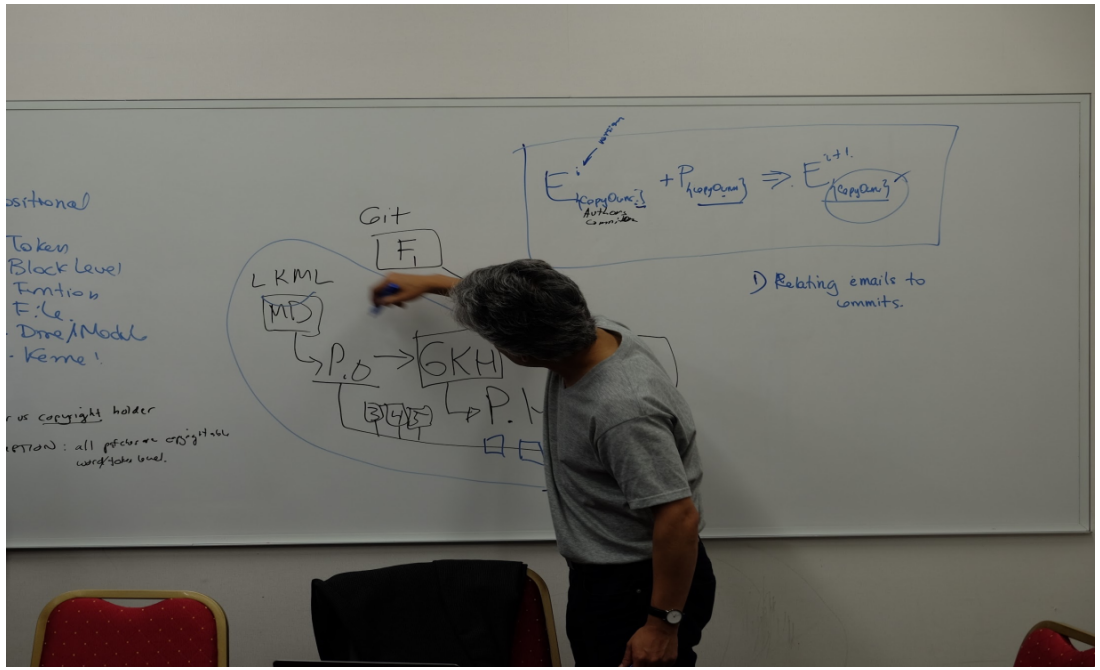
Hyperledger Fabric example

272 lines (244 sloc) | 12.9 KB

```
1  #
2  # Copyright IBM Corp. All Rights Reserved.
3  #
4  # SPDX-License-Identifier: Apache-2.0
5  #
```

Use this week to learn, improve and find ways to contribute back

- Our Summits are places to learn from others, share ideas and collaborate on how to improve
- Previous events led to the creation of OpenChain, cregit and evolved other ideas that turned into materials, tools or resources
- Join the communities working on improving compliance, adopt standards and tools internally and contribute back



Contact Us

The Linux Foundation

1 Letterman Drive
Building D, Suite D4700
San Francisco CA 94129
Phone/Fax: +1 415 7239709
www.linuxfoundation.org

General Inquiries

info@linuxfoundation.org

Membership

membership@linuxfoundation.org

Corporate Training

training@linuxfoundation.org

Event Sponsorship

sponsorships@linuxfoundation.org

THE LINUX FOUNDATION

Legal Notices

The Linux Foundation, The Linux Foundation logos, and other marks that may be used herein are owned by The Linux Foundation or its affiliated entities, and are subject to The Linux Foundation's Trademark Usage Policy at https://www.linuxfoundation.org/trademark-usage, as may be modified from time to time.

Linux is a registered trademark of Linus Torvalds. Please see the Linux Mark Institute's trademark usage page at https://lmi.linuxfoundation.org for details regarding use of this trademark.

Some marks that may be used herein are owned by projects operating as separately incorporated entities managed by The Linux Foundation, and have their own trademarks, policies and usage guidelines.

TWITTER, TWEET, RETWEET and the Twitter logo are trademarks of Twitter, Inc. or its affiliates.

Facebook and the "f" logo are trademarks of Facebook or its affiliates.

LinkedIn, the LinkedIn logo, the IN logo and InMail are registered trademarks or trademarks of LinkedIn Corporation and its affiliates in the United States and/or other countries.

YouTube and the YouTube icon are trademarks of YouTube or its affiliates.

All other trademarks are the property of their respective owners. Use of such marks herein does not represent affiliation with or authorization, sponsorship or approval by such owners unless otherwise expressly specified.

The Linux Foundation is subject to other policies, including without limitation its Privacy Policy at https://www.linuxfoundation.org/privacy and its Antitrust Policy at https://www.linuxfoundation.org/antitrust-policy. each as may be modified from time to time. More information about The Linux Foundation's policies is available at https://www.linuxfoundation.org.

Please email legal@linuxfoundation.org with any questions about The Linux Foundation's policies or the notices set forth on this slide.

THE LINUX FOUNDATION