From Request to ROI: Tracking your Company’s Investment in Open Source & Compliance

Krista Khare
David Grizzanti

Open FinTech Forum
October 11, 2018
Comcast Open Source Program Office

Opened beginning of 2017
4 person team
Support 7k+ developers
Support ~150 project repos on GitHub
Members of Open Source Advisory Council

• OSPO
• Legal
• Strategic IP/Patents
• Engineering
• Security
Open Source Contribution Request Flow

Request: Contributor submits request ticket in JIRA

Decide on Path: Filter for online review or OSAC

Consultation/Prep: Address Qs for online reviews or schedule prep call in advance of OSAC meeting

Due Diligence:
- Code review
- Security
- Dependency licenses
- Name/branding
-CLA

Contribute!
Create new project repo or give approval for 3rd-party contributions

For larger projects:
- blog announcement, tweets to generate awareness (Community Manager)
Pre-Approval Checkpoints

✓ Legal, Strategic IP – part of OSAC
✓ Security – code review part of OSAC; additional assessments as advised
✓ Compliance – dependency check and license scans
✓ Naming, branding/logo

Post-Approval Checkpoints

✓ Repo creation (Github management)
✓ Repo Linter from TODO Group
What’s the ROI?

✓ Releasing secure code
✓ Complying with license obligations
✓ Fostering better ecosystem for scanning tools, machine readability for our projects
Where engineers can help…

• Catching projects at the beginning is great, but what about projects that have been out there for years?
• How do we know which projects need attention or need help with attracting contributors?
• Looking at compliance from multiple angles helps us play catch up
Invest in review

• Invest in review as a way to improve our relationship with the OS community and be a good corporate citizen
• Make sure we have the correct licenses on existing OS projects (code of conduct, etc.)
• Make sure are internal projects are using the complying with our open source license policies
Metrics

• Which projects are we contributing to?

• Are people outside of Comcast contributing to our projects?

• Which developers are actively contributing to open source projects?

• Which Comcast projects have an active community?
Metrics

• Use to showcase our work and activity in open source to leadership
• Involvement in CHAOSS community
  - Presented our work at 2nd annual conference in Vancouver (co-located with the Open Source Summit)
Tools
Open-source developers all over the world are working on millions of projects: writing code & documentation, fixing & submitting bugs, and so forth. GH Archive is a project to record the public GitHub timeline, archive it, and make it easily accessible for further analysis.

GitHub provides 20+ event types, which range from new commits and fork events, to opening new tickets, commenting, and adding members to a project. These events are aggregated into hourly archives, which you can access with any HTTP client:

<table>
<thead>
<tr>
<th>Query</th>
<th>Command</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity for 1/1/2015 @ 3PM UTC</td>
<td>wget <a href="http://data.gharchive.org/2015-01-01-15.json.gz">http://data.gharchive.org/2015-01-01-15.json.gz</a></td>
</tr>
<tr>
<td>Activity for 1/1/2016</td>
<td>wget <a href="http://data.gharchive.org/2015-01-01-%7B0..23%7D.json.gz">http://data.gharchive.org/2015-01-01-{0..23}.json.gz</a></td>
</tr>
<tr>
<td>Activity for all of January 2015</td>
<td>wget <a href="http://data.gharchive.org/2015-01-%7B01..30%7D-%7B0..23%7D.json.gz">http://data.gharchive.org/2015-01-{01..30}-{0..23}.json.gz</a></td>
</tr>
</tbody>
</table>
GitHub archives, git, GitHub API -> Postgres databases -> Grafana dashboards  https://k8s.devstats.cncf.io/
Prometheus exporter for github metrics

50 commits  3 branches  4 releases  9 contributors  MIT
CloudWatch Event

Run every day at 1AM UTC

Get traffic insights

Lambda

Store time series values

EC2 InfluxDB
CloudWatch Event

- Run every day at 1AM UTC

Lambda

Lambda

Get traffic insights

EC2 InfluxDB

Store time series values

ECS/Fargate

- Run devstats

Commits, PRs, Author stats

RDS
CloudWatch Event

Lambda

Lambda

EC2 InfluxDB

ECS/Fargate

Commits, PRs, Author stats

RDS

Run devstats

Run every day at 1AM UTC

Get traffic insights

Store time series values

github exporter

Prometheus
Resources

- TODOgroup.org

- Repo Linter: https://github.com/todogroup/repolinter

- CHAOSS Community (Community Health Analytics Open Source Software): https://chaoss.community