Contributing to Open Source

Insights and Opportunities

Shaheem Azmal M MD @shaheemazmalmmd
Gaurav Mishra @GMishx
Kaushlendra Pratap Singh @Kaushl2208

Agenda

- **Who we are?**

- **Google Summer of Code**

- **Corporate & Open Source**

Introductionsss….

Hello everyone, I'm Shaheem Azmal..

I am a Lead Research Engineer at Siemens, with over 11 years of experience contributing to open source. As a maintainer, contributor, mentor, and technical lead for FOSSology project, I am passionate about fostering collaboration and innovation within the community. My work focuses on driving research and development while supporting and guiding new contributors.

Hey, this is Gaurav Mishra..

By day, I'm a Research Engineer at Siemens, and by night, I'm an Open-Source superhero! 🦸 Over the past 7 years, I've juggled roles like maintainer, community leader, owner, and mentor. My toolkit includes DevOps, Software Architecture, license compliance, cloud computing, GenAI, and Semantics. I'm here to share my adventures and geek out with you all!

Hii, I'm Kaushl.

I've been working as a Research Engineer at Siemens for the past three years, and I have been an active contributor to open source for over 6 years. My journey includes roles as a contributor, mentor, maintainer, and open-source advocate. I am passionate about driving innovation through open source. My toolkit includes DevOps, Software Architecture, Machine Learning & Artificial Intelligence, Generative-AI, and Semantics Excited to share my experiences and learn from this vibrant community!

fossology

Let's Dive Into → FOSSology

License compliance

Legal advice!

Example
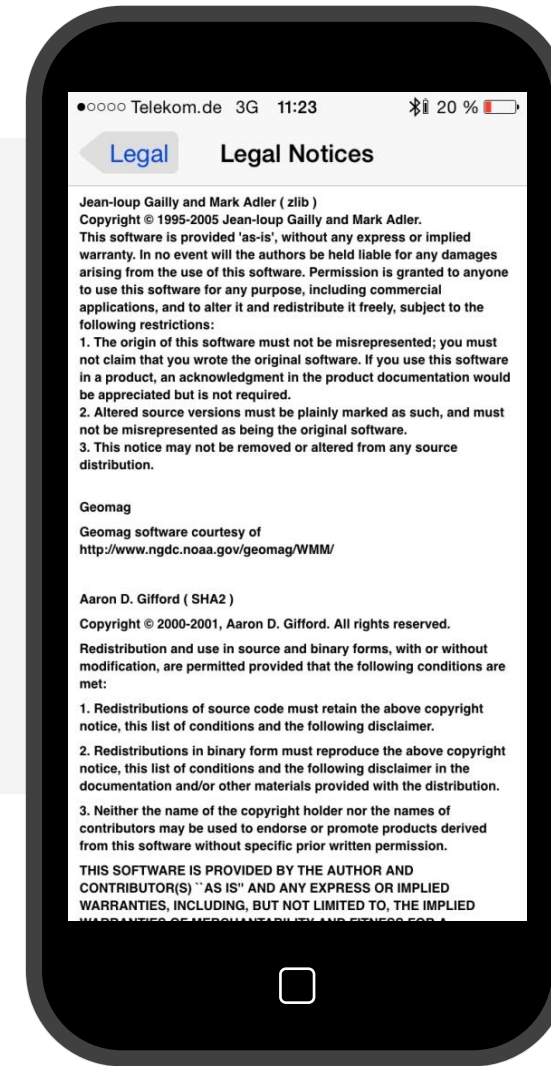
Example

No offense though!

Software's Perspective:

You know these examples
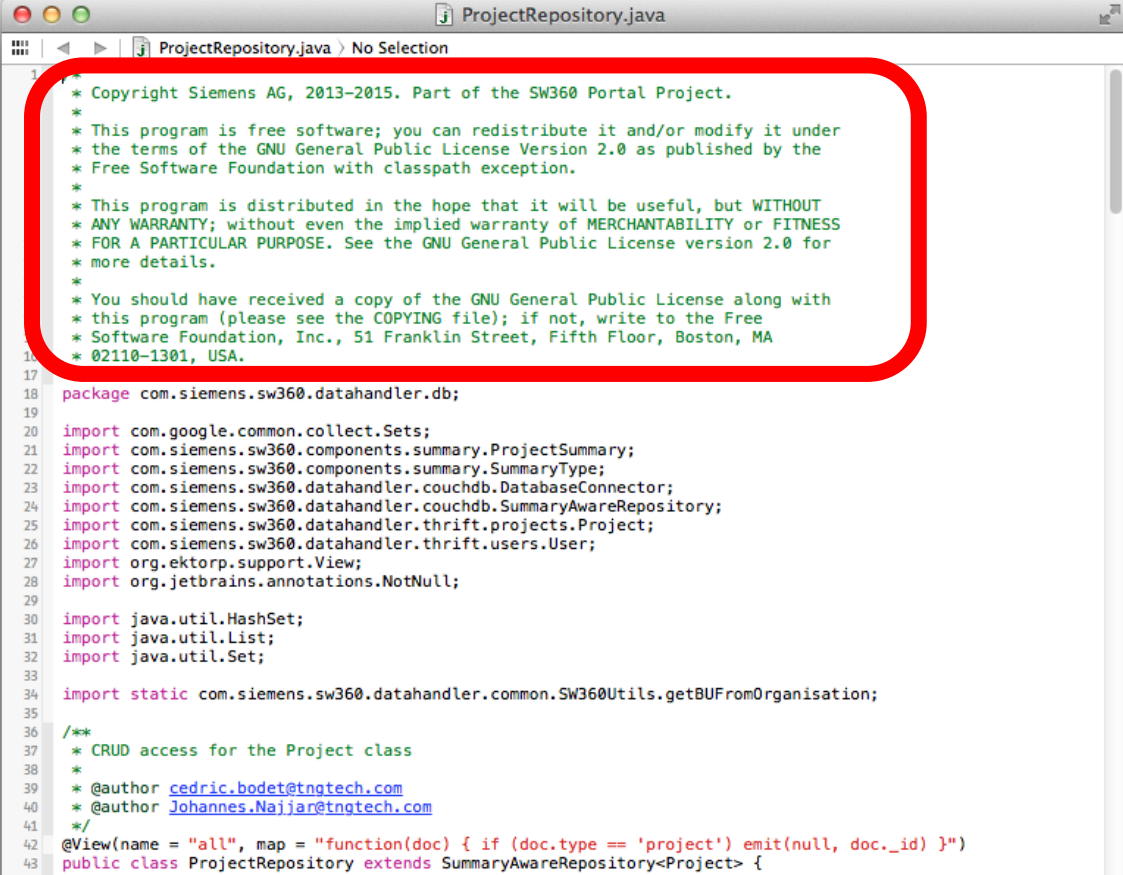
Distributing open-source software requires to

- Provide licenses of involved software

- Provide copyright statements of involved authors

- Provide disclaimers

- … and much more

It is about finding licenses

Finding Licenses

- License texts

- References to licenses

- Written texts explaining licensing

- License relevant statements

What is FOSSology?

A Web server application for license and copyright compliance of software components.

FOSSology Project https://www.fossology.org/	FOSSology Development https://www.github.com/fossology/fossology

▪ Published first in 2008, GPL-2.0
 	
▪ 2015: Linux Foundation collaboration project
 	
▪ Web server based and command line and REST interfaces
 	
▪ Scanning agents searching for license and copyright relevant hits (and more …)
 	
▪ A multi-user / multi-tenant Web UI for review organizing clearing job
	▪ Standard application stack:
	  ▪ Linux, Apache 2, PostgreSQL, PHP,
	  ▪ jQuery, Python, Go, C / C++.
	
	▪ Web-based UI & REST API in PHP, but most scanners written in C / C++
	
	▪ [Basic Workflow]
	
	▪ [Good first issues]
	
	▪ [Slack Channel]

How does FOSSology work?

Upload Component

Agents Scanning

Review Results

Generate Reporting

Pass Report to Client
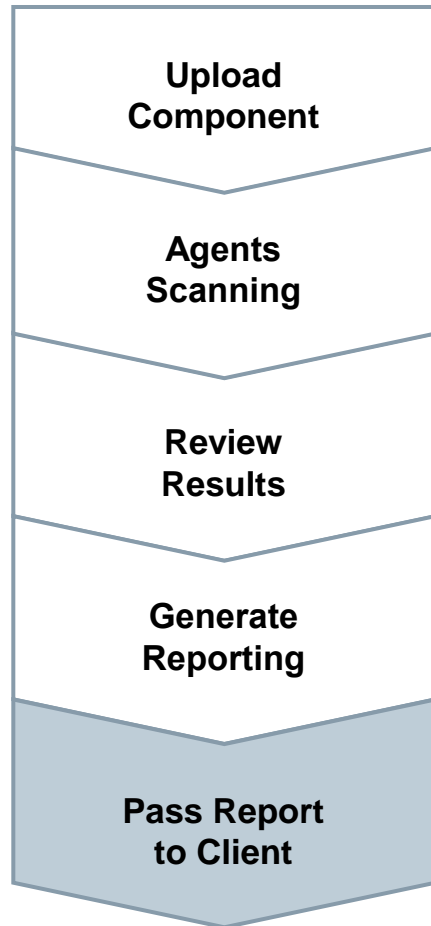
- Uploading source code archive (*.zip, *.tar.gz, etc)

- Agents scan for license relevant text (Nomos, Monk etc.)
- Copyrights, Export Control (ECC), your keywords to look for etc.

- Review scanner results for wrong license classification
- Review other scanner findings (copyrights, ECC)

- Result of the "clearing"
  - SPDX reporting
  - Generated notice or readme file
  - Debian-copyright
  - CycloneDX Reporting

fossology

FOSSology Feature Overview

A Web server application for license and copyright compliance of software components.

License Scan features	Other features

▪ Regular expression scanner
▪ Text similarity scanner
▪ License (text) management
▪ Aggregation of licenses in hierarchical view
▪ License histogram
▪ Supporting concluded vs. found license
▪ Bulk processing of files with same licensing
▪ Reusing of license conclusions
	▪ Copyright, authorship statements scanner
	▪ Export control and customs scanner
	▪ Command line & REST interfaces
	▪ Reporting
	  ▪ SPDX RDF and tag-value
	  ▪ Debian-copyright
	  ▪ Plain text output
	  ▪ CycloneDX Reporting and more..
	▪ User, group and upload management

Additions to FOSSology

- Newer agents like OJO and Compatibility agent

- Integrations with other projects like ClearlyDefined.io, ScanCode, ScanOSS

- Support for more data exchange formats like CLIXML, CycloneDx, Decision dump JSON

- New and Improved REST API v2

- Addition of projects like Safaa, LicenseDB, Minerva Dataset

- Many More….

Let's Jump to → Google Summer of Code

What is Google Summer of Code?

GSoC
is an international annual program in which Google awards stipends to students who successfully complete a software coding project for an open-source organization during the summer.

20K+	116	45M+
New Contributors	Countries	Lines of Code
1000+	19K+	19
Open Source Organizations	Mentors	Years

Impact:

• **Global Reach:** Over 20,000 students from 116 countries have participated in GSoC since its inception.

• **Open-Source Contributions:** Many projects initiated during GSoC have grown into key components of widely-used open-source software.

GSoC Timeline: Organization's POV

Gathering ideas

- Involvement of community.
- Using GitHub Discussions

Finalizing

Applying for GSoC

- Preparation of application
- Dedicated Space in slack

Proposals

Preparations

Application

Idea selection for GSoC

- Filtering idea
- Moving to Wiki
- Labelling issues
- Assigning Mentors

Collaboration on selection

- Gathering Proposals
- Reviewing Proposals
- Slack, email, GitHub issue

fossology

Perspective:

For Students	For Mentors
▪ Experience in software development ▪ Participation in Open-Source initiatives ▪ Proficiency in distributed work environments ▪ Internship experience ▪ Google stipend recipient	▪ Positive visibility ▪ Opportunity to meet new Students ▪ Opportunity to work on New Tech ▪ To help extend the OSS community ▪ Experience distributed collaboration

fossology

Let's talk Trends For FOSSology:
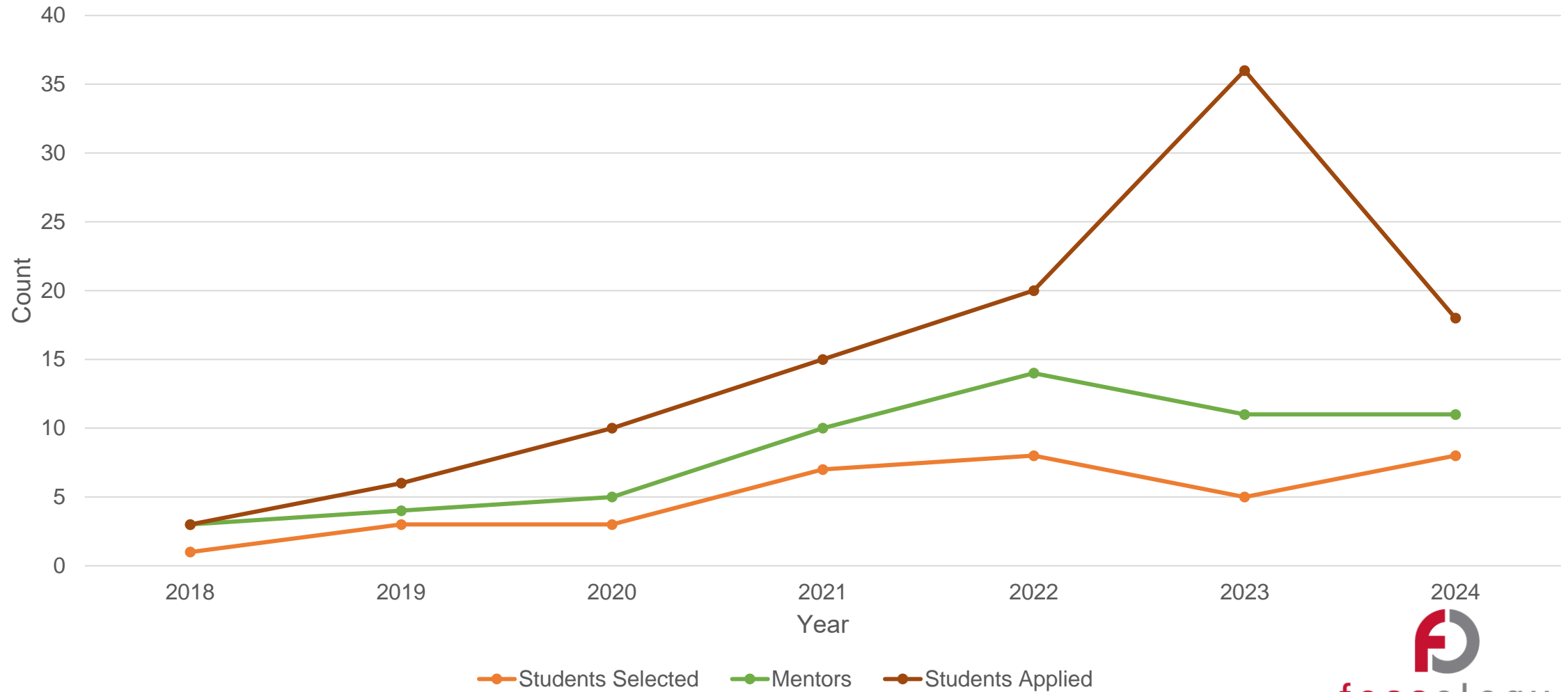
Atarashi

A Step towards non-rule based standalone command line scanner… (https://github.com/fossology/atarashi)

Different methods for scanning license statements

- Unlike rule-based approaches, like Nomos, Atarashi implements multiple text statistics and information retrieval algorithms.

Distance finding algorithms

- Word Frequency Similarity
- Term frequency-inverse document frequency (tf-idf)
- Damerau–Levenshtein distance
- N-grams

Similarity finding algorithms

- Score Similarity
- Cosine Similarity
- Dice Similarity
- Bi-gram Cosine Similarity

Safaa : Copyright False Positive Detection

Designed for handling false positive detection in copyright notices. Additionally, it can declutter copyright notices, removing unnecessary extra text.… (https://github.com/fossology/Safaa)

Features

Captured	Cleared
copyright, i.e., "Copyright (c) 1995-2001 Corporation for National Research Initiatives; All Rights Reserved" are retained in Python 1.6.1 alone or in any derivative version p…	Copyright (c) 1995-2001 Corporation for National Research Initiatives
© cessibles sur l'« Information », il les cède au « Réutilisateur » de façon non exclusive,…	False+ve

- Load pre-trained models or train your own.
- Integration with scikit-learn for training and prediction.
- Integrated with spaCy for named entity recognition and decluttering tasks.
- Preprocessing tools to ensure data consistency and quality.
- Ability to handle local or default model directories.

fossology

ScanCode : Integration with other community

Bringing copyright and license data from other scanners… (https://github.com/nexB/scancode-toolkit)

Integrating ScanCode to FOSSology

Nomos and **Monk** are the two leading scanners FOSSology uses for license detection and Copyright for scanning `copyright, URL, emails`, and `holders` name. FOSSology approach is to detect licenses with either a large (large: 2500 regexes) dataset of regex patterns (nomos) or a full string comparison against license full texts (large: ~400 texts) (monk). **Atarashi** license scanner implements multiple text statistics and information retrieval algorithms.

ScanCode Toolkit is a very established license scanner similar to Nomos or Monk. It is a simple python based command line scanner that runs on Windows, Linux, and Mac. It implements a number of different approaches and integrates these into one application for identifying and classifying license-relevant content in packages.

The **basic idea** is to use the command line interface from the ScanCode package in order to be called right from the FOSSology application. FOSSology will pass a single file and take the result from the ScanCode command line call. Scan result will include license name, the SPDX key, Score, Copyright and Holder name, Emails, and URLs present in the given code and as requested by the user.

Spasht : Sharing clearing data

Spasht allows users to share clearing information with community with help of ClearlyDefined.io project. The aim of ClearlyDefined is to provide clear license and compliance information in public domain. FOSSology excels in providing this information.

Post GSoC

- Student continue contributing to further issues

- Helping new joiners in community

- Bi-weekly meetings for interested students to discuss the community progress.

How are Students helping us post GSoC?

- Messenger of FOSSology

- Mentoring interested students

- And continuous collaboration

Post GSoC

- Student continue contributing to further issues

- Helping new joiners in community

- Bi-weekly meetings for interested students to discuss

How are Students helping us post GSoC?

- Messenger of FOSSology

- Mentoring interested students

- And continuous collaboration

Brainless Coders

KICKSTART TO
OPEN SOURCE
WITH YOUR FIRST
SUCCESSFUL **PR**

Discussion on **GSoC**, Open Source and Hacktoberfest

Register at -> bit.ly/kickstartopensource

For any queries
7663016721

7-8 PM 8 OCT 2020

Ayush Bhardwaj
GSOC 2019,2020, Fossology

Kaushlendra Pratap
GSOC 2020, Fossology

Sahil Jha
GSOC 2020, NRNB

May the source be with you

Post GSoC

- Student continue contributing to further issues

- Helping new joiners in community

- Bi-weekly meetings for interested students to discuss

How are Students helping us post G

- Messenger of FOSSology

- Mentoring interested students

- And continuous collaboration

7-8 PM 8 OCT 2020

Ayush Bhardwaj
GSOC 2019,2020, Fossology

Brainless Coders

KICKSTART TO
OPEN SOURCE

Google Summer of Code @ fossology

Integrating ScanCode Toolkit to FOSSology

Thumbnail from: Integrating ScanCode Toolkit by Sarita Singh

fossology

Po

- S... ...ssues
- H...
- E... ...ts to discuss...

How

- M...
- M...
- A...

GLUG NITH PRESENTS
in collaboration with
ROBOSOC, CSEC, ISTE

Contributing to Open Source Projects

What to do after choosing an organization?

How to start contributing?

Participation in Open Source Communities.

Good Contribution Tips.

AVINAL KUMAR
GSoC'21 @FOSSology
GSoD'20 @VideoLAN

ABHAY RAJ SINGH
GSoC'20 @ Inkscape

16TH OCTOBER 2021
4 PM IST

LIVE
BIT.LY/OSM-E3

OPEN SOURCE MONTH

PARTH PANT
GSoC'21 @ Inkscape

Brainless Coders

KICKSTART TO OPEN SOURCE WITH YOUR FIRST

7-8 PM 8 OCT 2020

Ayush Bhardwaj
GSOC 2019,2020, Fossology

...ndra Pratap
..., Fossology

...a
..., NRNB

Google Summer of Code @ fossology

Integrating ScanCode Toolkit to FOSSology

fossology

Corporate ❤️ Open Source

Open Source @ Siemens

SIEMENS Home   Manifesto   Team   Projects   Memberships   Events   Talks

Search 🔍

The Siemens Open Source Manifesto

Siemens has resolved to be an active and reliable member in the Open Source ecosystem. We aim to benefit our customers, employees, and the company, while contributing to the wider community. We recognize the central role of Open Source in promoting software reuse in our growing digital world, which is an increasingly important contribution in our effort to responsibly handle resources.

To ensure we can achieve and maintain our active participation, we have decided to follow these guiding principles in our commitment to Open Source:

> Are you a Siemens employee and want to learn how the Manifesto is embedded in our overall Open Source Strategy?
>
> **Learn more** 📖

Company Principles

1. **Encourage Open Source:** Siemens will support and empower its employees to use, contribute to, and create Open Source projects.
2. **Facilitate Participation:** Siemens will drive and foster active engagement in Open Source activities relevant to our company's projects.
3. **Grow Through Exposure:** Siemens will encourage and facilitate learning and advancement of its employees through Open Source activities.
4. **Promote Visibility:** Siemens will continuously improve and maintain a positive reputation in the Open Source community.

Employee Principles

1. **Prefer Open Source:** Employees will look for suitable Open Source solutions before writing or using custom code.
2. **Contribute:** Employees will favor contributing to Open Source projects over local changes within the scope of their day-to-day work.
3. **Foster Teamwork:** Employees will collaborate and support each other in driving Open Source efforts.
4. **Be Responsible Citizens:** Employees will always act responsibly in the Open Source community, with care and respect in both content and communication.

fossology

Open Source @ Siemens

SIEMENS Home   Manifesto   Team   Projects   Memberships   Events   Talks

Search

Upstream First ¶

Using and contributing to open source projects helps us provide more reliable products and develop faster. The Linux kernel, Debian, Yocto, coreboot, U-Boot, SWUpdate, and many other projects, have been a large part of the Siemens embedded story - and we want to give back!

This applies not just to our products, but the tools we use to develop them. Siemens is one of the largest contributors to GitLab, which powers our self-hosted social coding platform.

How Siemens created an open source DevOps culture with GitLab

6.4M+
Builds per month (>210,000 builds per day)

0
Support requests to GitLab

4+
Production deployments a month

Sharing Culture

Driving an Inner Source culture at Siemens helps us become better Open Source citizens.

This means following best practices from the open source world, and enabling both humans and software to work together seamlessly - with API-first thinking and development in the open.

As we do this, we can also more effectively collaborate with the larger community and bring more software to the open source world.

Open Standards

Siemens software and products are present across the world, and stay in the field for years or decades, from industrial IoT to large infrastructure projects. They also have to interact with a number of different technologies.

https://opensource.siemens.com We want to make them work together seamlessly by adopting and shaping open standards.

CIVIL INFRASTRUCTURE

Open Source @ Siemens

SIEMENS Home   Manifesto   Team   Projects   Memberships   Events   Talks

Search

Upstream First ¶

Using and contributing to open source projects helps us provide more reliable products and develop faster. The Linux kernel, Debian, Yocto, coreboot, U-Boot, SWUpdate, and many other projects, have been a large part of the Siemens embedded story - and we want to give back!

This applies not just to our products, but the tools we use to develop them. Siemens is one of the largest contributors to GitLab, which powers our self-hosted social coding platform.

How Siemens created an open source DevOps culture with GitLab

6.4M+
Builds per month (>210,000 builds per day)

0
Support requests to GitLab

4+
Production deployments a month

Sharing Culture

Driving an Inner Source culture at Siemens helps us become better Open Source citizens.

This means following best practices from the open source world, and enabling both humans and software to work together seamlessly - with API-first thinking and development in the open.

As we do this, we can also more effectively collaborate with the larger community and bring more software to the open source world.

Sponsor at multiple level:

https://opensource.siemens.com/memberships/

Open Standards

Siemens software and products are present across the world, and stay in the field for years or decades, from industrial IoT to large infrastructure projects. They also have to interact with a number of different technologies.

https://opensource.siemens.com We want to make them work together seamlessly by adopting and shaping open standards.

CIVIL INFRASTRUCTURE

SW360: Corporate collaboration

- SW360 is a software catalogue which helps companies manage their compliance data.

- SW360 is great example of collaboration between multiple corporations like Siemens, Cariad, Toshiba, Bosch, etc.

- Companies develop the tool together and share their knowledge across the board to help everyone grow.

- https://eclipse.dev/sw360/

Thank you for your attention!

© 2016-2024 Siemens AG, The Linux Foundation

CC-BY-SA 4.0
https://creativecommons.org/licenses/by-sa/4.0/

Internet
https://www.fossology.org

GitHub
https://github.com/fossology/fossology

Further Links
https://www.spdx.org
https://github.com/siemens
https://github.com/eclipse-sw360

Contact :

FOSSology Mailing list

- fossology@fossology.org

Email us
- shaheem.azmal@siemens.com
- mishra.gaurav@siemens.com
- kaushlendra-pratap.singh@siemens.com

fossology