



ClearlyDefined

Community Meeting - May 4, 2026



MAY THE \_\_\_\_\_  
**SOURCE**  
\_\_\_\_\_ BE WITH  
**YOU**



# Agenda

- Introducing Adam (AboutCode) and Duane (OSI ED)
- Upcoming events
  - Open Source Summit NA 2026
  - FOSSY 2026
- Google Summer of Code 2026
  - Two proposals accepted
- Tech governance
  - Approve PR to cover permissions below "maintainer" and expand maintainer expectations
- Roadmap 2026/2027/2028
  - Blog post / newsletter



# Introducing Adam & Duane

- Adam Herzog (AboutCode)
- Duane O'Brien (OSI ED)
  - [Hello From The New Executive Director](#)



# Open Source Summit NA 2026

- [Beyond SBOMs: Making License Data Actionable With ClearlyDefined - Jamie Magee, Microsoft](#)





## Upcoming events

- CFPs Open
  - [FOSSY 2026](#)
    - August 6th – 9th 2026
    - University of British Columbia, Vancouver, Canada
    - CFP deadline: May 22, 2026

**FOSSY 2026** [About](#) ▾ [Attend](#) ▾ [Program](#) ▾ [Sponsorship](#)

# FOSSY 2026

**August 6th – 9th 2026 — University of British  
Columbia, Canada**

The fourth Free and Open Source Software Yearly conference



# Google Summer of Code 2026

- Almost 100 proposals received for AboutCode / ScanCode
- Proposals reviews led by Ayan Mahapatra
- [Two proposals accepted](#)
  - [VulnerableCode Insights](#) - Sampurna Pyne (Samk)
  - [ScanCode-Toolkit : Mark required phrases for rules automatically using NLP/AI](#) - Kaushik Kumar



Google  
Summer of Code

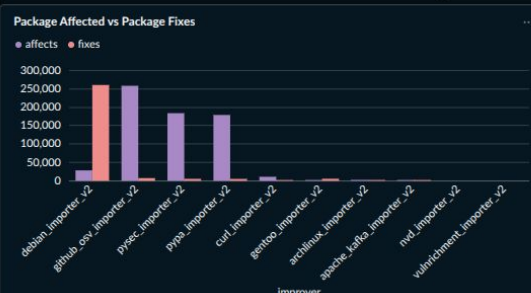
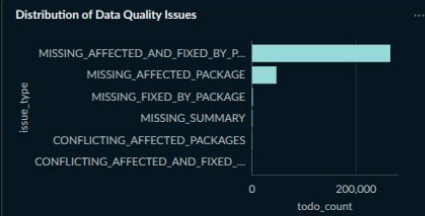
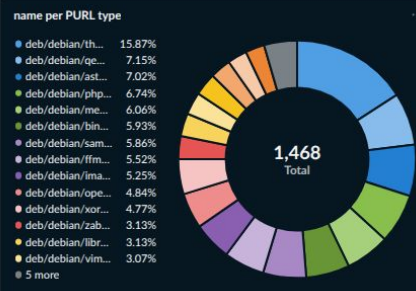
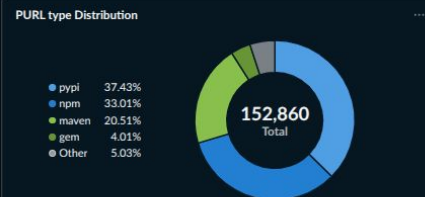
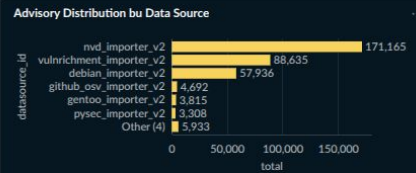


# GSoc 2026

## ● VulnerableCode Insights

Sampurna Pyne (Samk)

### PoC Vulnerable Insights





# GSoC 2026

- [ScanCode-Toolkit:](#)  
[Mark required phrases for rules automatically using NLP/AI](#)

Kaushik Kumar

## Scancode Required Phrase Extractor

by Kaushik Kumar

predicts the required phrase boundary in a `.RULE` file to prevent false positive license detections. made with a finetuned DeBERTa-v3-large model

Try an example:

LGPL-2 · GPL-2 (dual)

OLDAP-2.5

MIT Notice

LGPL (HTML rule)

GPL · LGPL · MPL  
(triple)

SPDX Tag

LGPL (OCaml  
comment)

Ambiguous

Paste rule text here...

Note: Paste only the plain text body. Exclude YAML frontmatter and the `---` separator.

Predict Required Phrase



# Tech Governance

- The Charter covers the steering committee and maintainers
- [PR](#) to cover adding permissions below "maintainer" and expand a bit on maintainer expectations.
  - tech-governance/README.md
  - tech-governance/CONTRIBUTORS.md
  - tech-governance/MAINTAINERS.md



# Roadmap (2026, 2027, 2028)

- Roadmap published at OSI blog and newsletter



AboutCode



## ClearlyDefined: A Three-Year Roadmap for Sustainability and Growth



# Roadmap (2026, 2027, 2028)

## Key points:

- Combine ClearlyDefined and AboutCode operations and backends when possible to enable scan data reuse and avoid duplicated, wasted compute resources across ecosystems
- Refactor ClearlyDefined backend and frontend to reduce the technical debts and increase its reach
- Enroll new sponsors to share the cloud compute costs
- Unlock data access for increased community adoption
- Enable new use cases around security and project health



# Roadmap (2026, 2027, 2028)

## Themes:

- 2026
  - Reduce technical debt and operating costs
  - Improve the curation and harvesting queue
  - Improve the UI to facilitate community contributions
- 2027
  - Migrate to a federated data architecture to further reduce hosting costs and unlock data access.
  - Initiate work on new use cases, vulnerabilities and project health and risks
- 2028
  - Deploy security and project health use cases



# Roadmap (2026, 2027, 2028)

## 1) Create better license and package scans, faster and more efficiently

- Enable reusing scans across ClearlyDefined harvests and AboutCode data collections
- Design a priority mechanism to scan packages based on actual usage
- Improve the performance of ScanCode for faster and more cost efficient scans
- Improve ScanCode license detection based from the analysis of ClearlyDefined curated data
- Adopt federated data design with data mirrored across multiple hosts to improve service availability and reduce costs, reusing existing scans



# Roadmap (2026, 2027, 2028)

## 2) Reduce technical debt

- Update dependencies to latest versions
- Refactor storage of scans and attachments
- Switch to the open source DocumentDB from the proprietary MongoDB
- Maintain and improve the ClearlyDefined code base, CI/CD, and infrastructure
- Update project documentation to reflect new and ongoing development
- Facilitate outreach and onboarding for new contributors and sponsors



## Roadmap (2026, 2027, 2028)

### 3) Shared scans for shared costs

- Expand the pool of organizations hosting scans to share the scanning costs
- Design new prioritized scan queue to scan packages in an order that aligns with business needs
- Integrate distributed scanning with federated data access with on-premises, private data usage
- This also includes direct collaboration with open source foundations, and large projects such as Linux distributions to help them scan and share scans of their upstream and downstream packages.



## Roadmap (2026, 2027, 2028)

### 4) Integrate PURL and SBOM for easier interoperability

- Create new API endpoints to access ClearlyDefined data keyed by PURL
- Create a new SBOM API endpoint to expose reference package data in SBOM formats
- Adopt PURL as a core identifier in the database and file store



## Roadmap (2026, 2027, 2028)

### 5) Improve curation processes and user experience

- Develop and deploy a new and improved UI for efficient data curation
- Make it easier for organization to contribute their curations back in bulk
- Improve the data models for curation storage
- Create a new specification for multi-stakeholder data curations exchange



# Roadmap (2026, 2027, 2028)

## 6) Support for more use cases

- Security with vulnerability data
- Data for project health and lifecycle events and analytics
- Usage to find the most used open source components and drive scan priorities



ClearlyDefined

Thank you