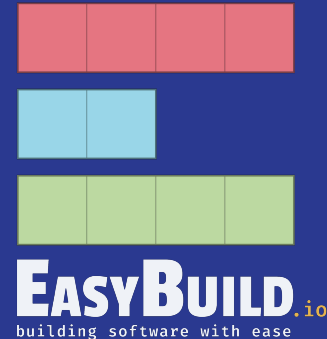


# Ecosyste.ms

## Exploring Open Source Software Landscapes

Andrew Nesbitt



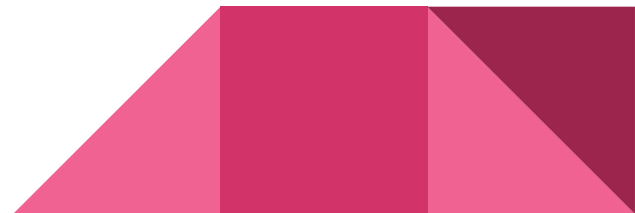
# About me

Open Source Software Developer

Package Management Enthusiast

Based in Somerset, United Kingdom

- Website: <https://nesbitt.io>
- GitHub: <https://github.com/andrew>
- Email: [andrew@ecosyste.ms](mailto:andrew@ecosyste.ms)
- Mastodon: <https://mastodon.social/@andrewnez>



# About my dogs



# Agenda

- What is Ecosyste.ms?
- Ecosyste.ms Services
- Case Studies
- Roadmap
- Questions



# Agenda

- **What is Ecosyste.ms?**
- Ecosyste.ms Services
- Case Studies
- Roadmap
- Questions



# Agenda

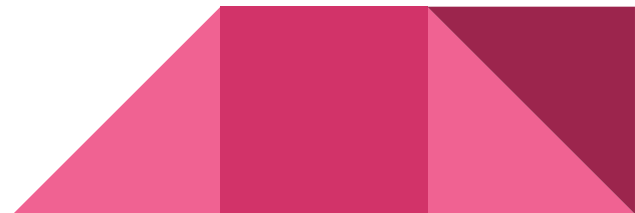
- **Why is Ecosyste.ms?**
- Ecosyste.ms Services
- Case Studies
- Roadmap
- Questions



# Exploring Open Source Software Ecosystems

There are all kinds of reasons to analyse open source

- Studying Open Source Software communities
- Comparing, sorting and categorizing OSS projects
- Comparing trends across software ecosystems
- Discovering interesting, critical or unusual projects
- Investigating security issues and trends
- Recognizing important maintenance work
- Finding and supporting overworked maintainers
- Enabling Data Based Decision Making
- Help make OSS Software Better



# Challenges in collecting OSS metadata

- Disparate sources of metadata spread across many services
- Many different data formats
- Different ecosystem registries expose different kinds of APIs
- Variety of rate limits and restrictions on keeping up to date
- Huge amounts of data
- PII compliance issues
- Spam, malicious code and other unwanted noise
- Diminishing returns for smaller software ecosystems



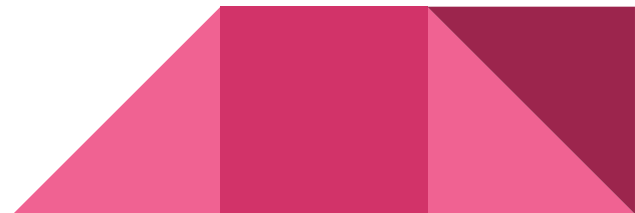


# Introducing Ecosyste.ms

Tools and open datasets to support, sustain, and secure critical digital infrastructure.

- Package manager metadata for 34 different software ecosystems
- Source Repository metadata from 785 different forges
- Issues, pull requests, commits and security advisory datasets
- Tools and APIs for analysing, parsing, diffing and scanning OSS
- Normalized data across many ecosystems and platforms
- Mining dependency graphs from packages, repos and containers
- All open source (AGPL) and open data (CC-BY-SA)

- Website: <https://ecosyste.ms>
- Code: <https://github.com/ecosyste-ms>



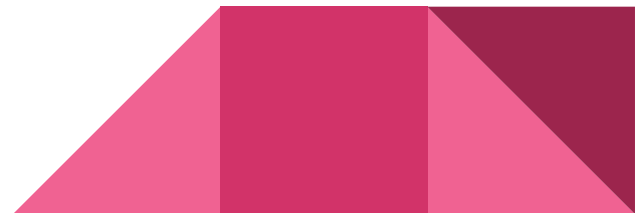
# Ecosyste.ms - The Numbers

- 9 million software packages
- 100 million package versions
- 200 million public software repositories
- 16 billion dependencies
- 27 million issues and pull requests
- 345 million commits
- 8 billion activity events
- 17 thousand security advisories
- 450 thousand docker image SBOMs
- 12TB of data in Postgres (~1TB indexes)
- 300 million API requests per month



# Ecosyste.ms - What can you do with it?

- Find Critical packages within an ecosystem
- Explore unseen infrastructure
- Discover key maintainers
- Look at cross-ecosystem dependency graphs
- Large scale analysis of software communities
- Connect with other kinds of data
  - Scientific papers
  - Funding data
  - Software Foundations
  - And more!



# Agenda

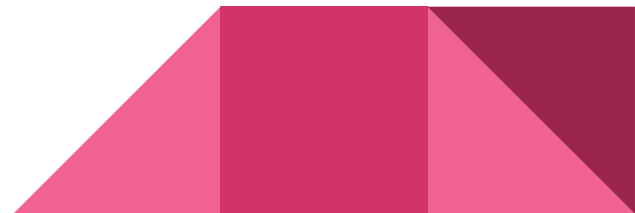
- What is Ecosyste.ms?
- **Ecosyste.ms Services**
- Case Studies
- Roadmap
- Questions



# Ecosyste.ms Services

Individual services for parsing, normalizing and aggregating OSS metadata

- Packages
- Timeline
- Parser
- Archives
- Digest
- Diff
- Licenses
- Repos
- Open Collective
- SBOM
- Resolve
- Advisories
- Commits
- Docker
- Summary
- Issues
- OST
- Papers
- Awesome



# Ecosyste.ms Services: Packages

Normalized package manager metadata from many software ecosystems

- 34 software ecosystems
  - 59 package manager registries
  - 9.5 million packages
  - 101 million versions
  - 1.2 billion dependencies
- 
- Website: <https://packages.ecosyste.ms>
  - Code: <https://github.com/ecosyste-ms/packages>



# Ecosyste.ms Services: Parser

Web service to parse dependency metadata from manifest files

98 file types supported from 30 different software ecosystems:



\*.cabal, \*.csproj, \*.gemspec, \*.nuspec, \*.podspec, \*.podspec.json, .github/workflows/\*.yaml, .github/workflows/\*.yml, Brewfile, Brewfile.lock.json, Cargo.lock, Cargo.toml, Cartfile, Cartfile.private, Cartfile.resolved, DESCRIPTION, Dockerfile, Gemfile, Gemfile.lock, Godeps, Godeps/Godeps.json, Gopkg.lock, Gopkg.toml, META.json, META.yml, Package.resolved, Package.swift, Pipfile, Pipfile.lock, Podfile, Podfile.lock, Project.json, Project.lock.json, REQUIRE, action.yaml, action.yml, bower.json, build.gradle, build.gradle.kts, cabal.config, composer.json, composer.lock, cyclonedx.json, cyclonedx.xml, docker-compose.yml, dub.json, dub.sdl, elm-package.json, elm-stuff/exact-dependencies.json, elm\_dependencies.json, environment.yaml, environment.yaml.lock, environment.yml, environment.yml.lock, gems.locked, gems.rb, glide.lock, glide.yaml, go-resolved-dependencies.json, go.mod, go.sum, gradle-dependencies-q.txt, haxelib.json, ivy.xml, maven-dependency-tree.txt, maven-resolved-dependencies.txt, mix.exs, mix.lock, npm-ls.json, npm-shrinkwrap.json, package-lock.json, package.json, packages.config, packages.lock.json, paket.lock, pip-resolved-dependencies.txt, pnpm-lock.yaml, poetry.lock, pom.xml, project.assets.json, project.clj, pubspec.lock, pubspec.yaml, pyproject.toml, req\*.pip, req\*.txt, requirements.frozen, requirements/\*.pip, requirements/\*.txt, sbt-update-full.txt, setup.py, shard.lock, shard.yml, vcpkg.json, vendor/manifest, vendor/vendor.json, versions.json, yarn.lock

- Website: <https://parser.ecosyste.ms>
- Code: <https://github.com/ecosyste-ms/parser>

# Ecosyste.ms Services: Repos

Repository metadata from a variety of software forges such as GitHub, GitLab, BitBucket, Codeberg, Gitea and Forgejo instances.

- 785 forges
  - 205 million repositories
  - 195 million tags
  - 236 million manifest files
  - 17 billion dependencies
- 
- Website: <https://repos.ecosyste.ms>
  - Code: <https://github.com/ecosyste-ms/repos>

 <b>github</b> 1 instance - 198,947,992 repositories
<a href="#">github</a> 198,947,992 repositories
 <b>gitlab</b> 12 instances - 4,641,345 repositories
<a href="#">gitlab.com</a> 4,512,565 repositories
<a href="#">salsa.debian.org</a> 66,869 repositories
<a href="#">framagit.org</a> 31,793 repositories
<a href="#">gitlab.freedesktop.org</a> 13,534 repositories
<a href="#">invent.kde.org</a> 10,975 repositories
<a href="#">gitlab.alpinelinux.org</a> 1,867 repositories
<a href="#">gitlab.haskell.org</a> 1,276 repositories
<a href="#">git.pleroma.social</a> 1,014 repositories
<a href="#">gitlab.xfce.org</a> 830 repositories
<a href="#">git.drupalcode.org</a> 399 repositories
<a href="#">gitlab.arm.com</a> 181 repositories
<a href="#">gitlab.gentoo.org</a> 42 repositories
 <b>bitbucket</b> 1 instance - 1,704,869 repositories
<a href="#">bitbucket.org</a> 1,704,869 repositories
 <b>forgejo</b> 176 instances - 42,332 repositories
<a href="#">codeberg.org</a> 33,848 repositories
<a href="#">git.disroot.org</a> 291 repositories



# Ecosyste.ms Services: Advisories

Security Advisory metadata connecting packages and repositories

- 17,500 advisories
- 12 ecosystems
- 8,150 affected packages
- 500,000+ affected versions
- 1,000,000+ affected open source repositories
  
- Website: <https://advisories.ecosyste.ms>
- Code: <https://github.com/ecosyste-ms/advisories>

Filter by Severity	
Moderate	7,509
High	6,132
Critical	2,699
Low	961

Filter by Ecosystem	
maven	5,501
pypi	3,537
npm	3,519
packagist	3,218
go	1,835
nuget	1,388
rubygems	812
cargo	772
swift	31
hex	29
actions	16
pub	8

# Ecosyste.ms Services: Issues

## Issue and Pull Request metadata aggregated

- 789 forges
- 2.8 million repositories indexed
- 12 million issues
- 26 million pull requests
- 71 million comments
- 3.2 million authors
- 26% of all issues and pull requests created by bots

- Website: <https://issues.ecosyste.ms>
- Code: <https://github.com/ecosyste-ms/issues>

### [electron/electron](#)

Total issues: 20,799

Total pull requests: 22,673

Last synced: about 23 hours ago

### [microsoft/vscode](#)

Total issues: 17,901

Total pull requests: 4,513

Last synced: 1 day ago

### [AleoHQ/leo](#)

Total issues: 17,095

Total pull requests: 938

Last synced: 1 day ago

### [WordPress/gutenberg](#)

Total issues: 11,833

Total pull requests: 12,016

Last synced: 4 days ago

### [rust-lang/rust](#)

Total issues: 11,393

Total pull requests: 16,365

Last synced: 1 day ago

# Ecosyste.ms Services: Commits

Commit metadata aggregated and summarized

- 789 forges
- 1.4 million repositories indexed
- 345 million commits counted
- 6.2% commits authored by a bot
- Average 223 commits per repository
- Average 9.6 committers per repository
  
- Website: <https://commits.ecosyste.ms>
- Code: <https://github.com/ecosyste-ms/commits>

## [CocoaPods/Specs](#)

Branch: master

Total Commits: 798,446

Total Committers: 46,278

Avg Commits per committer: 17.253

Development Distribution Score (DDS): 0.991

Last synced: 5 days ago

## [CocoaPods/CDN-Specs](#)

Branch: master

Total Commits: 340,206

Total Committers: 27,658

Avg Commits per committer: 12.3

Development Distribution Score (DDS): 0.995

Last synced: 5 days ago

## [ygit/Specs](#)

Branch: master

Total Commits: 231,844

Total Committers: 20,619

Avg Commits per committer: 11.244

Development Distribution Score (DDS): 0.994

Last synced: 4 days ago

# Ecosyste.ms Services: Docker

Index of dependencies inside public docker images using [syft](#) to create SBOMs of each image.

- 450,000 docker images indexed
  - 324 Billion downloads
  - 983,000 unique dependencies from 27 ecosystems
  - 130 Million dependencies
  - Includes system dependency usage metrics
- 
- Website: <https://docker.ecosyste.ms>
  - Code: <https://github.com/ecosyste-ms/docker>

## [debian/base-files](#)

95,725 dependent images - 104,680,610,177 total downloads

## [debian/libcrypt20](#)

91,336 dependent images - 101,465,021,528 total downloads

## [debian/libss2](#)

90,102 dependent images - 84,296,354,346 total downloads

## [debian/ncurses-bin](#)

90,057 dependent images - 84,367,279,374 total downloads

## [debian/libp11-kit0](#)

89,819 dependent images - 101,129,807,438 total downloads

## [debian/libgmp10](#)

89,815 dependent images - 101,148,232,400 total downloads

## [debian/tzdata](#)

83,202 dependent images - 76,797,119,733 total downloads

## [debian/libc6](#)

81,084 dependent images - 76,188,542,194 total downloads

# Agenda

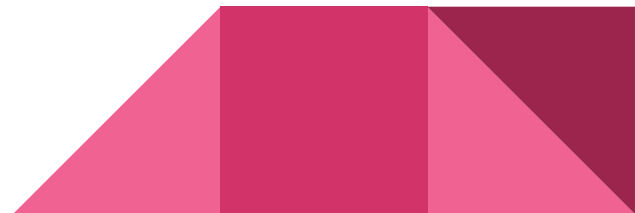
- What is Ecosyste.ms?
- Ecosyste.ms Services
- **Case Studies**
- Roadmap
- Questions



# Ecosyste.ms Case Study - Mapping Software Mentions

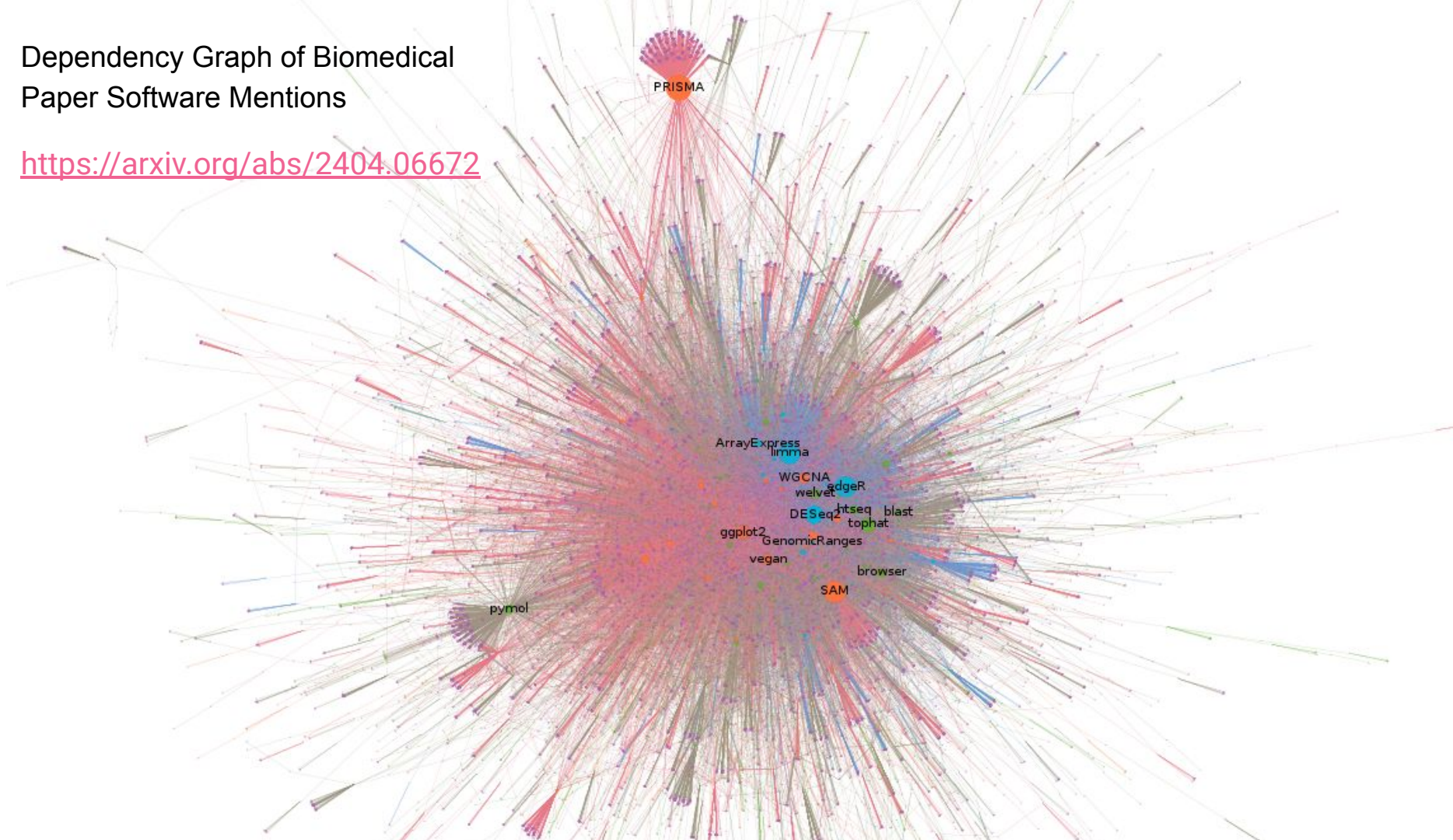
Mapping dependency graphs from software mentions in Biomedical Papers in the CZI Software Mentions dataset.

- Resolve full dependency tree for software mentioned in papers
  - Highlight credit to hidden contributors and credit
  - Connect all biomedical papers by their shared dependencies
- 
- Paper: <https://arxiv.org/abs/2404.06672>
  - Website: <https://papers.ecosyste.ms>
  - Code: <https://github.com/ecosyste.ms/papers>



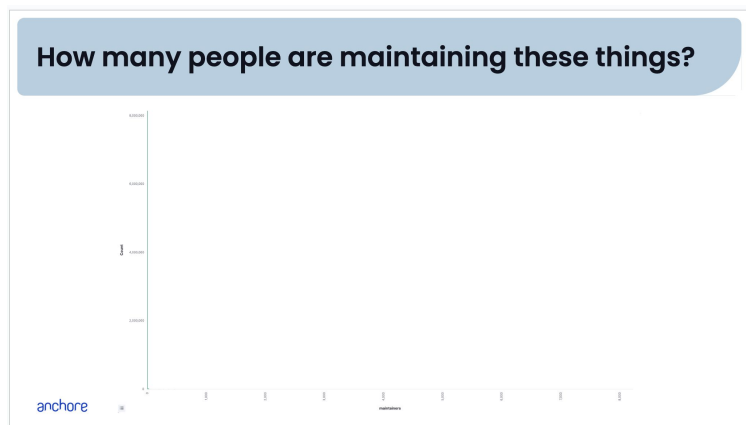
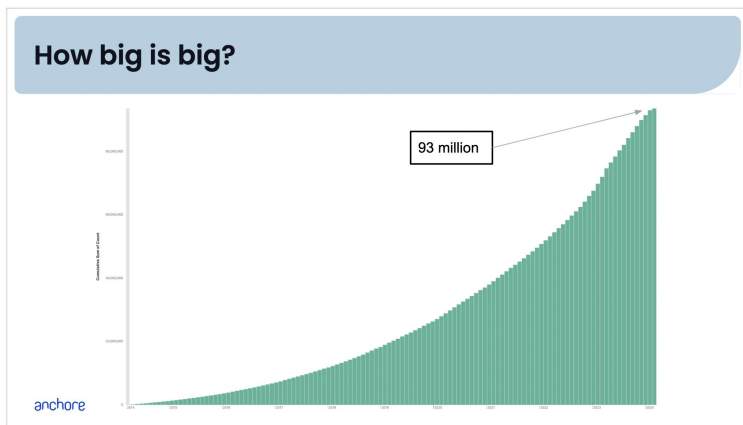
# Dependency Graph of Biomedical Paper Software Mentions

<https://arxiv.org/abs/2404.06672>



# Ecosyste.ms Case Study - Critical OSS

Discovering both the visible and invisible core pieces of open source software across every ecosystem.

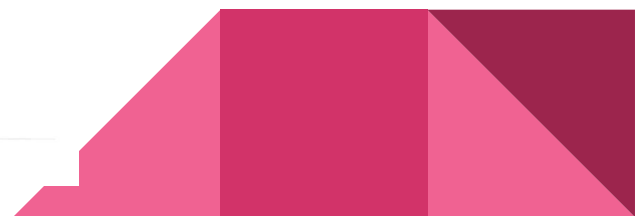
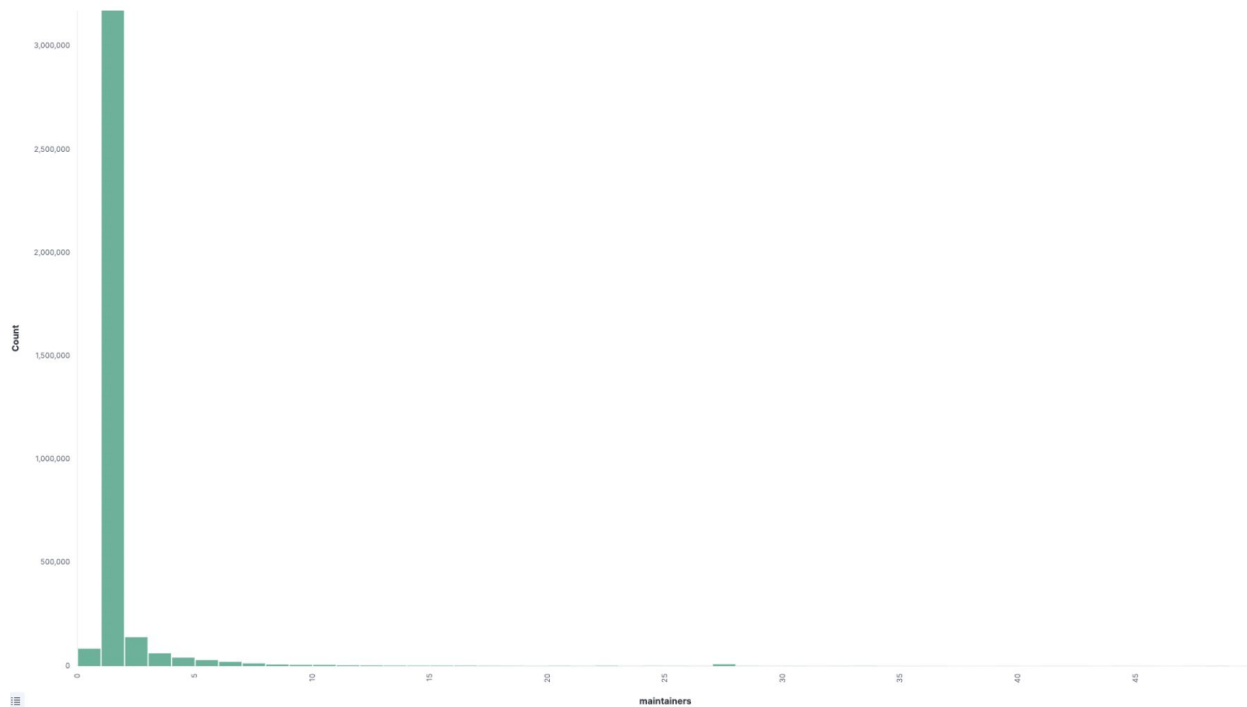


- Slides: <https://tinyurl.com/joshbressers>
- Data: <https://packages.ecosyste.ms/open-data>
- Related website: <https://packages.ecosyste.ms/critical>



# Ecosyste.ms Case Study - Critical OSS

86% of “critical” open source projects only have one maintainer



# Ecosyste.ms Case Study - Funding.yml

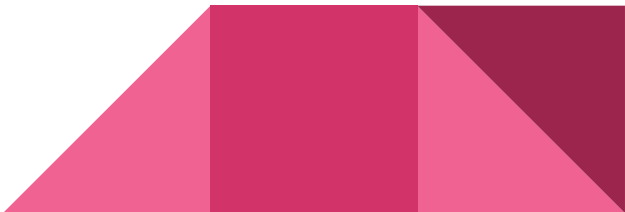
286,425 packages (3.03%) have declared a way to fund their development via a funding platform in their metadata.

22% of “Critical” packages and 14% of the “Top 1%” of packages have funding metadata.

Funded packages are detected via a funding url on their registry, via a funding.yml file in their source repository or the owner of the repository is part of GitHub Sponsors.

Soon to be expanded with metadata of if they belong to a Foundation.

- Website: <https://packages.ecosyste.ms/funding>
- Code: <https://github.com/ecosyste-ms/packages>



# Ecosyste.ms Case Study - Funding by Ecosystem

[npmjs.org](https://npmjs.org)

119,947 funded packages (3.05%)



[crates.io](https://crates.io)

12,565 funded packages (8.58%)



[pypi.org](https://pypi.org)

37,434 funded packages (6.71%)



[rubygems.org](https://rubygems.org)

6,134 funded packages (3.19%)



[packagist.org](https://packagist.org)

23,835 funded packages (5.75%)



[github actions](https://github.com/actions)

3,538 funded packages (11.21%)



[repo1.maven.org](https://repo1.maven.org)

23,190 funded packages (4.64%)



[bower.io](https://bower.io)

2,738 funded packages (3.9%)



[proxy.golang.org](https://proxy.golang.org)

19,706 funded packages (1.45%)



[pub.dev](https://pub.dev)

2,273 funded packages (4.39%)



[nuget.org](https://nuget.org)

12,678 funded packages (2.01%)



[alpine](https://alpine)

1,864 funded packages (5.74%)



# Ecosyste.ms Case Study - Funding by Platform

[github.com](https://github.com)

240,017 funded packages (83.8%)



[buymeacoffee.com](https://buymeacoffee.com)

6,010 funded packages (2.1%)



[opencollective.com](https://opencollective.com)

39,861 funded packages (13.92%)



[liberapay.com](https://liberapay.com)

4,406 funded packages (1.54%)



[patreon.com](https://patreon.com)

21,603 funded packages (7.54%)



[paypal.com](https://paypal.com)

2,625 funded packages (0.92%)



[ko-fi.com](https://ko-fi.com)

11,488 funded packages (4.01%)



[sindresorhus.com](https://sindresorhus.com)

1,905 funded packages (0.67%)



[paypal.me](https://paypal.me)

11,185 funded packages (3.91%)



[mate-desktop.org](https://mate-desktop.org)

1,644 funded packages (0.57%)



[tidelift.com](https://tidelift.com)

9,203 funded packages (3.21%)



[issuehunt.io](https://issuehunt.io)

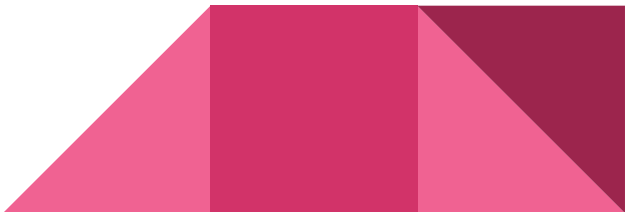
1,620 funded packages (0.57%)



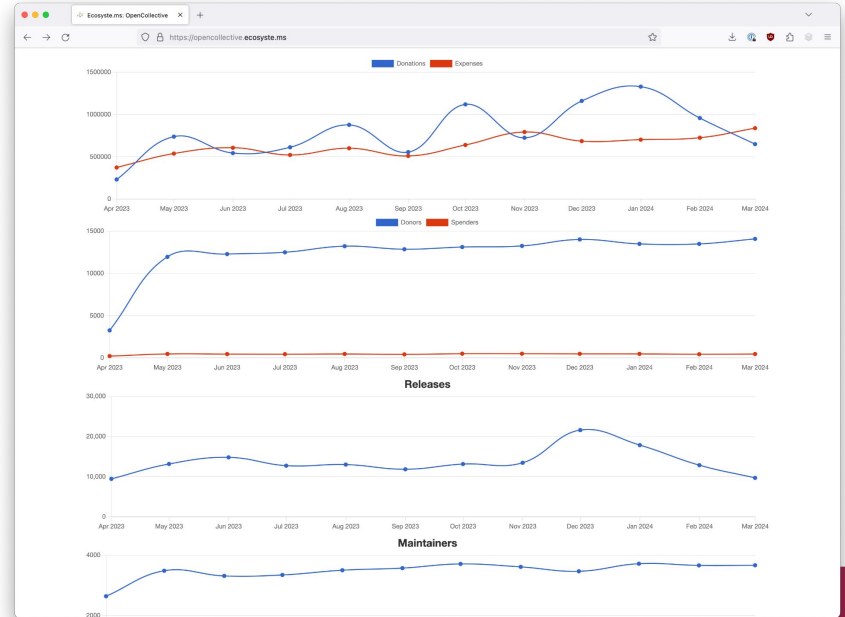
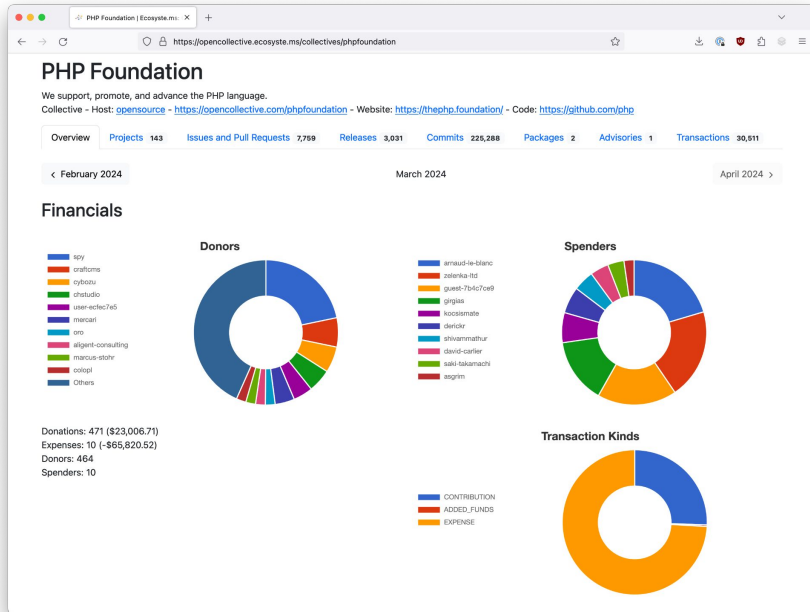
# Ecosyste.ms Case Study - Open Source Collective

Joining together the transaction data of donations, expenses and funders on Open Source Collective with the activity data from the open source projects being funded.

- Looking for correlations between funding and contributions
  - Allow funders to see the state of the projects they've supported
  - How are dependencies of OC projects also funded?
  - Fund your whole SBOM (coming soon)
- 
- Website: <https://opencollective.ecosyste.ms>
  - Code: <https://github.com/ecosyste-ms/opencollective>



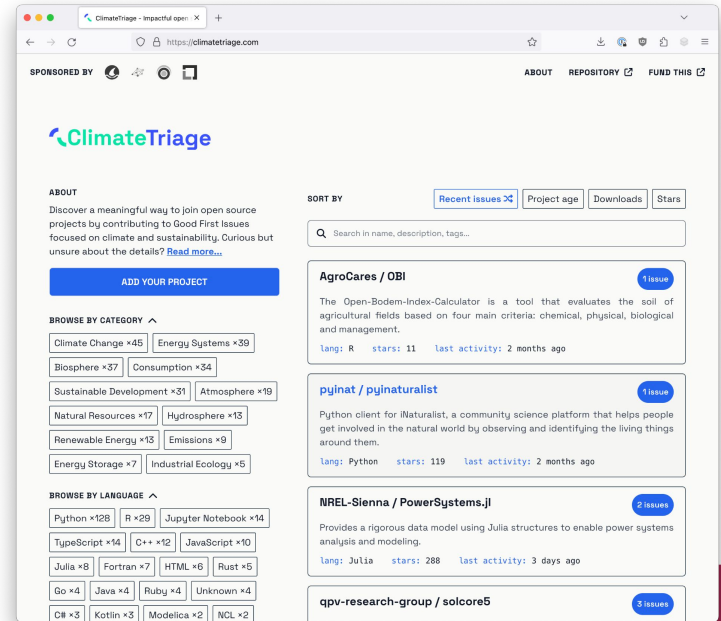
# Ecosyste.ms Case Study - Open Source Collective



# Ecosyste.ms Case Study - climatetriage.com

Highlight “good first issues” and “help wanted” issues from open source software projects in the areas of climate change, sustainable energy, biodiversity and natural resources from [opensustain.tech](https://opensustain.tech)

- Website: <https://climatetriage.com>
- Code: <https://github.com/prototypes/climate-triage>



# Agenda

- What is Ecosyste.ms?
- Ecosyste.ms Services
- Case Studies
- **Roadmap**
- Questions





# Ecosyste.ms Roadmap

What's next?

- Version and file level copyright and license data
- Changelogs and release notes per version
- OpenSSF Scorecards
- Project classification
- Search
- More system package manager support
- Software Foundations via <https://fossfoundation.info>
- Reverse Dependency Tooling

Propose ideas on <https://github.com/ecosyste-ms/roadmap>



# Reverse Dependency Tooling

Who depends on my open source project?

What versions of my software are people depending upon?

Are they direct dependents or transitive? Which packages are pulling in my library as a transitive dependency?

Who is affected by a security advisory I'm about to publish?

Are there packages holding back a version upgrade of transitive dependencies?

Are people actually merging automated updates from Dependabot?

Can I check I'm not making breaking changes against downstream users?



# Reverse Dependency Tooling

**Radar** Home Packages

## Packages: global-modules-path

Latest version: 3.0.3  
Returns path to globally installed package  
<https://github.com/roosefdavid/global-modules-path>

7 versions - Latest release: over 1 year ago - 45,668 dependent repositories - 52 dependent packages

### Version Breakdown

Major Versions	Minor Versions	Patch Versions
<ul style="list-style-type: none"><li>2.x.x: 43,025 repositories</li></ul>	<ul style="list-style-type: none"><li>2.3.x: 40,223 repositories</li><li>2.3.0: 2,918 repositories</li><li>2.3.0: 87 repositories</li><li>2.2.x: 64 repositories</li></ul>	<ul style="list-style-type: none"><li>2.3.0: 27,815 repositories</li><li>2.3.1: 12,800 repositories</li><li>2.1.0: 2,818 repositories</li><li>2.0.0: 87 repositories</li><li>2.2.0: 64 repositories</li></ul>

### Advisories

- [global-modules-path Command Injection vulnerability](#)  
Affected versions: 1.0.0, 2.0.0, 2.1.0, 2.2.0, 2.3.0, 2.3.1

### Repositories Breakdown

Direct/Transitive	Forks	Archived
<ul style="list-style-type: none"><li>Transitive: 45,428 repositories</li><li>Direct: 240 repositories</li></ul>	<ul style="list-style-type: none"><li>false: 45,428 repositories</li><li>true: 238 repositories</li></ul>	<ul style="list-style-type: none"><li>false: 42,517 repositories</li><li>true: 771 repositories</li></ul>

### Dependent Repositories 45,666

Repository	Owners
<a href="#">ElmEE/Elmement</a> <a href="#">github.com</a> A Vue.js 2.0 UI Toolkit for Web Stars: 83,019 Last pushed: 2023-05-07T10:08:52.000Z Received versions: 2.3.1 Direct dependency: false Maintainer: yam-lock	<a href="#">CoffeeCakes</a> 15 <a href="#">batian</a> 15 <a href="#">elme71x</a> 15 <a href="#">plattrethdee</a> 15 <a href="#">andreferrier</a> 15 <a href="#">cybal</a> 15 <a href="#">oukeraho</a> 15 <a href="#">microsoft</a> 15 <a href="#">Josh-star</a> 15 <a href="#">timselt</a> 15
<a href="#">date-fns/date-fns</a> <a href="#">github.com</a> Modern JavaScript date utility library Stars: 30,757 Last pushed: 2023-01-27T11:54:23.000Z Received versions: 2.3.0 Direct dependency: false Maintainer: examples/webpack-4.x/yarn-lock	

## Ecosystems: Issues

An open API service for providing issue and pull request metadata for open source projects.

### GitHub / Dependabot issues and pull requests

- [#456 - Bump bibliotherapy from 'a922b5d' to '571ec9f'](#)  
Pull Request - State: closed - Opened by [dependabot\[bot\]](#) 9 months ago  
Labels: dependencies, ruby
- [#364 - Bump bibliotherapy from 'a922b5d' to '571ec9f'](#)  
Pull Request - State: closed - Opened by [dependabot\[bot\]](#) 9 months ago  
Labels: dependencies, ruby
- [#3405 - Bump rack-protection from 3.0.6 to 3.1.0](#)  
Pull Request - State: closed - Opened by [dependabot\[bot\]](#) 9 months ago  
Labels: dependencies, ruby
- [#3404 - Bump rails-dom-testing from 2.1.1 to 2.2.0](#)  
Pull Request - State: closed - Opened by [dependabot\[bot\]](#) 9 months ago  
Labels: dependencies, ruby
- [#3403 - Bump addressable from 2.8.4 to 2.8.5](#)  
Pull Request - State: closed - Opened by [dependabot\[bot\]](#) 9 months ago  
Labels: dependencies, ruby
- [#303 - Bump zeitwerk from 2.6.9 to 2.6.11](#)  
Pull Request - State: closed - Opened by [dependabot\[bot\]](#) 9 months ago - 1 comment  
Labels: dependencies, ruby
- [#3402 - Bump zeitwerk from 2.6.10 to 2.6.11](#)  
Pull Request - State: closed - Opened by [dependabot\[bot\]](#) 9 months ago  
Labels: dependencies, ruby
- [#3401 - Bump redis-client from 0.14.1 to 0.15.0](#)  
Pull Request - State: closed - Opened by [dependabot\[bot\]](#) 9 months ago  
Labels: dependencies, ruby

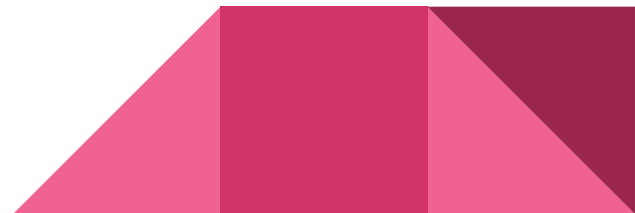
Code: <https://github.com/ecosyste-ms/radar>

# Thanks

Let's collaborate!

Code and data is all free to use and share.

- Website: <https://ecosyste.ms>
- Code: <https://github.com/ecosyste-ms>
- Mastodon: <https://mastodon.social/@ecosystems>
- Email: [andrew@ecosyste.ms](mailto:andrew@ecosyste.ms)



# Questions

In person or via #eum in slack

- Website: <https://ecosyste.ms>
- Code: <https://github.com/ecosyste-ms>
- Mastodon: <https://mastodon.social/@ecosystems>
- Email: [andrew@ecosyste.ms](mailto:andrew@ecosyste.ms)

