Airflow Beach Cleaning

Collaborative effort on securing Airflow ecosystem with Alpha-Omega, PSF & ASF

Jarek Potiuk

Apache Airflow PMC
member & committer

Member of Apache
Software Foundation
Security Committee

Michael Winser

Alpha-Omega
Co-founder

Eclipse Foundation Security
Strategy Ambassador

Alpha-Omega Mission

Catalyze sustainable security improvements within the most critical open source projects and ecosystems.

Alpha-Omega Grants

$23M Raised

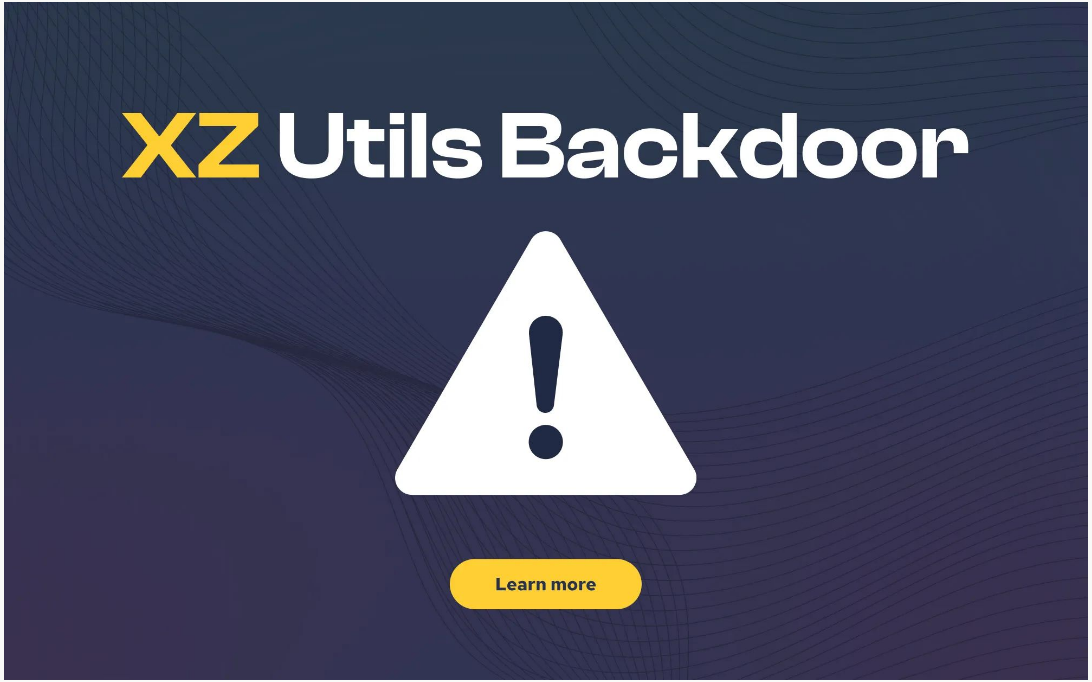
AI Libraries Airflow Apache (ASF) Clang Eclipse FreeBSD
Homebrew Jenkins JQuery Linux Kernel LLVM NodeJS
Omega OpenJS OpenSSL PHP Python Rubygems Rust

Alpha-Omega Explained

α → Leverage

Ω → Scale

A Wake Up Call



Open Source Supply Chain Risk

ALL MODERN DIGITAL INFRASTRUCTURE
A PROJECT SOME RANDOM PERSON IN NEBRASKA HAS BEEN THANKLESSLY MAINTAINING SINCE 2003

Size != Risk

Free Puppies Not Free Beer

YOUR CODE
OPEN-SOURCE SOFTWARE

Most of your code is OSS

Scaled Open Source Security Is Very Hard

Automation is noisy

Maintainers are wary

Impact is diffuse

From Pacific Garbage Patch to Beach Cleaning

Narrow Focus

Actionable Impact

Community Momentum

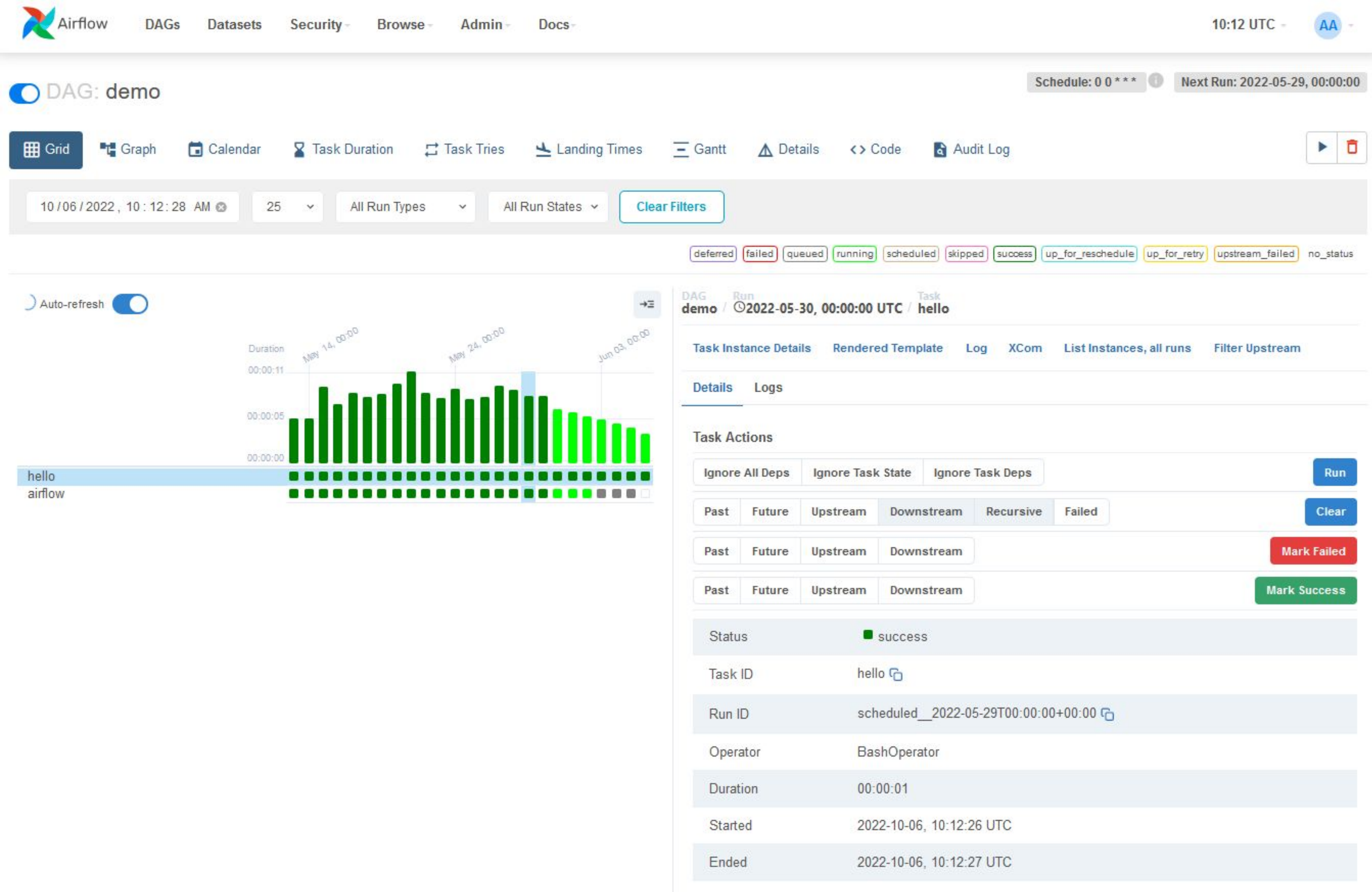
Selecting The Right Beach

Well Resourced Project **Community**

Security **Culture**

Supply Chain **Inventory**

So let’s talk about Airflow Security



DAG: demo

Schedule: 0 0 * * *

Next Run: 2022-05-29, 00:00:00

10 / 06 / 2022 , 10 : 12 : 28 AM

25

All Run Types

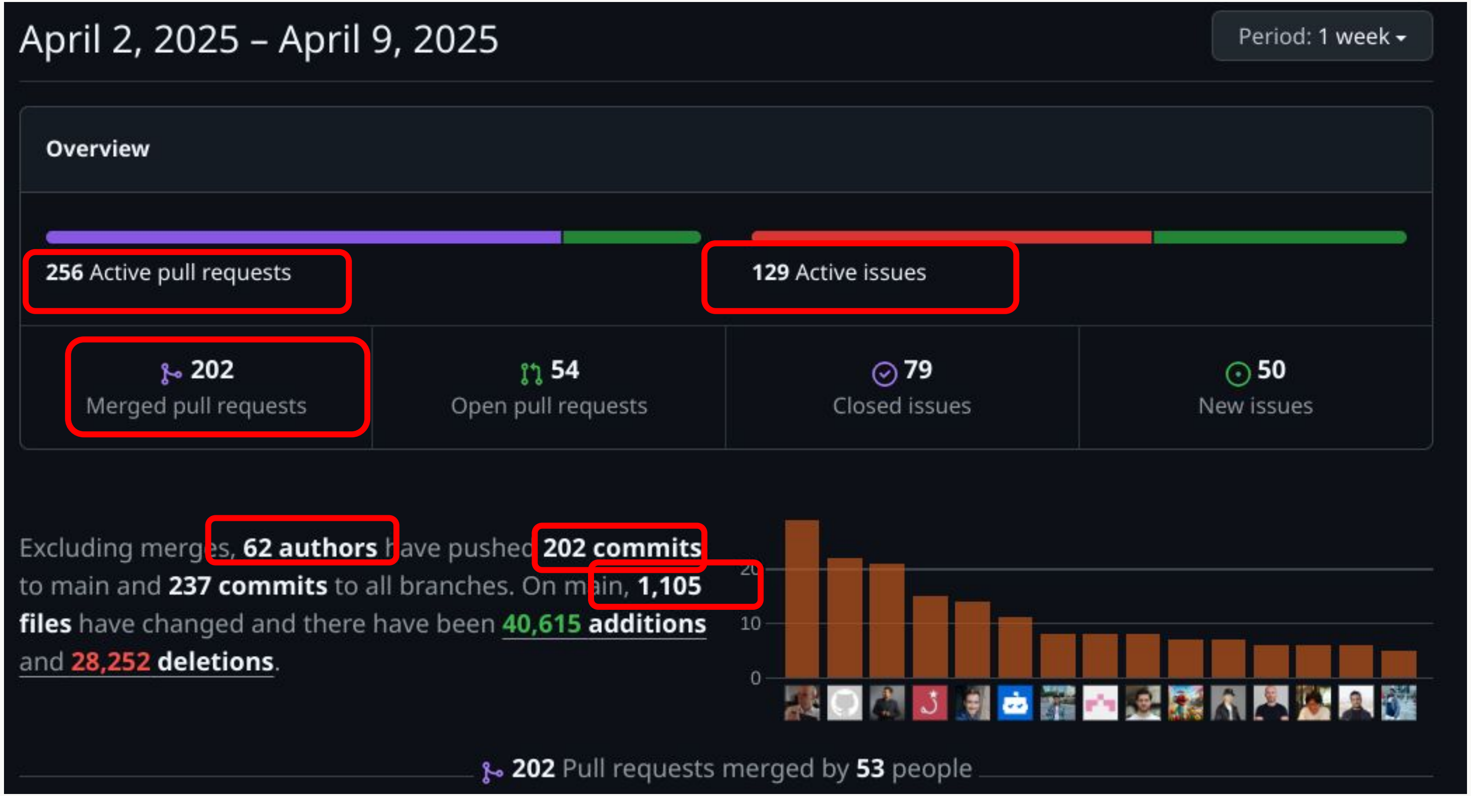
All Run States

DAG demo / Run 2022-05-30, 00:00:00 UTC / Task hello

Task Actions

Status	🟩 success
Task ID	hello
Run ID	scheduled__2022-05-29T00:00:00+00:00
Operator	BashOperator
Duration	00:00:01
Started	2022-10-06, 10:12:26 UTC
Ended	2022-10-06, 10:12:27 UTC

Airflow by numbers



April 2, 2025 – April 9, 2025

Period: 1 week ▾

Overview

256 Active pull requests

129 Active issues

202 Merged pull requests

54 Open pull requests

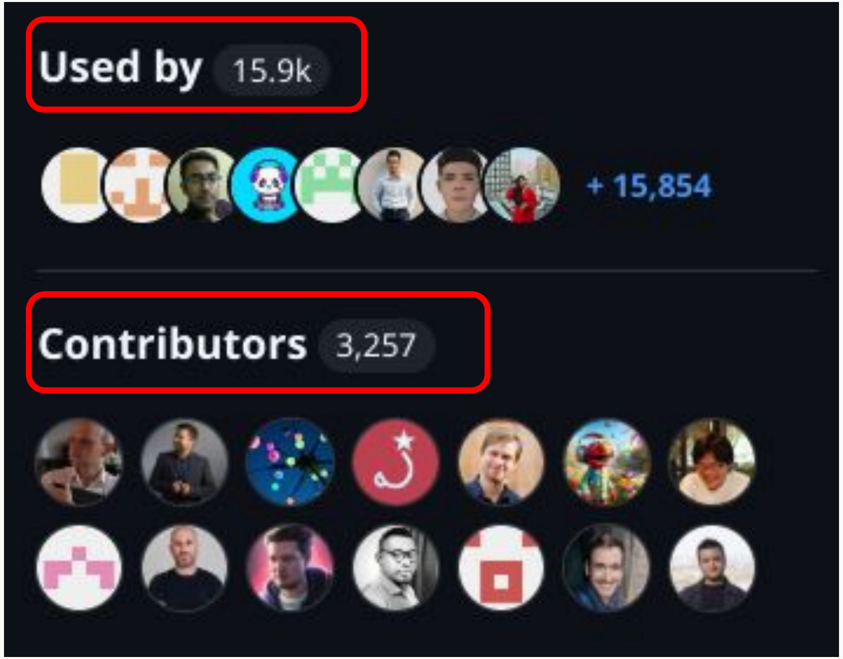
79 Closed issues

50 New issues

Excluding merges, **62 authors** have pushed **202 commits** to main and **237 commits** to all branches. On main, **1,105 files** have changed and there have been **40,615 additions** and **28,252 deletions**.

202 Pull requests merged by **53** people

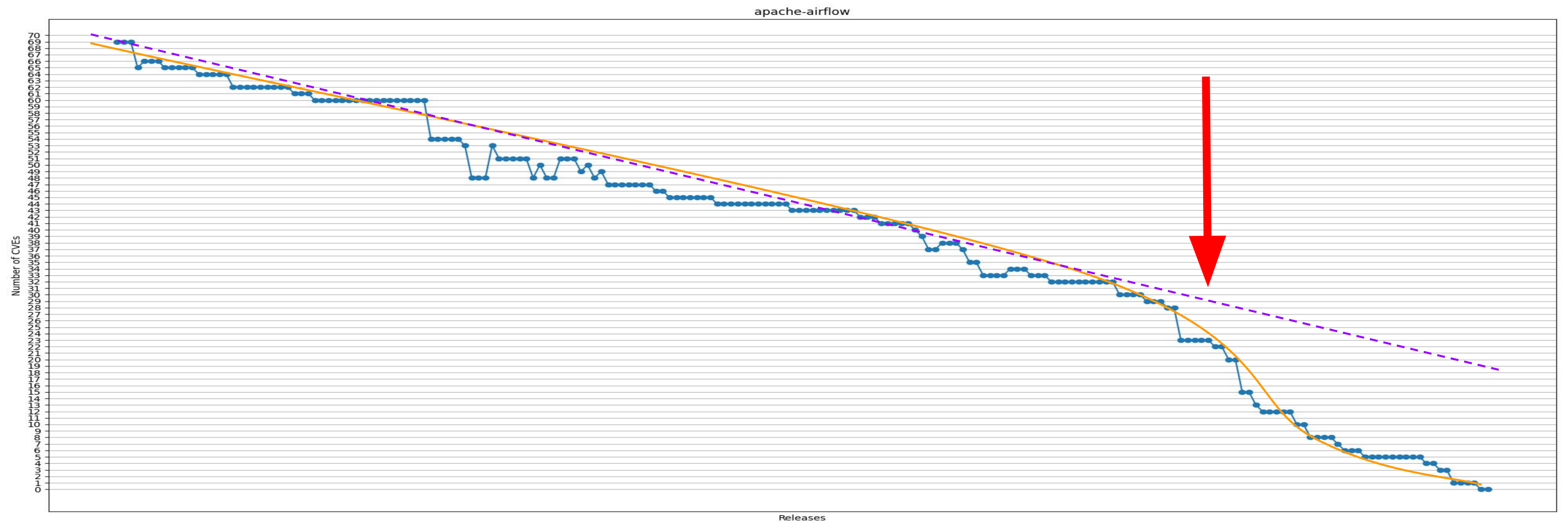
Users and Contributors



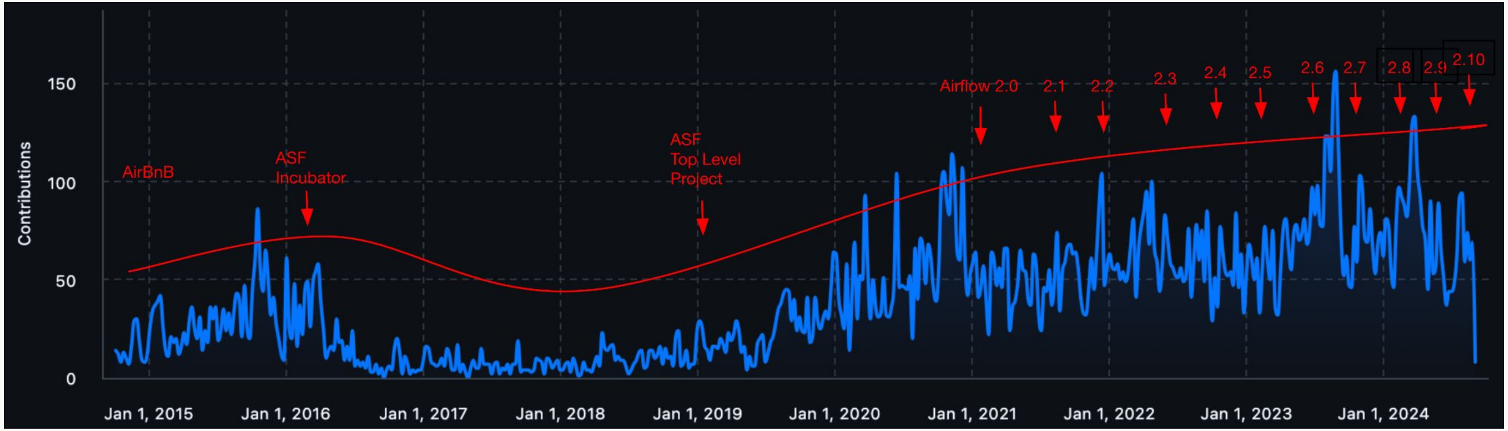
Used by 15.9k

+ 15,854

Contributors 3,257



apache-airflow
Number of CVEs
Releases
70
69
68
67
66
65
64
63
62
61
60
59
58
57
56
55
54
53
52
51
50
49
48
47
46
45
44
43
42
41
40
39
38
37
36
35
34
33
32
31
30
29
28
27
26
25
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0



Contributions
150
100
50
0
Jan 1, 2015
Jan 1, 2016
Jan 1, 2017
Jan 1, 2018
Jan 1, 2019
Jan 1, 2020
Jan 1, 2021
Jan 1, 2022
Jan 1, 2023
Jan 1, 2024
AirBnB
ASF
Incubator
ASF
Top Level
Project
Airflow 2.0
2.1
2.2
2.3
2.4
2.5
2.6
2.7
2.8
2.9
2.10

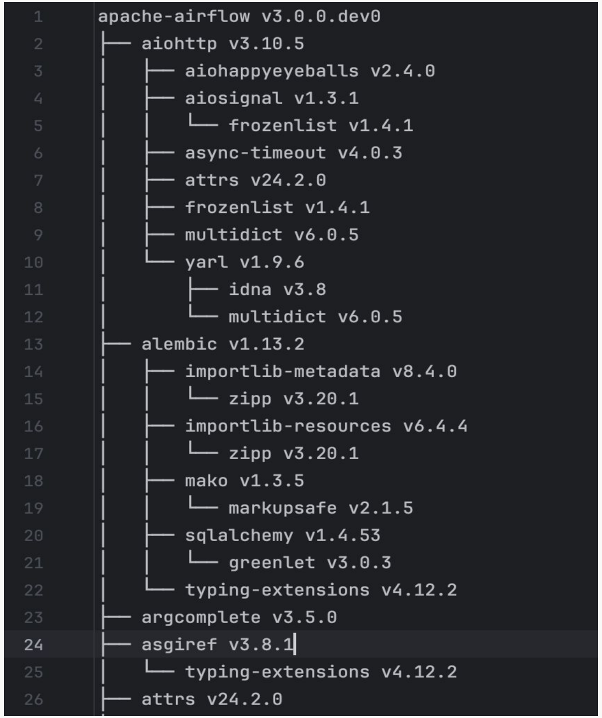
Security improvements in 2023-2024

- Dedicated security team
- Created and documented detailed process
- Introduced security model
- Canned responses to issues
- Disabled inherently insecure features
- Hardened CI workflows
- Introduced reproducible builds (provenance)

Airflow

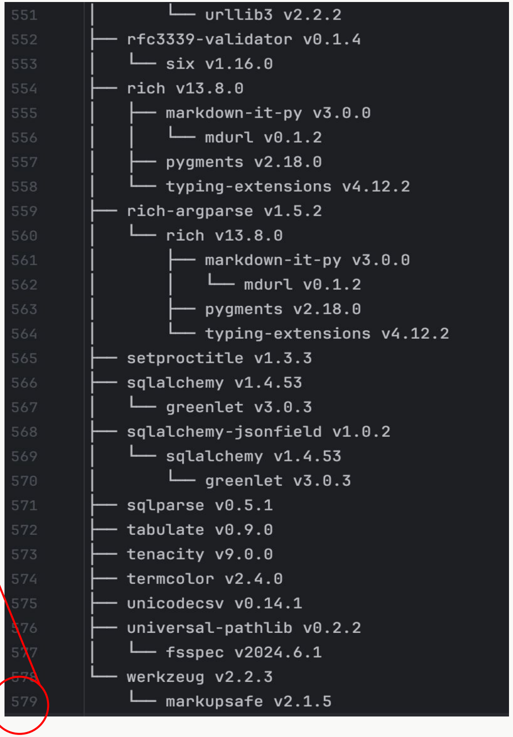
- Active security team (15 people, ~ 5 more active)
- Airflow: 66 committers, 33 PMC members
- 3200+ (!!!) contributors
- Airflow is big “enough” to attract funding
    - Stakeholders
    - Sovereign Tech Fund : 2023
    - Alpha-Omega Fund: 2024

Dependency tree



```
apache-airflow v3.0.0.dev0
├── aiohttp v3.10.5
│   ├── aiohappyeyeballs v2.4.0
│   ├── aiosignal v1.3.1
│   │   └── frozenlist v1.4.1
│   ├── async-timeout v4.0.3
│   ├── attrs v24.2.0
│   ├── frozenlist v1.4.1
│   ├── multidict v6.0.5
│   └── yarl v1.9.6
│       ├── idna v3.8
│       └── multidict v6.0.5
├── alembic v1.13.2
│   ├── importlib-metadata v8.4.0
│   │   └── zipp v3.20.1
│   ├── importlib-resources v6.4.4
│   │   └── zipp v3.20.1
│   ├── mako v1.3.5
│   │   └── markupsafe v2.1.5
│   ├── sqlalchemy v1.4.53
│   │   └── greenlet v3.0.3
│   └── typing-extensions v4.12.2
├── argcomplete v3.5.0
├── asgiref v3.8.1
│   └── typing-extensions v4.12.2
├── attrs v24.2.0
```

579



```
│       └── urllib3 v2.2.2
├── rfc3339-validator v0.1.4
│   └── six v1.16.0
├── rich v13.8.0
│   ├── markdown-it-py v3.0.0
│   │   └── mdurl v0.1.2
│   ├── pygments v2.18.0
│   └── typing-extensions v4.12.2
├── rich-argparse v1.5.2
│   └── rich v13.8.0
│       ├── markdown-it-py v3.0.0
│       │   └── mdurl v0.1.2
│       ├── pygments v2.18.0
│       └── typing-extensions v4.12.2
├── setproctitle v1.3.3
├── sqlalchemy v1.4.53
│   └── greenlet v3.0.3
├── sqlalchemy-jsonfield v1.0.2
│   └── sqlalchemy v1.4.53
│       └── greenlet v3.0.3
├── sqlparse v0.5.1
├── tabulate v0.9.0
├── tenacity v9.0.0
├── termcolor v2.4.0
├── unicodecsv v0.14.1
├── universal-pathlib v0.2.2
│   └── fsspec v2024.6.1
└── werkzeug v2.2.3
    └── markupsafe v2.1.5
```

Supply chain in OSS

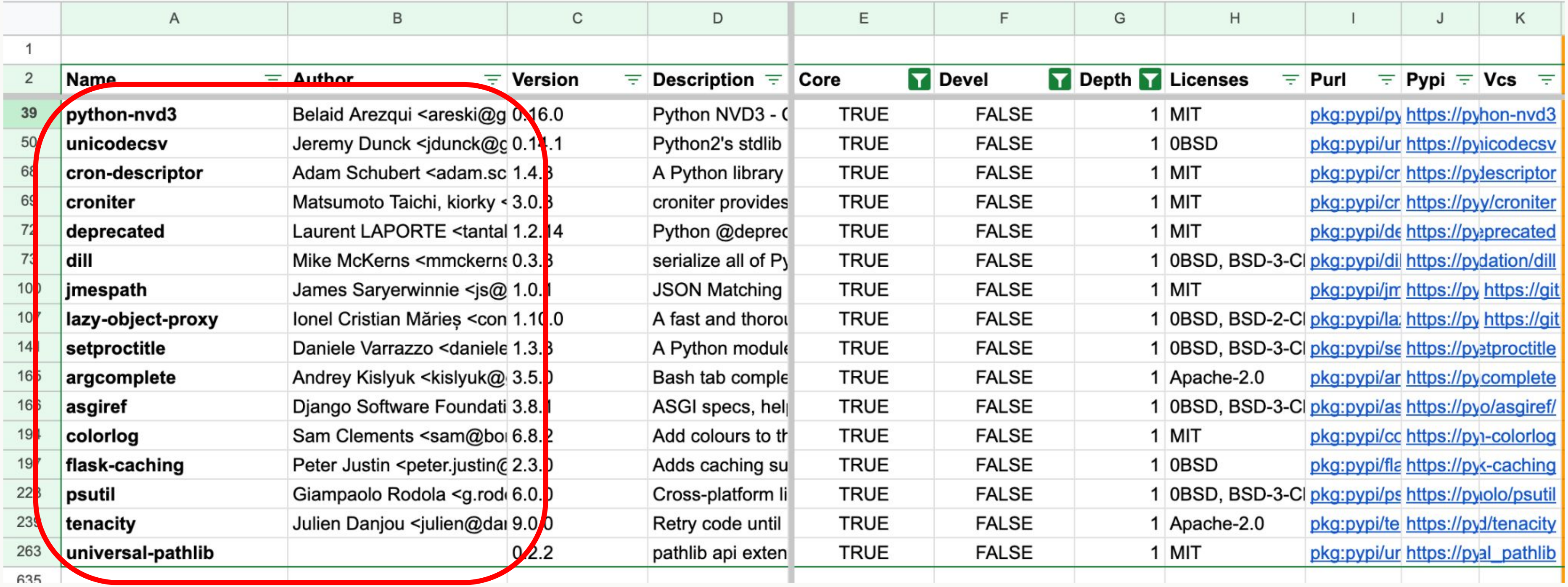
United effort

- Apache Software Foundation
- Airflow PMC
- Python Software Foundation
- Alpha-Omega Fund

Goals and Principles

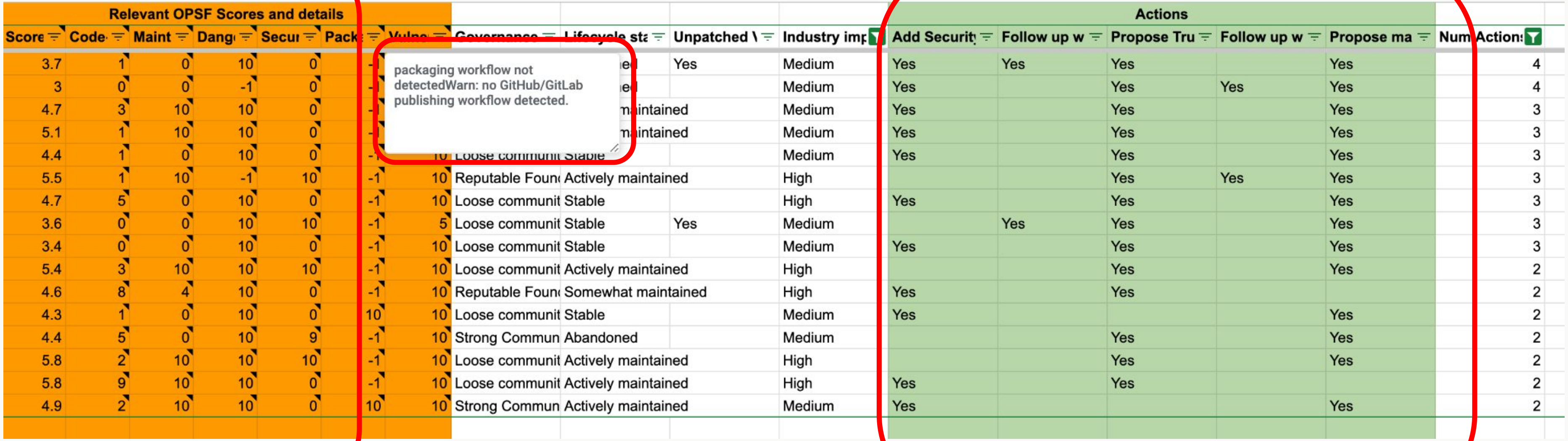
- We want to review ALL our dependencies (700+ !)
- We are leading and learning and adapting
- Automation to scale
- Your project’s “perspective” is important
- Always remember the people

Inventory - automated in about 50%



	A	B	C	D	E	F	G	H	I	J	K
1											
2	**Name**	**Author**	**Version**	**Description**	**Core**	**Devel**	**Depth**	**Licenses**	**Purl**	**Pypi**	**Vcs**
39	**python-nvd3**	Belaid Arezqui <areski@g	0.16.0	Python NVD3 - (	TRUE	FALSE	1	MIT	pkg:pypi/py	https://pyhon-nvd3	
50	**unicodecsv**	Jeremy Dunck <jdunck@g	0.14.1	Python2's stdlib	TRUE	FALSE	1	0BSD	pkg:pypi/ur	https://pyicodecsv	
68	**cron-descriptor**	Adam Schubert <adam.sc	1.4.3	A Python library	TRUE	FALSE	1	MIT	pkg:pypi/cr	https://pydescriptor	
69	**croniter**	Matsumoto Taichi, kiorky <	3.0.3	croniter provides	TRUE	FALSE	1	MIT	pkg:pypi/cr	https://pyy/croniter	
72	**deprecated**	Laurent LAPORTE <tantal	1.2.14	Python @deprec	TRUE	FALSE	1	MIT	pkg:pypi/de	https://pyeprecated	
73	**dill**	Mike McKerns <mmckerns	0.3.8	serialize all of Py	TRUE	FALSE	1	0BSD, BSD-3-Cl	pkg:pypi/dil	https://pydation/dill	
100	**jmespath**	James Saryerwinnie <js@	1.0.1	JSON Matching	TRUE	FALSE	1	MIT	pkg:pypi/jm	https://py	https://git
107	**lazy-object-proxy**	Ionel Cristian Mărieș <con	1.10.0	A fast and thorou	TRUE	FALSE	1	0BSD, BSD-2-Cl	pkg:pypi/la:	https://py	https://git
141	**setproctitle**	Daniele Varrazzo <daniele	1.3.3	A Python module	TRUE	FALSE	1	0BSD, BSD-3-Cl	pkg:pypi/se	https://pyetproctitle	
165	**argcomplete**	Andrey Kislyuk <kislyuk@	3.5.0	Bash tab comple	TRUE	FALSE	1	Apache-2.0	pkg:pypi/ar	https://pycomplete	
166	**asgiref**	Django Software Foundati	3.8.1	ASGI specs, hel	TRUE	FALSE	1	0BSD, BSD-3-Cl	pkg:pypi/as	https://pyo/asgiref/	
194	**colorlog**	Sam Clements <sam@bo	6.8.2	Add colours to th	TRUE	FALSE	1	MIT	pkg:pypi/cc	https://pyn-colorlog	
197	**flask-caching**	Peter Justin <peter.justin@	2.3.0	Adds caching su	TRUE	FALSE	1	0BSD	pkg:pypi/fla	https://pyk-caching	
223	**psutil**	Giampaolo Rodola <g.rod	6.0.0	Cross-platform li	TRUE	FALSE	1	0BSD, BSD-3-Cl	pkg:pypi/ps	https://pyolo/psutil	
239	**tenacity**	Julien Danjou <julien@da	9.0.0	Retry code until	TRUE	FALSE	1	Apache-2.0	pkg:pypi/te	https://pyd/tenacity	
263	**universal-pathlib**		0.2.2	pathlib api exten	TRUE	FALSE	1	MIT	pkg:pypi/ur	https://pyal_pathlib	
635											

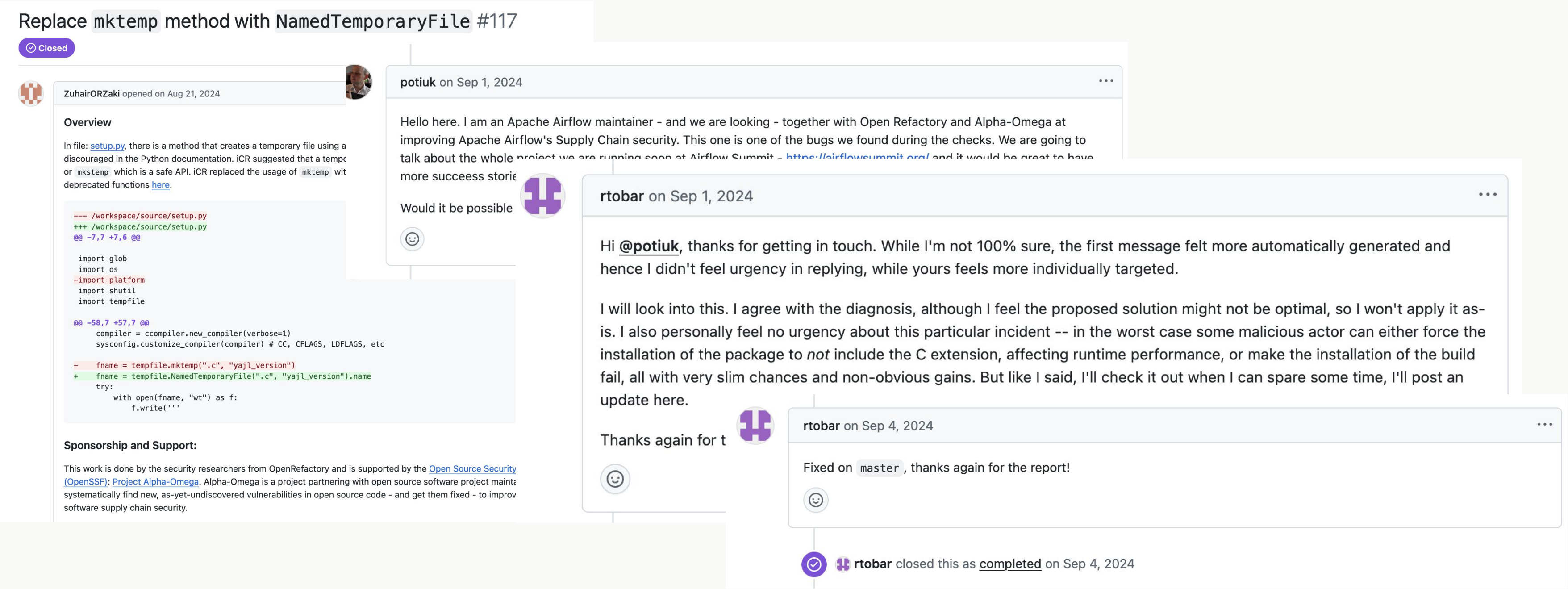
Assessment



Relevant OPSF Scores and details											Actions					
Score	Code	Maint	Dang	Secur	Pack	Vulne	Governance	Lifecycle sta	Unpatched V	Industry imp	Add Security	Follow up w	Propose Tru	Follow up w	Propose ma	Num Actions
3.7	1	0	10	0	-1	[illegible]	[illegible]	ed	Yes	Medium	Yes	Yes	Yes		Yes	4
3	0	0	-1	0	-1	[illegible]	[illegible]	ed		Medium	Yes		Yes	Yes	Yes	4
4.7	3	10	10	0	-1	[illegible]	[illegible]	maintained		Medium	Yes		Yes		Yes	3
5.1	1	10	10	0	-1	[illegible]	[illegible]	maintained		Medium	Yes		Yes		Yes	3
4.4	1	0	10	0	-1	10	Loose communit	Stable		Medium	Yes		Yes		Yes	3
5.5	1	10	-1	10	-1	10	Reputable Foun	Actively maintained		High			Yes	Yes	Yes	3
4.7	5	0	10	0	-1	10	Loose communit	Stable		High	Yes		Yes		Yes	3
3.6	0	0	10	10	-1	5	Loose communit	Stable	Yes	Medium		Yes	Yes		Yes	3
3.4	0	0	10	0	-1	10	Loose communit	Stable		Medium	Yes		Yes		Yes	3
5.4	3	10	10	10	-1	10	Loose communit	Actively maintained		High			Yes		Yes	2
4.6	8	4	10	0	-1	10	Reputable Foun	Somewhat maintained		High	Yes		Yes			2
4.3	1	0	10	0	10	10	Loose communit	Stable		Medium	Yes				Yes	2
4.4	5	0	10	9	-1	10	Strong Commun	Abandoned		Medium			Yes		Yes	2
5.8	2	10	10	10	-1	10	Loose communit	Actively maintained		High			Yes		Yes	2
5.8	9	10	10	0	-1	10	Loose communit	Actively maintained		High	Yes		Yes			2
4.9	2	10	10	0	10	10	Strong Commun	Actively maintained		Medium	Yes				Yes	2

packaging workflow not detectedWarn: no GitHub/GitLab publishing workflow detected.

The Human Component In Security Bug Reporting



Replace `mktemp` method with `NamedTemporaryFile` #117

ZuhairORZaki opened on Aug 21, 2024

Overview

In file: setup.py, there is a method that creates a temporary file using a discouraged in the Python documentation. iCR suggested that a tempc or `mkstemp` which is a safe API. iCR replaced the usage of `mktemp` wit deprecated functions here.

```
--- /workspace/source/setup.py
+++ /workspace/source/setup.py
@@ -7,7 +7,6 @@

 import glob
 import os
-import platform
 import shutil
 import tempfile

@@ -58,7 +57,7 @@
     compiler = ccompiler.new_compiler(verbose=1)
     sysconfig.customize_compiler(compiler) # CC, CFLAGS, LDFLAGS, etc

-    fname = tempfile.mktemp(".c", "yajl_version")
+    fname = tempfile.NamedTemporaryFile(".c", "yajl_version").name
     try:
         with open(fname, "wt") as f:
             f.write('''
```

Sponsorship and Support:

This work is done by the security researchers from OpenRefactory and is supported by the Open Source Security (OpenSSF): Project Alpha-Omega. Alpha-Omega is a project partnering with open source software project mainta systematically find new, as-yet-undiscovered vulnerabilities in open source code - and get them fixed - to improv software supply chain security.

potiuk on Sep 1, 2024

Hello here. I am an Apache Airflow maintainer - and we are looking - together with Open Refactory and Alpha-Omega at improving Apache Airflow's Supply Chain security. This one is one of the bugs we found during the checks. We are going to talk about the whole project we are running soon at Airflow Summit - https://airflowsummit.org/ and it would be great to have more success storie

Would it be possible

rtobar on Sep 1, 2024

Hi **@potiuk**, thanks for getting in touch. While I'm not 100% sure, the first message felt more automatically generated and hence I didn't feel urgency in replying, while yours feels more individually targeted.

I will look into this. I agree with the diagnosis, although I feel the proposed solution might not be optimal, so I won't apply it as-is. I also personally feel no urgency about this particular incident -- in the worst case some malicious actor can either force the installation of the package to *not* include the C extension, affecting runtime performance, or make the installation of the build fail, all with very slim chances and non-obvious gains. But like I said, I'll check it out when I can spare some time, I'll post an update here.

Thanks again for t

rtobar on Sep 4, 2024

Fixed on `master`, thanks again for the report!

rtobar closed this as completed on Sep 4, 2024

Experiment in progress …

Actions

- 16 projects to start with
- Add security policies
- Follow up on unsecure workflows
- Propose Trusted Publishing
- Follow up on unpatched vulnerabilities
- Propose mandatory code reviews

Responses from maintainers

- No response
- Happy to get help
- Great to get help
- I am also a security freak
- Croniter

The 3 Fs of your supply chain

Fix

Fork

Forgo

Example cases

Croniter (Fork)

- Maintainer threatening to remove the project
- Discussion started (public issue)
- People volunteered to fork
- Eventually - moved to Pallets-eco and we became maintainers

Werkzeug (Fix)

- Transitive dependency
- Held by few other dependencies
- Google security team works on patching

Connexion (Forego)

- Difficult to upgrade
- Google security team works on patching it
- We switched to Fast API

Long term targets

- Full automation (AI?) and coverage
- Target ALL projects
- Regular, incremental process
- Spread the methodology / findings
- Contribute to other efforts (PSF)

Learnings?

One Thing

OS project’s security depends on Their engagement with their supply chain

Takeaways

Supply chain relationships are a **human problem**

The transitive problem: every (new) dependency creates **exponential risk over time**

Current vulnerabilities are less important than **sustained security** handling and project health

Make **security a first-class priority** in every project plan

Q&A

[SSLSA]

[SSDF]

[Alpha-Omega Project]

[Case Study Eclipse Temurin: Pioneering Software Supply Chain Security]

[Open Source Software Foundation]

[Airflow Security]

[Python Security]