Security United

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

AIRFLOW SUMMIT 2024
10 YEARS OF AIRFLOW
SF·USA

Jarek Potiuk

Apache Airflow PMC member & committer

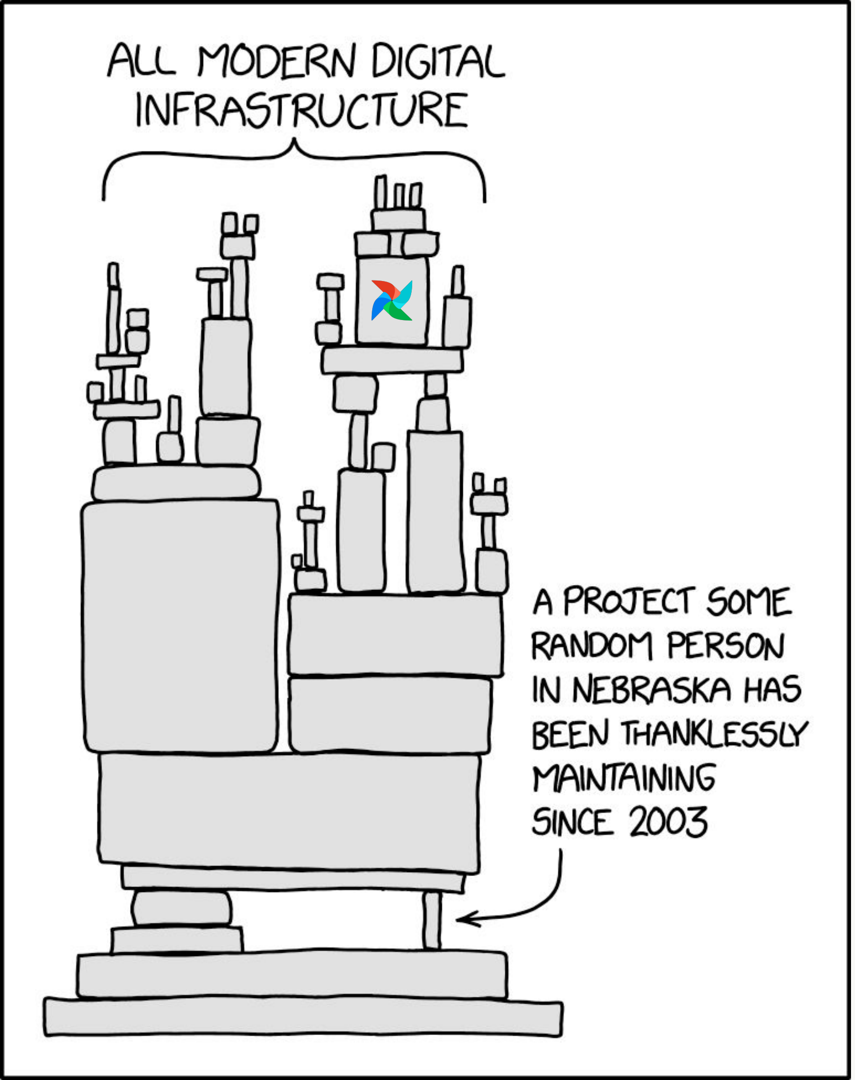
Member of Apache Software Foundation Security Committee

Michael Winser

Alpha-Omega co-founder

Eclipse Foundation Security Strategy Ambassador

What is Supply Chain Risk


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

https://xkcd.com/2347/

Supply Chain Risk Model

Vulnerabilities

Tampering

Availability

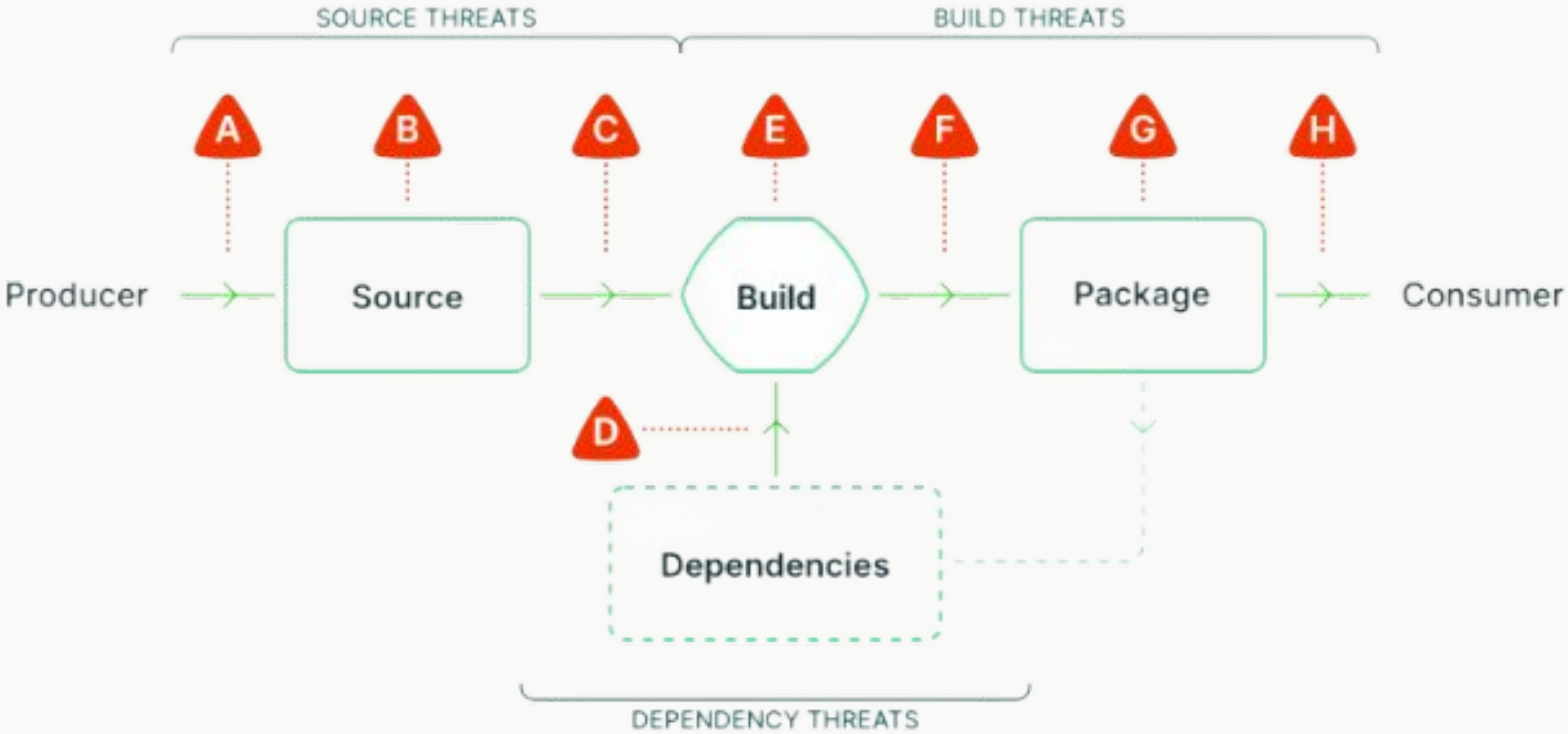
X

Culture

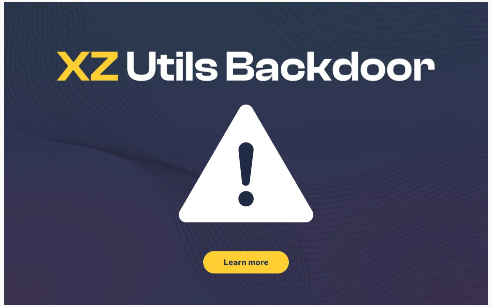
Process

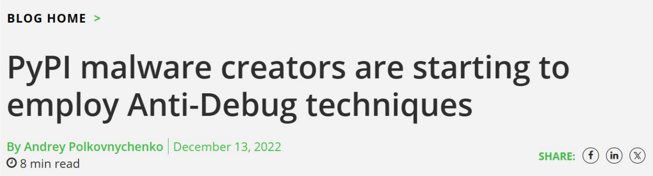
Solutions

SLSA Threat Model



Why Now



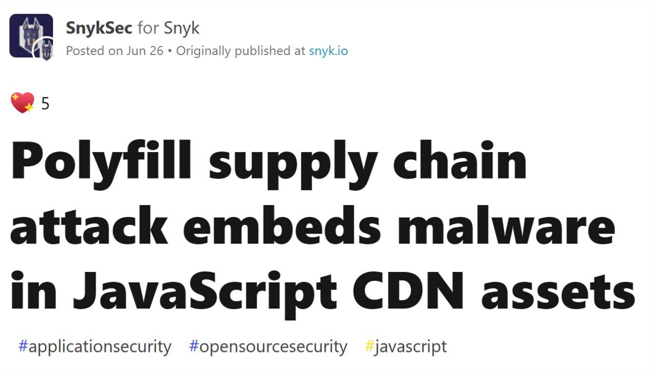

BLOG HOME >

PyPI malware creators are starting to employ Anti-Debug techniques

By Andrey Polkovnychenko | December 13, 2022

8 min read

SHARE:


SnykSec for Snyk

Posted on Jun 26 • Originally published at snyk.io

5

Polyfill supply chain attack embeds malware in JavaScript CDN assets

#applicationsecurity #opensourcesecurity #javascript

Social Engineering Campaign Targeting Tech Employees Spreading Through
npm
Malware

This Stuff is Hard

The 3 Fs of your supply chain

Fix

Fork

Forgo

One More F

Funding

Alpha-Omega Mission

Protect society by catalyzing sustainable security improvements to the most critical open source software projects and ecosystems

Alpha-Omega Strategy

A

Security Staffing

B

Package Repositories

C

Audits and Remediation

D

Innovation & Experiments

Alpha-Omega By the Numbers

13

Full Time Engineers Hired

9

Security Audits

$8.3M

in Grants

Orgs Funded

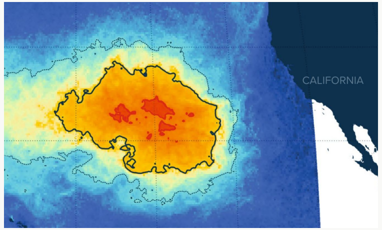
14

Origin Story: It's Always About People

Seth Larson and Michael Winser present at PyCon

Jarek & Michael meet

From the Pacific Garbage Patch to cleaning the Airflow beach



Alpha-Omega - Airflow PMC collaboration

Beach Cleaning

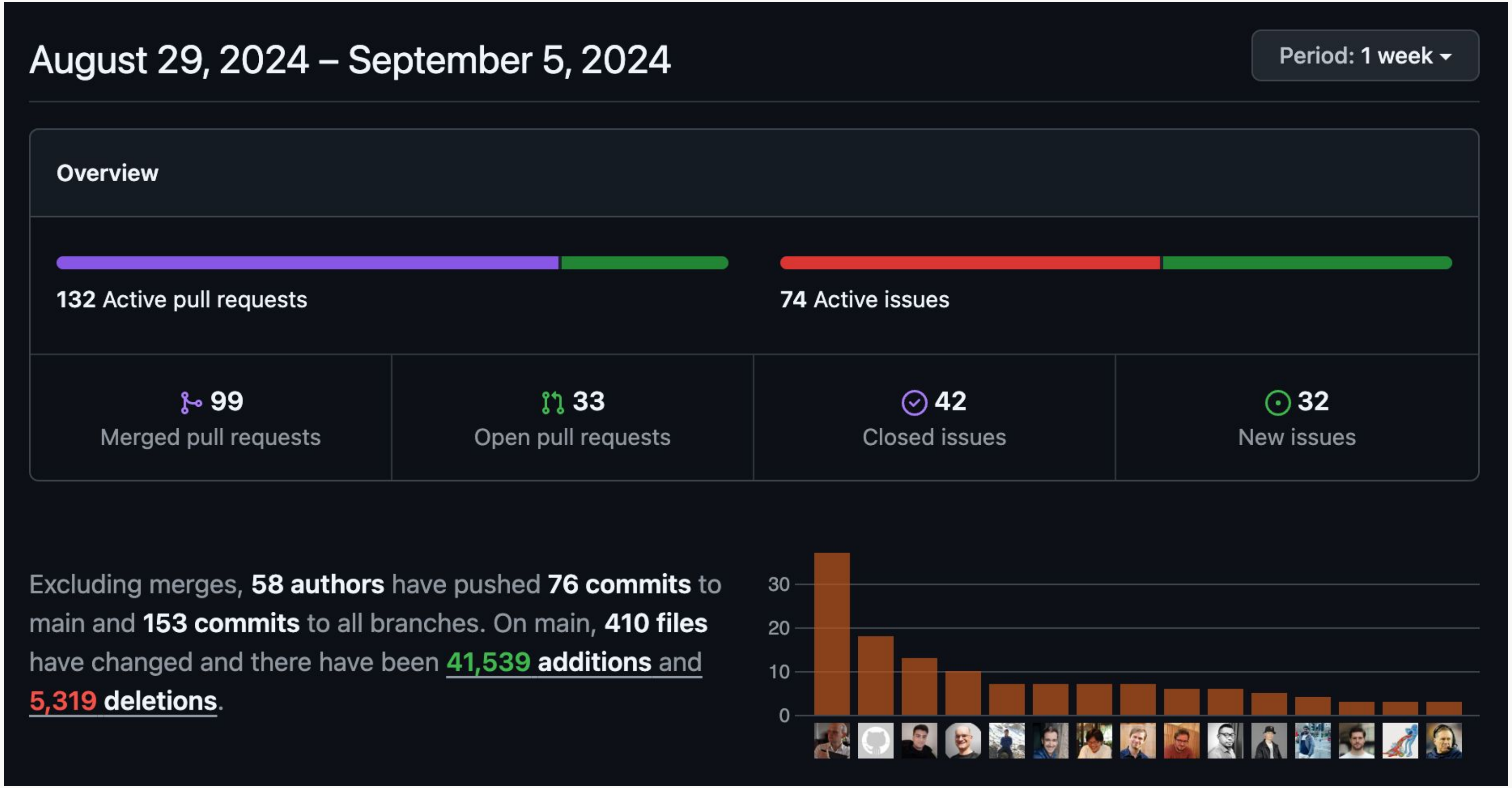
So let’s talk about Airflow Security

Airflow is secure

~~Airflow is secure~~

Airflow is active in its security

Airflow is active



August 29, 2024 – September 5, 2024

Period: 1 week ▾

Overview

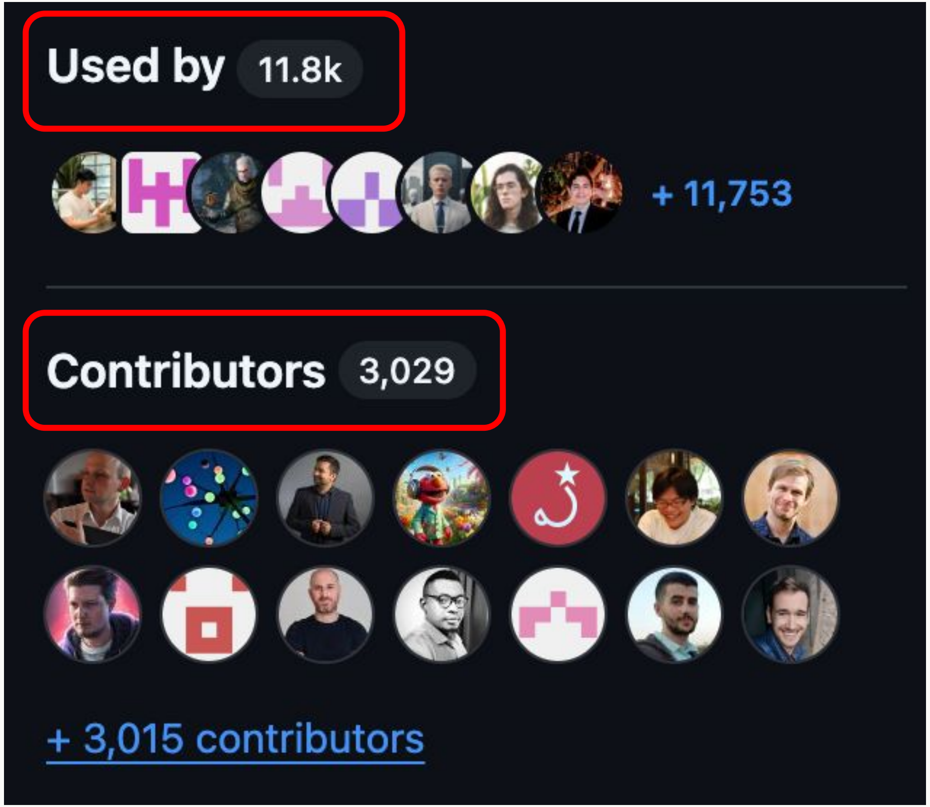
132 Active pull requests

74 Active issues

99	33	42	32
Merged pull requests	Open pull requests	Closed issues	New issues

Excluding merges, **58 authors** have pushed **76 commits** to main and **153 commits** to all branches. On main, **410 files** have changed and there have been **41,539 additions** and **5,319 deletions**.

Users and Contributors

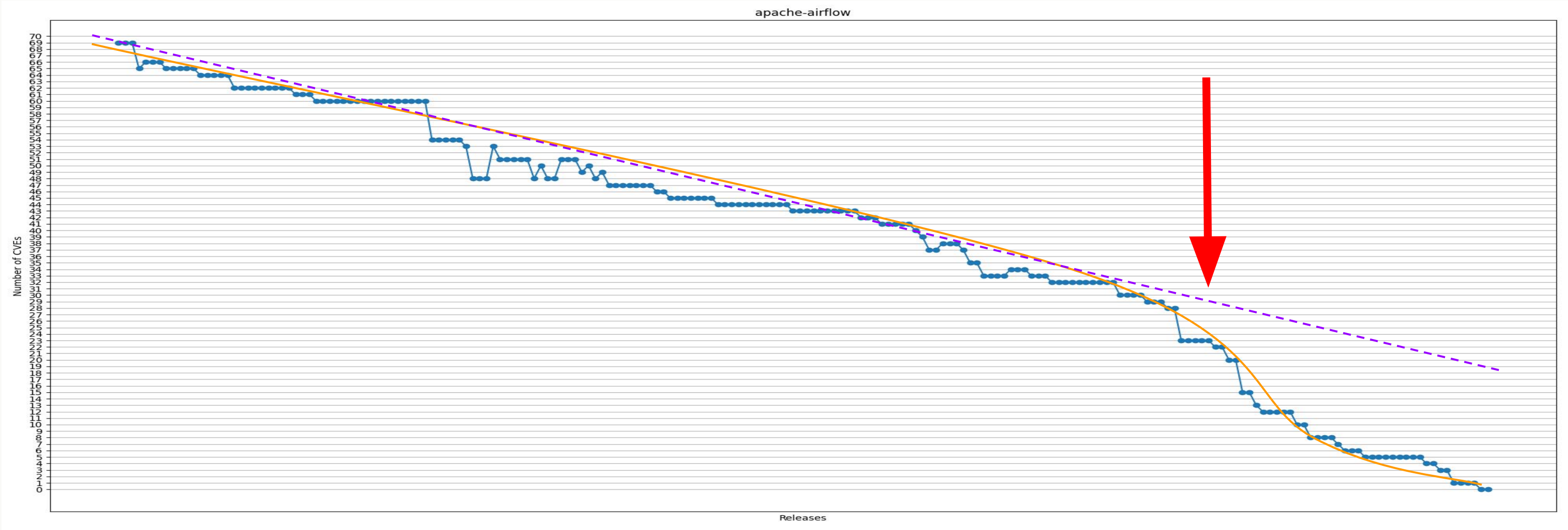


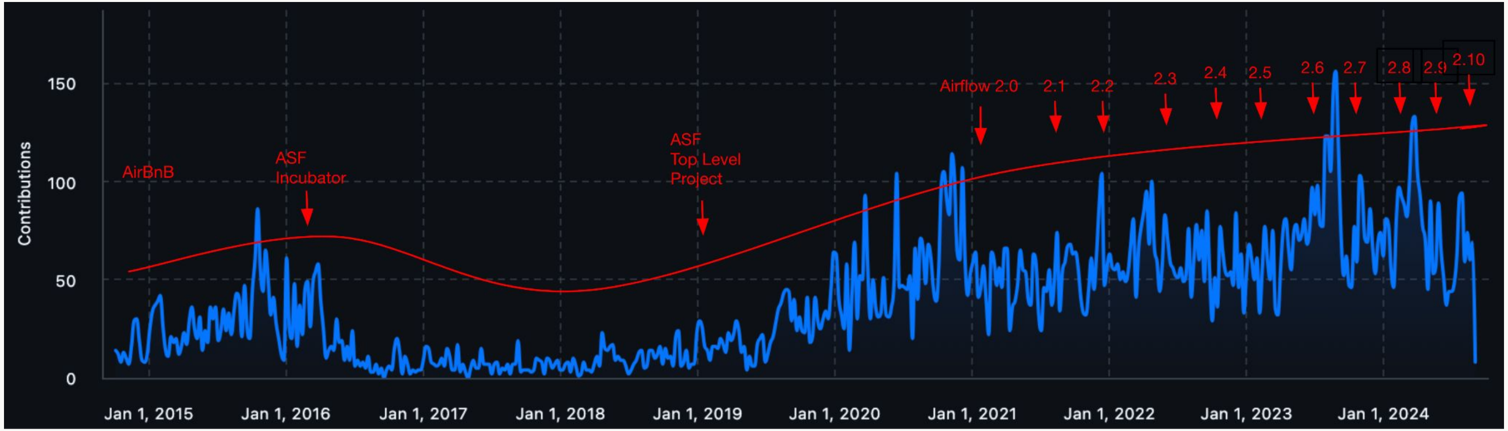
Used by 11.8k

+ 11,753

Contributors 3,029

+ 3,015 contributors


apache-airflow
Number of CVEs
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
Releases


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

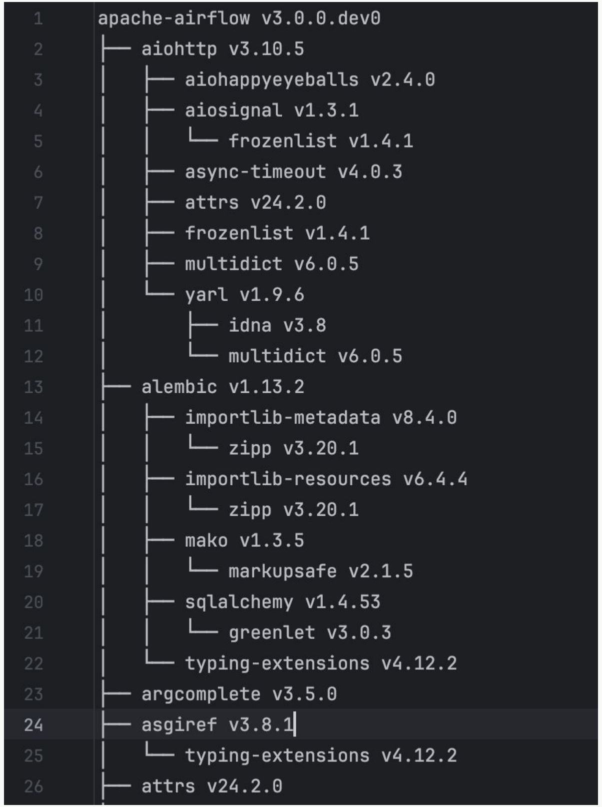
Security improvements in 2023-2024

- 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

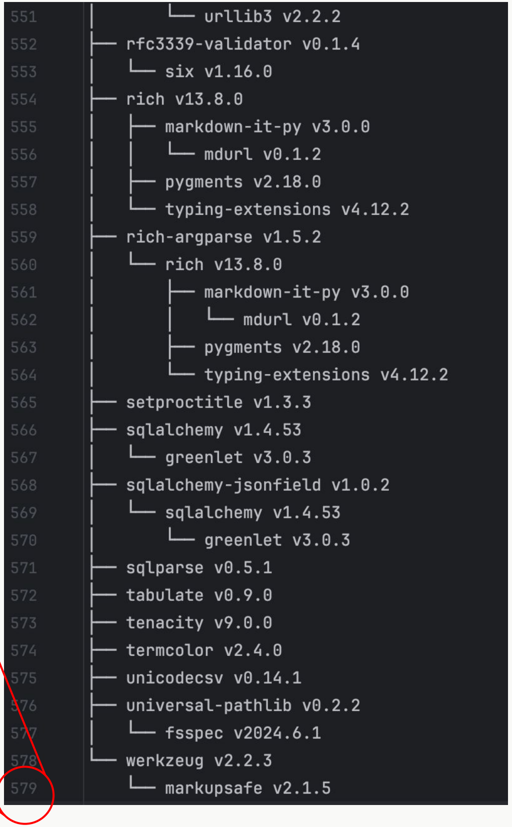
- Active security team (15 people, ~ 5 more active)
- Airflow: 62 committers, 33 PMC members
- 3000+ (!!!) contributors
- Airflow is big “enough” to attract funding
  - Stakeholders
  - Sovereign Tech Fund : 2023
  - Alpha-Omega Fund: 2024

Dependency tree

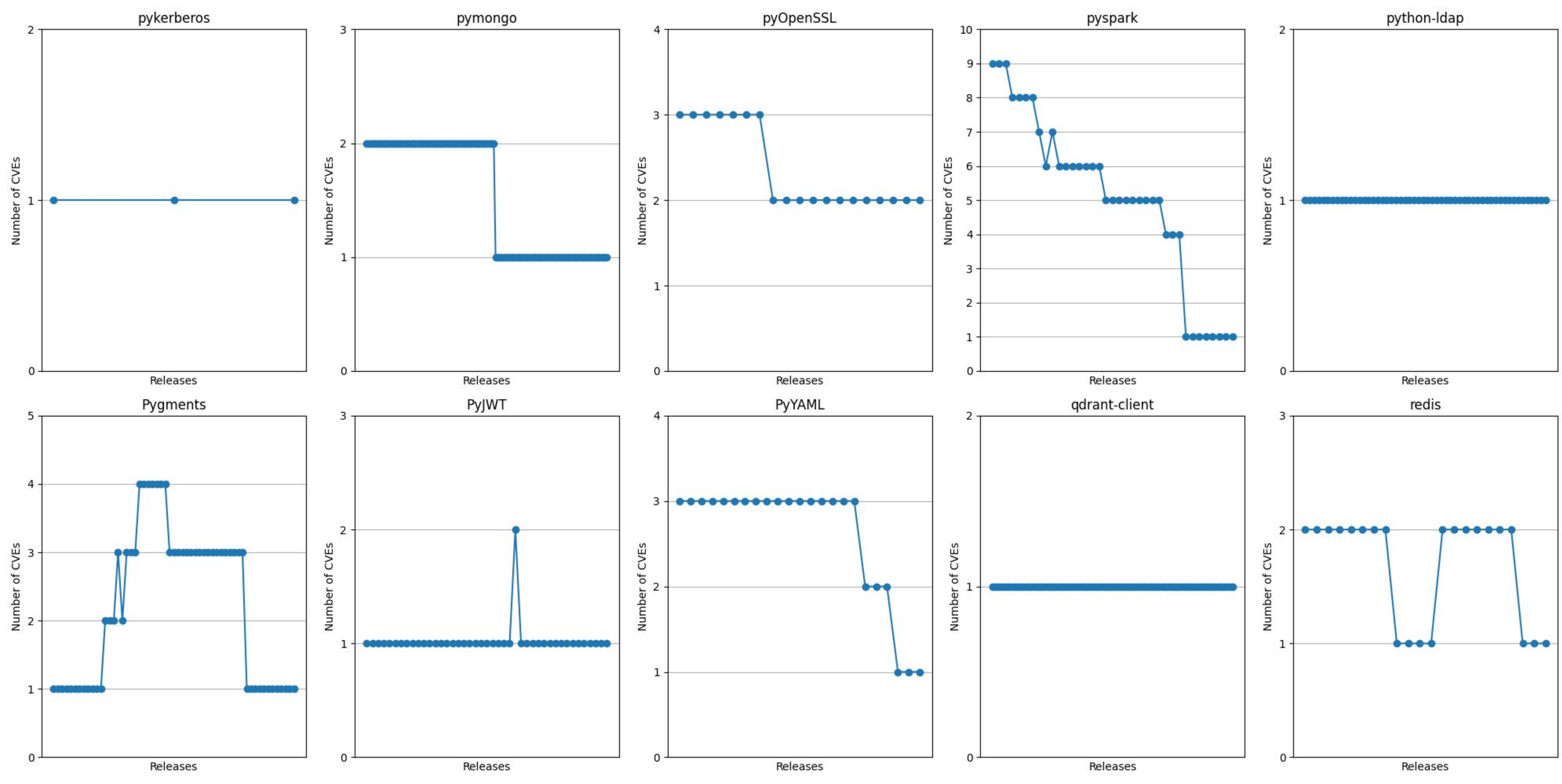


```
apache-airflow v3.0.0.dev0
├── aiohttp v3.10.5
│   ├── aiohappyeyeballs v2.4.0
│   ├── aiosignal v1.3.1
│   │   └── frozenlist v1.4.1
│   ├── async-timeout v4.0.3
│   ├── attrs v24.2.0
│   ├── frozenlist v1.4.1
│   ├── multidict v6.0.5
│   └── yarl v1.9.6
│       ├── idna v3.8
│       └── multidict 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
```


pykerberos
pymongo
pyOpenSSL
pyspark
python-ldap
Pygments
PyJWT
PyYAML
qdrant-client
redis
Number of CVEs
Releases
0
1
2
3
4
5
6
7
8
9
10

Airflow Security Ecosystem

- Users
- Committers and PMC members
- Airflow Dependencies

Security regulation

- Regulations are coming
- We have less than 2 years
- Everyone is impacted
- Everyone needs to be involved

Experiment starts ...

United effort

- Apache Software Foundation
- Airflow PMC
- Python Software Foundation
- Alpha-Omega Fund
- Some users (indirectly)

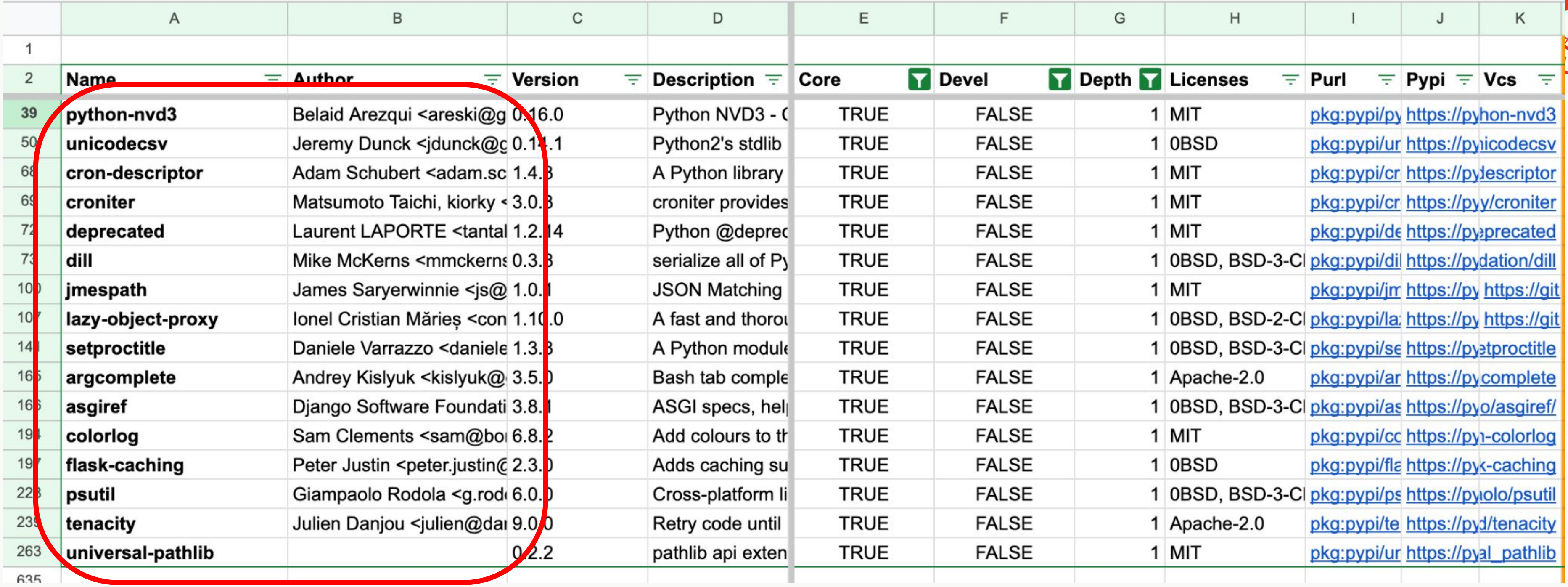
Involved parties

- OpenRefactory (analysing source code for bugs)
- CDXGen (generating SBOMs and other inventories)
- OSTIF (security audits)
- External researchers/security specialists

Goals and Principles

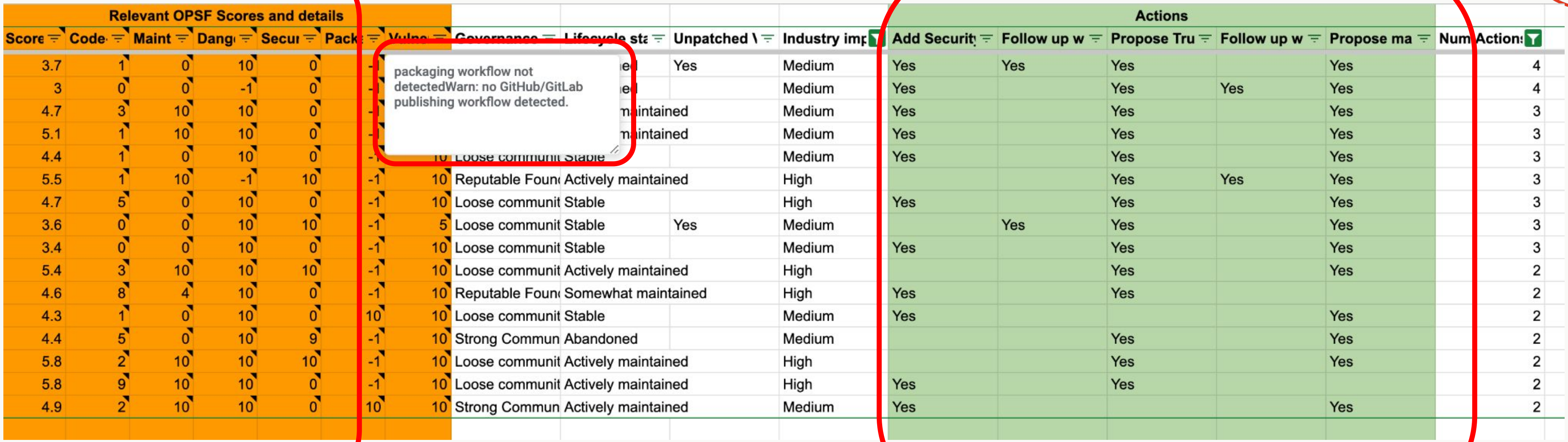
- We want to review ALL our dependencies (700+ !)
- We are leading and learning and adapting
- Automation to scale
- Always remember the people

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/jn	https://py	https://git
107	**lazy-object-proxy**	Ionel Cristian Mărieș <con	1.10.0	A fast and thoro	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·	Secu	Pack	Vulne	Governance	Lifecycle sta	Unpatched \	Industry imp	Add Securit	Follow up w	Propose Tru	Follow up w	Propose ma	Num Action:
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.

Open Refactory: Bugs analysis

- 2 months
- 719 Packages
- 14 Bugs Reported
- 3 High, 6 Medium, 5 Low severity

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

Long term targets

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

What YOU can do?

Think security

- Contribute back security reports
- Know your dependencies
- Support security efforts of similar initiatives

Learnings?

Conclusion

Airflow’s security depends on its engagement with its 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

Let’s continue the conversation

COMMUNITY

THE ASF CONFERENCE

CODE

https://communityovercode.org/

Denver, Colorado, October 7-10, 2024

Thank You!

We’ll be around at Grand Ballroom at
2 pm to talk to us (Birds of a Feather)

Resources

SLSA

SSDF

Alpha-Omega Project

Case Study Eclipse Temurin: Pioneering Software Supply Chain Security

Open Source Software Foundation

Airflow Security

Python Security