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

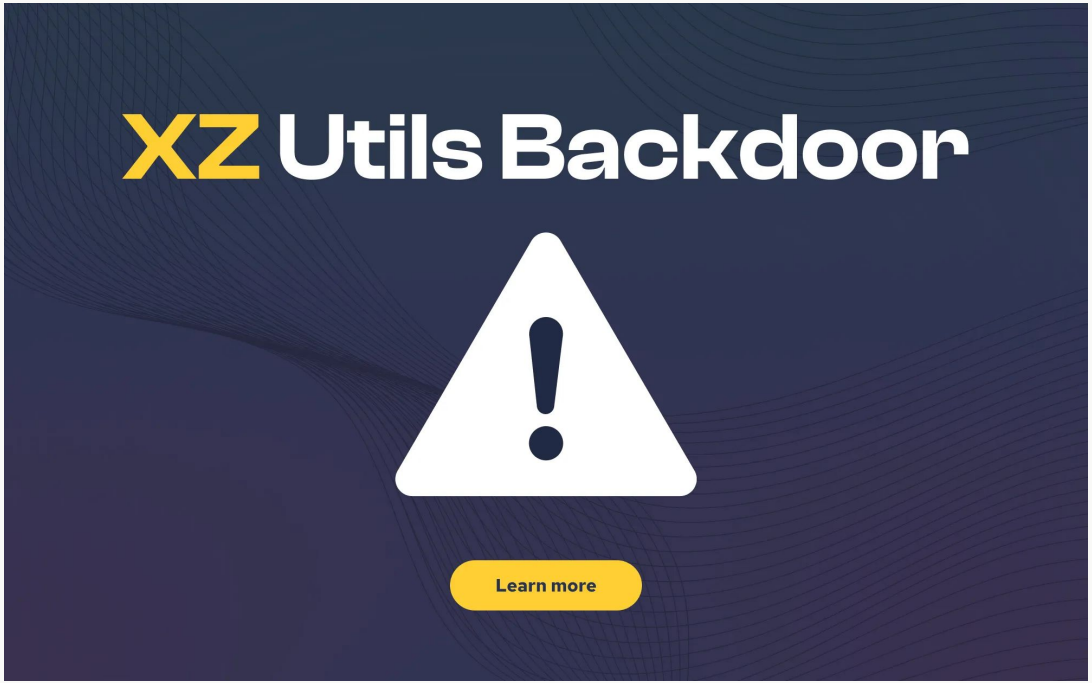
AI Libraries Airflow Apache (ASF) Clang Eclipse FreeBSD Homebrew Jenkins JQuery Linux Kernel LLVM NodeJS Omega OpenJS OpenSSL PHP Python Rubygems Rust

Alpha-Omega Explained

α → Leverage

Ω → Scale

A Wake Up Call



Open Source Supply Chain Risk

Size != Risk

Free Puppies Not Free Beer

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

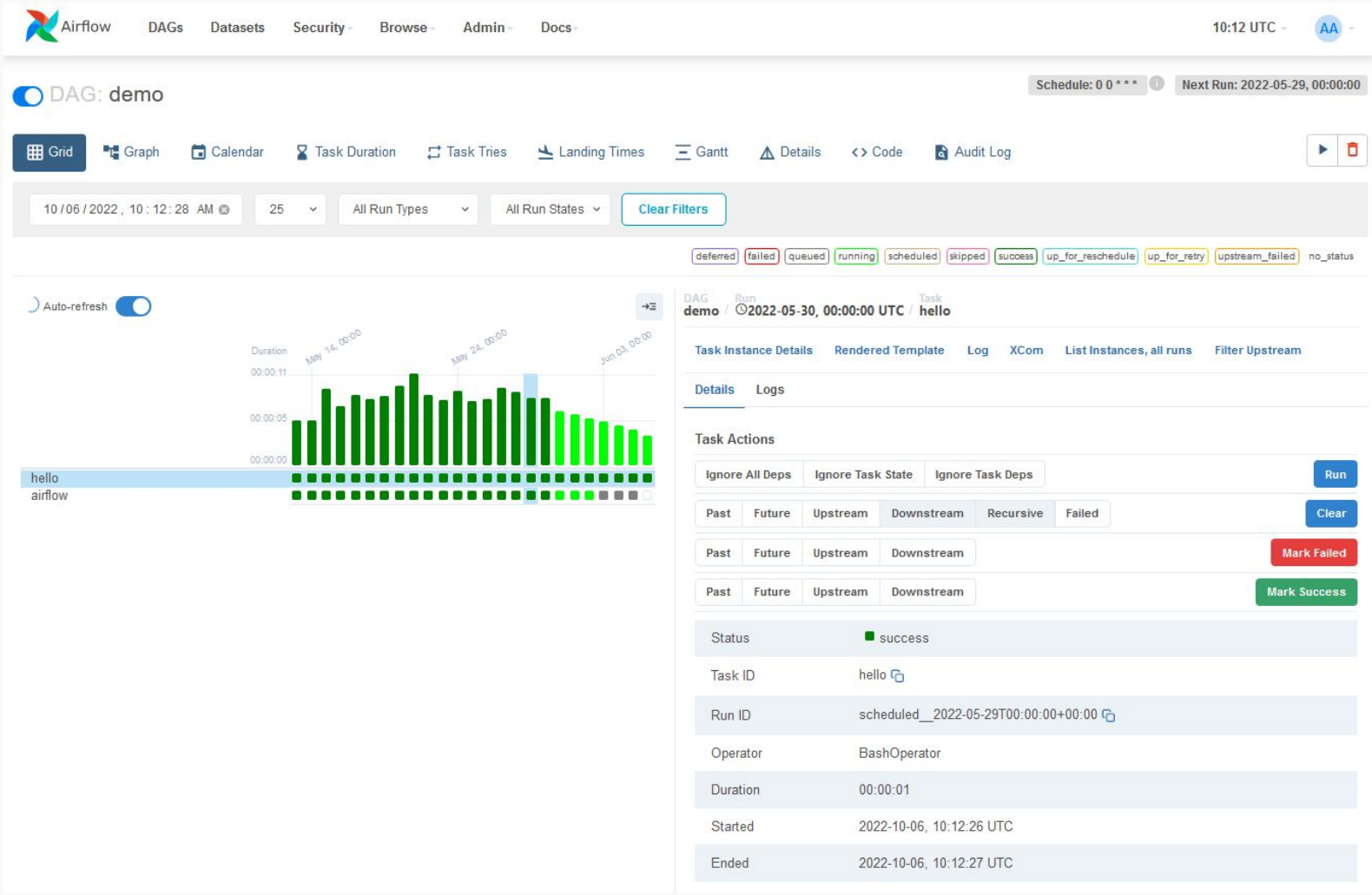
Selecting The Right Beach

Well Resourced Project **Community**

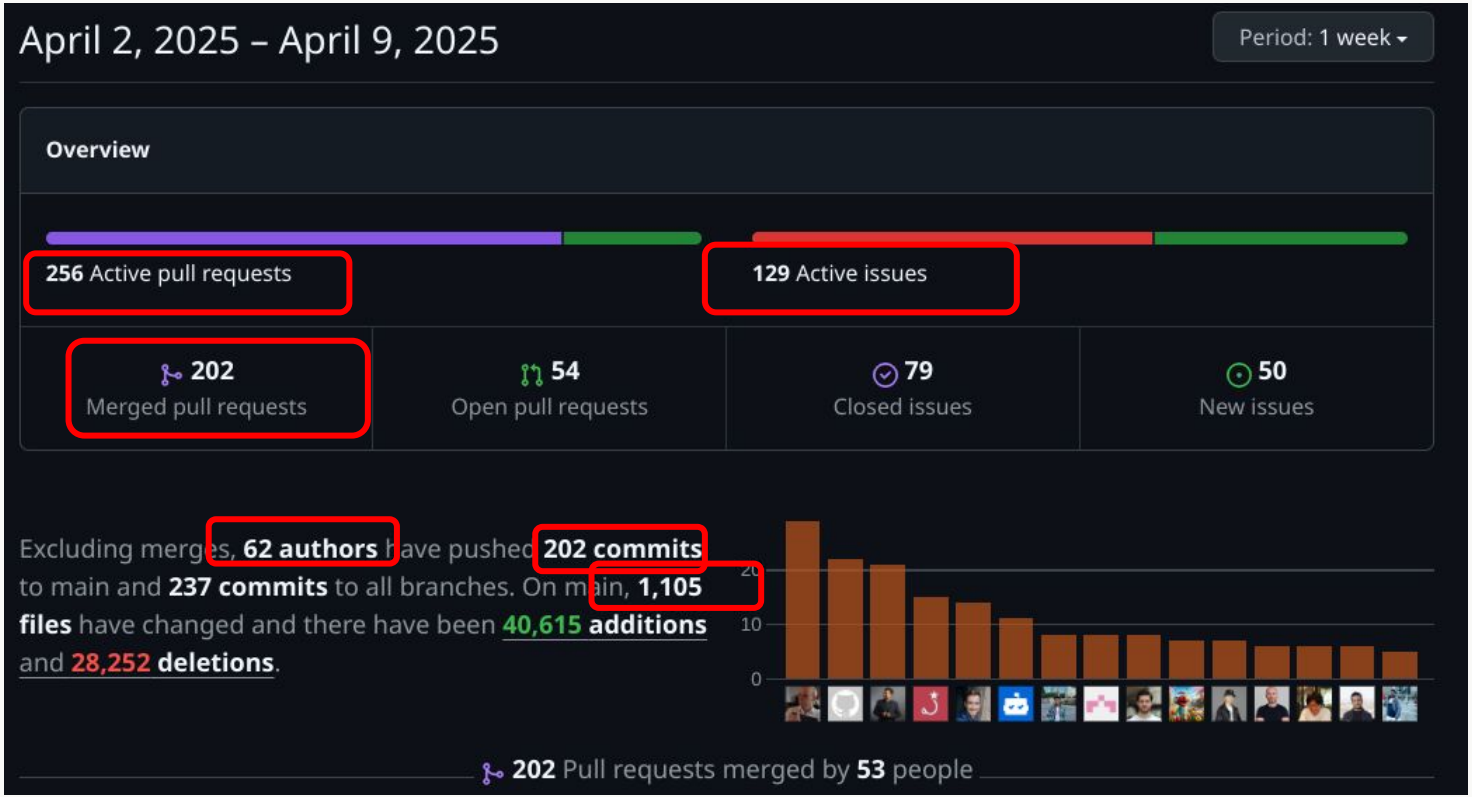
Security **Culture**

Supply Chain **Inventory**

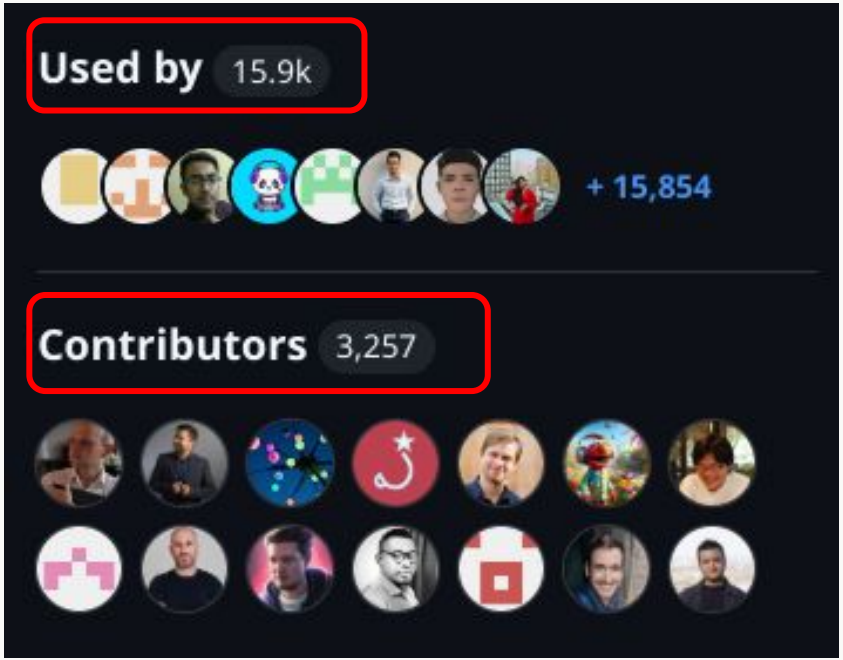
So let’s talk about Airflow Security

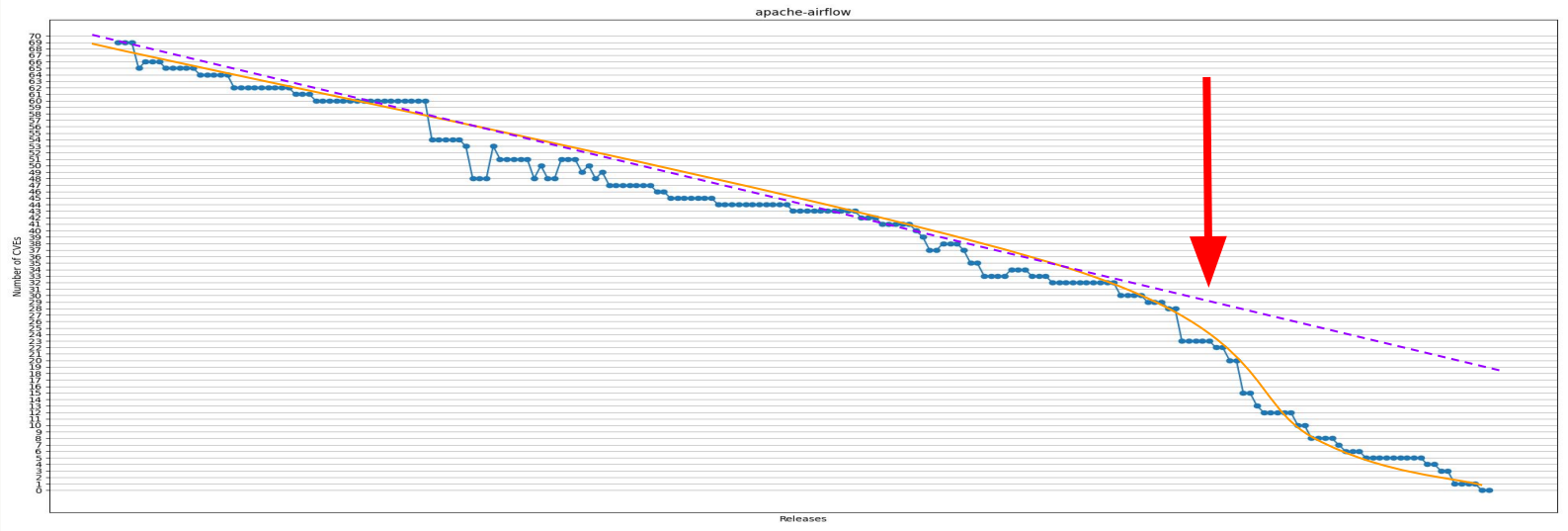

Airflow
DAGs
Datasets
Security
Browse
Admin
Docs
10:12 UTC
AA
DAG: demo
Schedule: 0 0 * * *
Next Run: 2022-05-29, 00:00:00
Grid
Graph
Calendar
Task Duration
Task Tries
Landing Times
Gantt
Details
Code
Audit Log
10 / 06 / 2022 , 10 : 12 : 28 AM
25
All Run Types
All Run States
Clear Filters
deferred
failed
queued
running
scheduled
skipped
success
up_for_reschedule
up_for_retry
upstream_failed
no_status
Auto-refresh
Duration
May 14, 00:00
May 24, 00:00
Jun 03, 00:00
00:00:11
00:00:05
00:00:00
hello
airflow
DAG
demo
Run
2022-05-30, 00:00:00 UTC
Task
hello
Task Instance Details
Rendered Template
Log
XCom
List Instances, all runs
Filter Upstream
Details
Logs
Task Actions
Ignore All Deps
Ignore Task State
Ignore Task Deps
Run
Past
Future
Upstream
Downstream
Recursive
Failed
Clear
Past
Future
Upstream
Downstream
Mark Failed
Past
Future
Upstream
Downstream
Mark Success
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

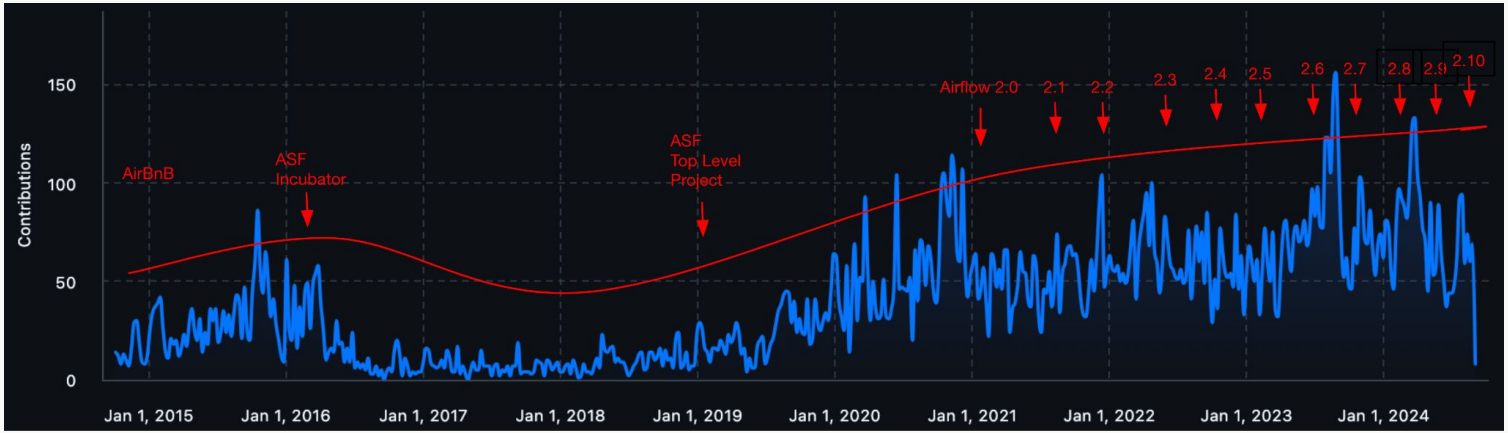
Airflow by numbers



Users and Contributors




apache-airflow
Number of CVEs
Releases


Contributions
150
100
50
0
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
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

Security improvements in 2023-2024

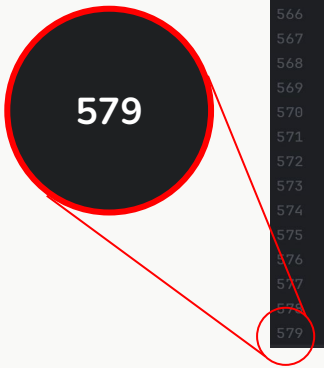
- Dedicated security team
- Created and documented detailed process
- Introduced security model
- Canned responses to issues
- Disabled inherently insecure features
- Hardened CI workflows
- Introduced reproducible builds (provenance)

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
```




```
│       └── 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%

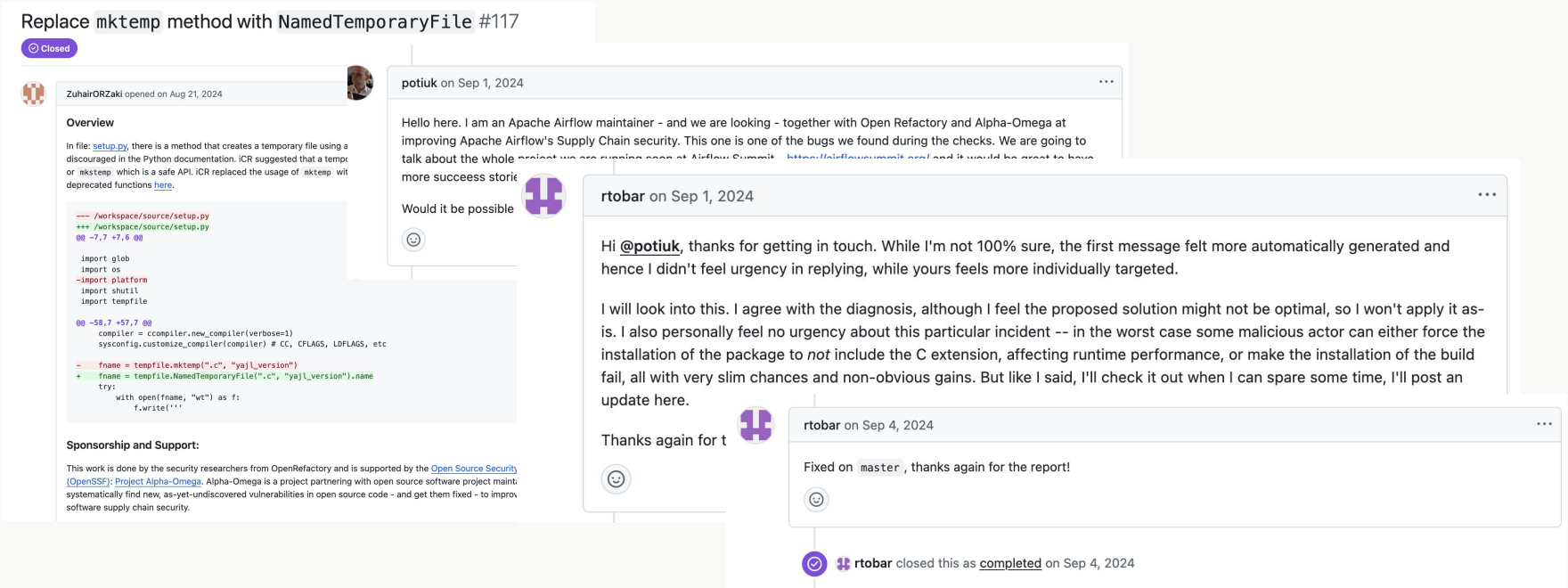
	A	B	C	D	E	F	G	H	I	J	K
1											
2	**Name**	**Author**	**Version**	**Description**	**Core**	**Devel**	**Depth**	**Licenses**	**Purl**	**Pypi**	**Vcs**
39	**python-nvd3**	Belaid Arezqui <areski@g	0.16.0	Python NVD3 - (	TRUE	FALSE	1	MIT	pkg:pypi/py	https://py	hon-nvd3
50	**unicodecsv**	Jeremy Dunck <jdunck@g	0.14.1	Python2's stdlib	TRUE	FALSE	1	0BSD	pkg:pypi/ur	https://py	icodecsv
68	**cron-descriptor**	Adam Schubert <adam.sc	1.4.3	A Python library	TRUE	FALSE	1	MIT	pkg:pypi/cr	https://py	escriptor
69	**croniter**	Matsumoto Taichi, kiorky <	3.0.3	croniter provides	TRUE	FALSE	1	MIT	pkg:pypi/cr	https://py	y/croniter
72	**deprecated**	Laurent LAPORTE <tantal	1.2.14	Python @deprec	TRUE	FALSE	1	MIT	pkg:pypi/de	https://py	precated
73	**dill**	Mike McKerns <mmckerns	0.3.8	serialize all of Py	TRUE	FALSE	1	0BSD, BSD-3-Cl	pkg:pypi/di	https://py	dation/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://py	tproctitle
165	**argcomplete**	Andrey Kislyuk <kislyuk@	3.5.0	Bash tab comple	TRUE	FALSE	1	Apache-2.0	pkg:pypi/ar	https://py	complete
166	**asgiref**	Django Software Foundati	3.8.1	ASGI specs, hel	TRUE	FALSE	1	0BSD, BSD-3-Cl	pkg:pypi/as	https://py	o/asgiref/
194	**colorlog**	Sam Clements <sam@bo	6.8.2	Add colours to th	TRUE	FALSE	1	MIT	pkg:pypi/cc	https://py	-colorlog
197	**flask-caching**	Peter Justin <peter.justin@	2.3.0	Adds caching su	TRUE	FALSE	1	0BSD	pkg:pypi/fla	https://py	-caching
228	**psutil**	Giampaolo Rodola <g.rod	6.0.0	Cross-platform li	TRUE	FALSE	1	0BSD, BSD-3-Cl	pkg:pypi/ps	https://py	olo/psutil
235	**tenacity**	Julien Danjou <julien@dai	9.0.0	Retry code until	TRUE	FALSE	1	Apache-2.0	pkg:pypi/te	https://py	d/tenacity
263	**universal-pathlib**		0.2.2	pathlib api exten	TRUE	FALSE	1	MIT	pkg:pypi/ur	https://py	al_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					Yes	Medium	Yes	Yes	Yes		Yes	4
3	0	0	-1	0						Medium	Yes		Yes	Yes	Yes	4
4.7	3	10	10	0				maintained		Medium	Yes		Yes		Yes	3
5.1	1	10	10	0				maintained		Medium	Yes		Yes		Yes	3
4.4	1	0	10	0						Medium	Yes		Yes		Yes	3
5.5	1	10	-1	10		-1	10	Reputable Foun	Actively maintained		High			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



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

SLSA

SSDF

Alpha-Omega Project

Case Study Eclipse Temurin: Pioneering Software Supply Chain Security

Open Source Software Foundation

Airflow Security

Python Security