



Google
Open Source
Security Team

The Trials and Tribulations of Bulk Converting CVEs to OSV

Andrew Pollock

I dent i fy, def i ne, and cat a lo g pu b li c ly di s cl o s e d cy b e r s e c u r i t y
v u l n e r a b i l i t i e s.

Why?

- So there's a unique identifier for vulnerabilities
- So everyone realizes they're talking about the same vulnerability
- So people can prioritize the response to the existence of the vulnerability in their environment
 - eliminate the risk by upgrading
 - eliminate the risk by mitigation
 - accept the risk or there is no actual risk
- Because vulnerabilities can have security implications

CVE-2024-23725

```
"affected": [  
  {  
    "vendor": "n/a",  
    "product": "n/a",  
    "versions": [  
      {  
        "version": "n/a",  
        "status": "affected"  
      }  
    ]  
  }  
],
```

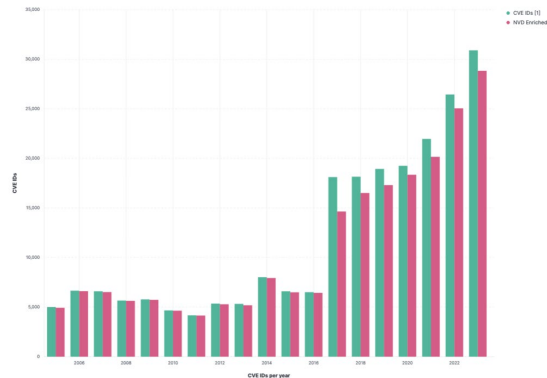
Source: <https://cveawg.mitre.org/api/cve/CVE-2024-23725>

CVE-2024-23725

```
"descriptions": [  
  {  
    "lang": "en",  
    "value": "Ghost before 5.76.0 allows XSS via a post excerpt in  
excerpt.js. An XSS payload can be rendered in post summaries."  
  }  
]
```

Source: <https://cveawg.mitre.org/api/cve/CVE-2024-23725>

Identify, define, and catalog publicly disclosed cybersecurity vulnerabilities.



Source: <https://anchore.com/blog/national-vulnerability-database-opaque-changes-and-unanswered-questions/>

Bottom line up front

- **CNAs**
 - Think about the CVEs you're authoring and their fitness for purpose, in aggregate
- **CVE Program**
 - Make it easy for CNAs to do the right thing, and harder for them to do the wrong things

\$ whoami

- Software Engineer
- Google Open Source Security Team
 - OSV
- Before
 - Systems Administrator ⇨ Site Reliability Engineer ⇨ Security Engineer
- Currently based in Brisbane, Australia
 - The rest of the OSV team is based in Sydney

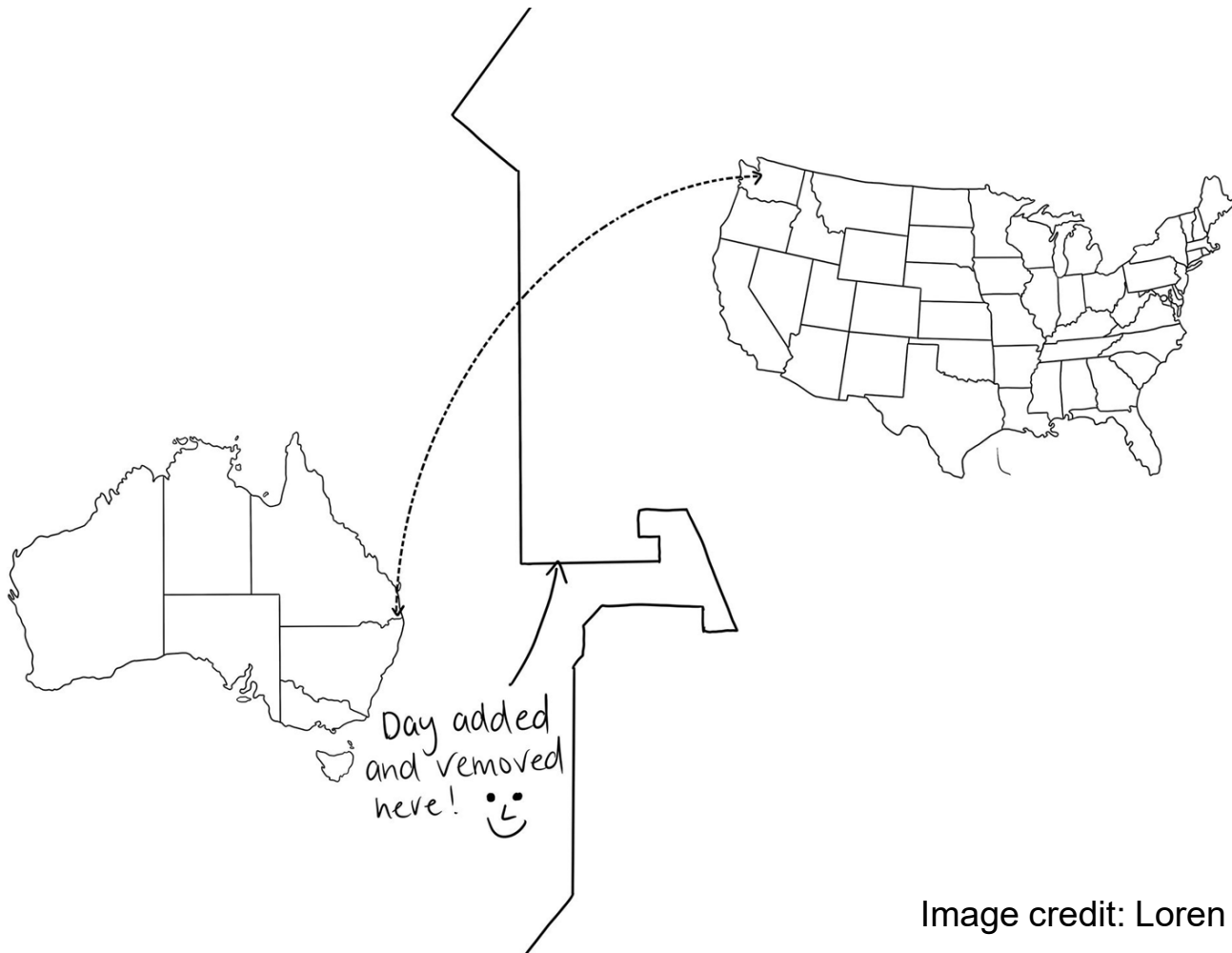
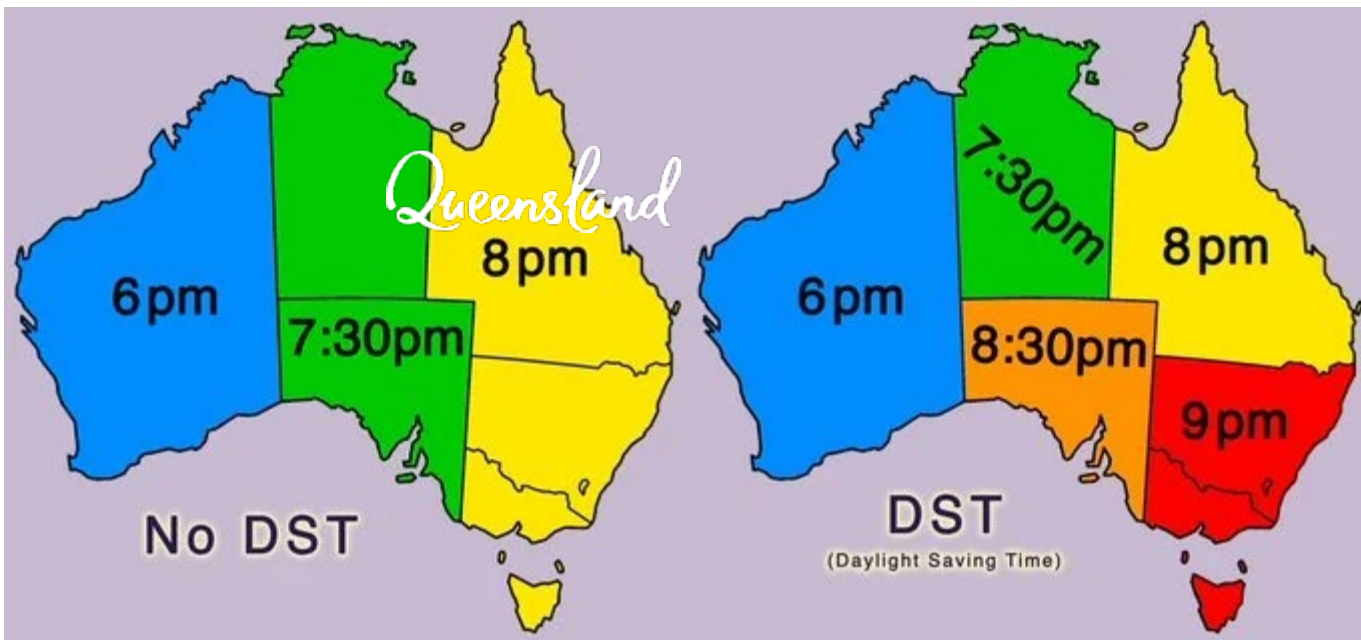


Image credit: Loren Morssinkhof



Come visit!

- AusCERT
- BSides
 - Brisbane, Melbourne, Canberra
- CrikeyCon



HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)



xkcd.com/927/

Enable developers to reduce security risk arising from known vulnerabilities in open source components they use.

OSV: Background

- Precise identification of vulnerabilities in **open source software**
 - Why? 😊 To enable prioritized vulnerability remediation
- It all began in 2021 with OSS Fuzz...
 - CVE 4 (and CPEs) couldn't express findings
 - CVE ID allocation was manual and slow
 - Automated record creation and submission was 😬
- OSV schema donated to the OpenSSF
 - github.com/ossf/osv-schema
- OSV.dev and OSV-Scanner are Google-sponsored infrastructure and open source projects
 - github.com/google/osv.dev
 - github.com/google/osv-scanner

Fun Fact

- OSV schema influenced CVE 5 schema
 - “computable open-source version information”
 - github.com/CVEProject/cve-schema/issues/87
 - theoretical interoperability between the two formats

OSV: Today

- Organic OSV schema adoption
 - 20+ ecosystems
- OSV.dev
 - Aggregates, enriches and provides a central API

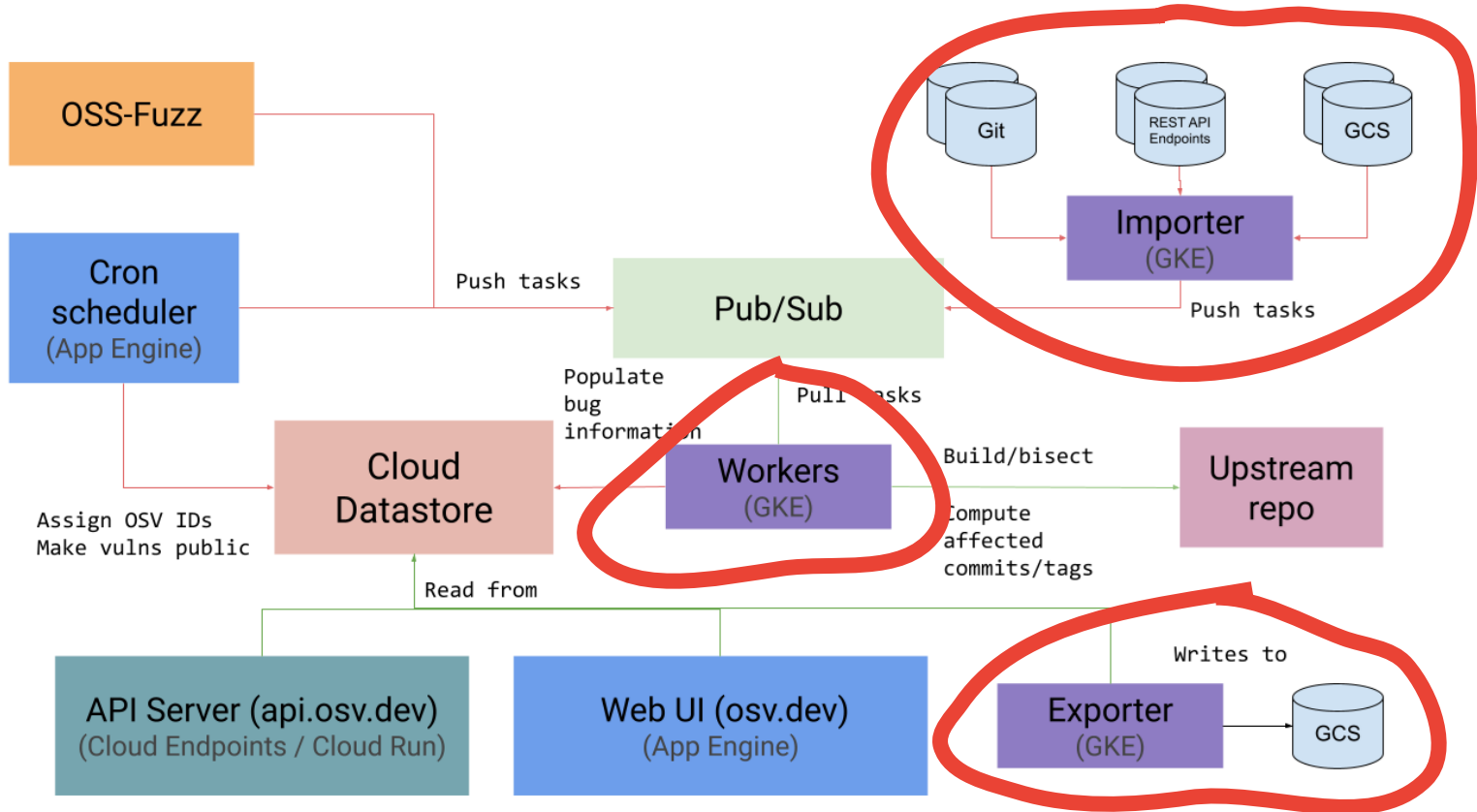
OSV: Today

- Audience: open source software developers and consumers
- First-party
 - OSV-Scanner
 - First-party API client
 - Source code vulnerability and licence scanning
 - Container image scanning (focus area for 2024)
 - Enabling end-to-end vulnerability remediation using OSV.dev's data
- Third-party
 - A growing number of integrations
 - Trivy
 - Renovate
 - Dependency-Track

OSV: Tomorrow

- Comprehensive, accurate and timely database of known vulnerabilities:
 - Scalable data quality
 - Broader symbol-level vulnerability detail
- Client-side tooling:
 - accurate vulnerability identification and prioritization
 - integrate into the software development lifecycle
 - record generation and management

OSV.dev infrastructure



“The OSV database contains 100% of vulnerabilities from NVD/CVE since 2016 that are determined to relate to OSS”

- Why? Close the coverage gap on C and C++ vulnerabilities
- Use the existence of a Git repository as a proxy for open source software
- Use the NVD as the source of CVE data
- Use the NVD as the primary source of CPE to Git repositories
 - Use Debian’s copyright metadata as a fallback

Challenge #1: figuring out **the** repository for the vulnerable software

- Pre-process the CPE Dictionary into a simpler to use Vendor/Product ⇒ Git repo mapping
- Fall back to individual CVE references that are Git repositories
- Just because it's a Git repo doesn't mean it's **the** Git repo
 - <https://github.com/keru6k/Online-Admission-System-RCE-PoC>
 - <https://github.com/laoquanshi/BILLING-SOFTWARE-SQL-injection-vulnerability>
 - <https://github.com/leekenghwa/CVE-2023-34830---Reflected-XSS-found-in-I-doit-Open-v24-and-below>

Challenge #1: figuring out **the** repository for the vulnerable software

```
"chromiumembedded:chromium_embedded_framework": [  
  "https://github.com/chromiumembedded/cef"  
]  
"gnu:glibc": [  
  "git://sourceware.org/git/glibc.git"  
],  
"isc:bind": [  
  "https://github.com/isc-projects/bind9",  
  "https://gitlab.isc.org/isc-projects/bind9"  
]
```

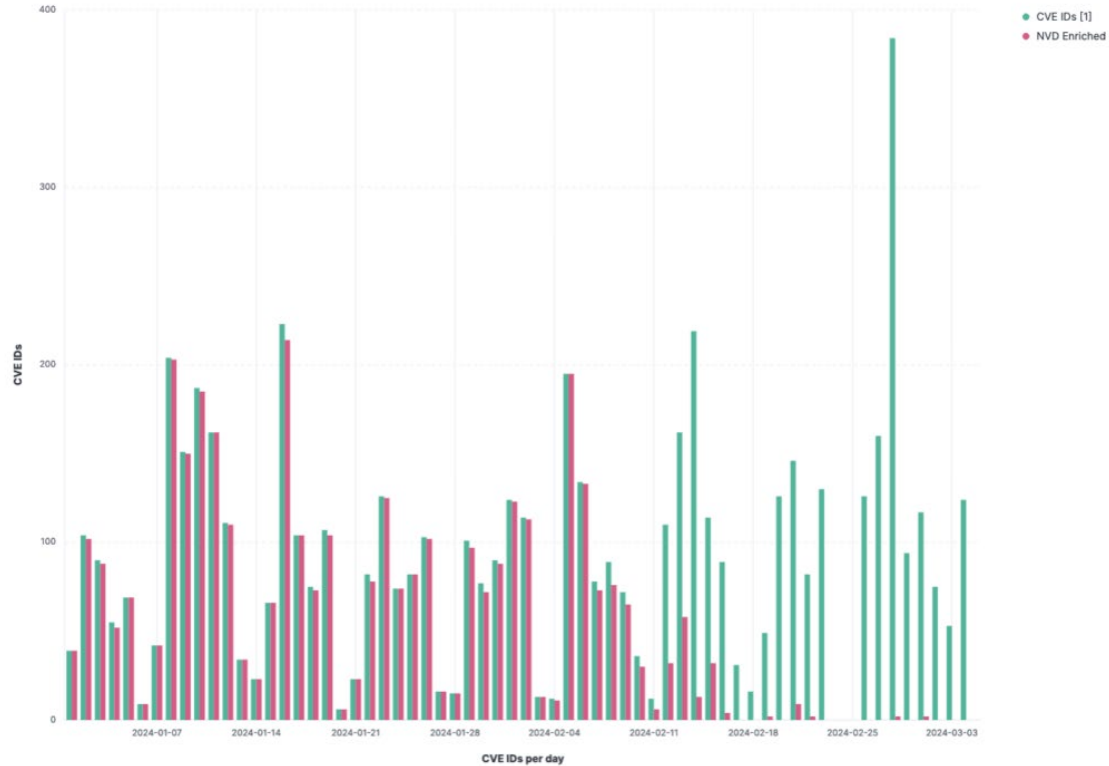
Challenge #1: figuring out *the* repository for the vulnerable software

- CPE Dictionary has its downsides
 - The good old “naming problem”
 - We’ve been contributing improvements to the CPE Dictionary
- github.com/scanoss/purl2cpe is an interesting late-breaking discovery
 - Preliminary spot checks haven’t revealed any significant gaps
- The denylist of garbage repos requires manual curation

NOTICE

NIST is currently working to establish a consortium to address challenges in the NVD program and develop improved tools and methods. You will temporarily see delays in analysis efforts during this transition. We apologize for the inconvenience and ask for your patience as we work to improve the NVD program.

Challenge #1: figuring out *the* repository for the vulnerable software



<https://anchore.com/blog/national-vulnerability-database-opaque-changes-and-unanswered-questions/>

Challenge #2: figuring out the versions

- Versions in OSV records
 - `introduced`
 - `fixed`
 - `last_affected` (less preferred)
- OSV opts for false negatives over false positives
- The NVD CVE record's applicability statement
 - the richer version details in the applicability statement
 - the CPE string
- Falling back to parsing the CVE description text 🤖

Challenge #2: figuring out the versions

Using the applicability statement's JSON

```
"cve": {
  "id": "CVE-2024-1250",
  "vulnStatus": "Analyzed",
  "descriptions":
    {
      "value": "An issue has been discovered in GitLab EE affecting all versions starting from 16.8 before 16.8.2. When a user is assigned a custom role with manage_group_access_tokens permission, they may be able to create group access tokens with Owner privileges, which may lead to privilege escalation."
    },
  "configurations": [
    "cpeMatch": [
      {
        "vulnerable": true,
        "criteria": "cpe:2.3:a:gitlab:gitlab:*:*:*:*:enterprise:*:*:*",
        "versionStartIncluding": "16.8.0",
        "versionEndExcluding": "16.8.2",
      }
    ]
  ],
}
```

Challenge #2: figuring out the versions

Using the CPE string

```
"cve": {
  "id": "CVE-2024-0207",
  "vulnStatus": "Analyzed",
  "descriptions": [
    {
      "value": "HTTP3 dissector crash in Wireshark 4.2.0 allows denial of service via packet injection or
crafted capture file"
    },
  ],
  "configurations": [
    "cpeMatch": [
      {
        "vulnerable": true,
        "criteria": "cpe:2.3:a:wireshark:wireshark:4.2.0:*:*:*:*:*:*:*:*"
      }
    ]
  ]
}
```

Challenge #2: figuring out the versions

Resorting to parsing the description text 🤖

```
"cve": {
  "id": "CVE-2024-28757",
  "vulnStatus": "Awaiting Analysis",
  "descriptions": [
    {
      "value": "libexpat through 2.6.1 allows an XML Entity Expansion attack when there is isolated use of external parsers (created via XML_ExternalEntityParserCreate).",
    },
  ],
  "references": [
    {
      "url": "https://github.com/libexpat/libexpat/issues/839",
    },
    {
      "url": "https://github.com/libexpat/libexpat/pull/842",
    }
  ]
}
```

Challenge #3: getting to a commit

- Sometimes there's a (presumed) fix commit as a reference on the CVE
- The moral equivalent of a whole lot of `git ls-remote`
- Fuzzy matching on tags
 - If they're being used 🤔

Challenge #3: getting to a commit

for CVE-2024-1250 wanted: 16.8.0 and 16.8.2

```
$ git ls-remote -t https://gitlab.com/gitlab-org/gitlab
1e912d57d5a8f1135f4d41e25469069790134d41      refs/tags/v16.8.0-ee^{ }
e3c23d67e9ce2f074cd79b753ef95291da459a93      refs/tags/v16.8.2-ee^{ }
```

for CVE-2024-0207 wanted: 4.2.0


```
$ git ls-remote -t https://github.com/wireshark/wireshark
54eedfc63953c8180b5a9c60015917cce7a2548a      refs/tags/v4.2.0^{ }
```

for CVE-2024-28757 wanted: 2.6.1

```
$ git ls-remote -t https://github.com/libexpat/libexpat
a590b2d5846865412182805b853dd91d18f38c8d      refs/tags/R_2_6_1^{ }
```

Challenge #3a: the perils of jumping ahead to a commit

- There are commit hashes and there are commit hashes
- Often the commit hashes in CVE references are from a fork
- This makes our programmatic Git repository analysis very sad

 This commit does not belong to any branch on this repository, and may belong to a fork outside of the repository.

CNAs: ensuring that references belong to the actual repository and not a fixer's fork of it would be 🙌

Putting it all together

```
{
  "id": "CVE-2024-1250",
  "details": "An issue has been discovered in GitLab EE affecting all versions starting from 16.8 before 16.8.2. When a user is assigned a custom role with manage_group_access_tokens permission, they may be able to create group access tokens with Owner privileges, which may lead to privilege escalation.",
  "affected": [
    {
      "ranges": [
        {
          "type": "GIT",
          "repo": "https://gitlab.com/gitlab-org/gitlab",
          "events": [
            {
              "introduced": "1e912d57d5a8f1135f4d41e25469069790134d41"
            },
            {
              "fixed": "e3c23d67e9ce2f074cd79b753ef95291da459a93"
            }
          ]
        }
      ]
    }
  ],
}
```


Putting it all together

```
"id": "CVE-2024-28757",
  "details": "libexpat through 2.6.1 allows an XML Entity Expansion attack when there is
isolated use of external parsers (created via XML_ExternalEntityParserCreate).",
  "affected": [
    {
      "ranges": [
        {
          "type": "GIT",
          "repo": "https://github.com/libexpat/libexpat",
          "events": [
            {
              "introduced": "0"
            },
            {
              "last_affected": "a590b2d5846865412182805b853dd91d18f38c8d"
            }
          ]
        }
      ]
    }
  ]
}
```

Putting it all together

```
{
  "id": "CVE-2024-0207",
  "details": "HTTP3 dissector crash in Wireshark 4.2.0 allows denial of service via packet
injection or crafted capture file",
  "affected": [
    {
      "ranges": [
        {
          "type": "GIT",
          "repo": "https://github.com/wireshark/wireshark",
          "events": [
            {
              "introduced": "0"
            },
            {
              "last_affected": "54eedfc63953c8180b5a9c60015917cce7a2548a"
            }
          ]
        }
      ]
    }
  ],
}
```

Current Conversion Metrics

- Precomputed CPE Vendor/Product combinations with repos
 - 9,467
- Unique precomputed repos
 - 8,812

Current Conversion Metrics

Year	Total	In Scope	Converted	Percentage
2016	10,547	2,608	1,759	67%
2017	16,979	4,215	2,802	66%
2018	17,347	5,557	3,153	57%
2019	16,973	4,890	3,219	66%
2020	20,439	6,140	4,144	67%
2021	22,096	6,533	4,450	68%
2022	24,549	7,112	5,026	71%
2023	25,492	7,076	4,838	68%
2024	2,329	994	529	53%

Potential further work

- Resolving GitHub Pull Requests to commits
- Looking up the Pull Request from a GitHub Issue
- Similar things for GitLab

Andrew's Data Quality Wishlist

- “Zaroo Boogs” thanks to better software development practices
- The NVD outputs CVE 5 records
- CNAs are uniformly using CVE 5 to its full potential
 - CPEs
 - Any versions/commits in the description or references are also in the **affected** field
 - Canonical source repository details are in the **affected** field
 - Schema validation and RSUS enforce a higher minimum quality bar
 - RSUS performs lint checks
- References to GitHub commits are from within the canonical repository for the related software
- Comprehensive, open and free mapping between CPEs, Purls and canonical Git repositories
- The CVE Program has a standardised way to improve existing records

Bright spots: CVE-2024-21733

```
"descriptions": [  
  {  
    "lang": "en",  
    "value": "Generation of Error Message Containing Sensitive Information  
vulnerability in Apache Tomcat.This issue affects Apache Tomcat: from  
8.5.7 through 8.5.63, from 9.0.0-M11 through 9.0.43.\n\nUsers are  
recommended to upgrade to version 8.5.64 onwards or 9.0.44 onwards, which  
contain a fix for the issue.\n\n"  
  }  
],
```

Source: <https://cveawg.mitre.org/api/cve/CVE-2024-21733>

Bright spots: CVE-2024-21733

```
"affected": [  
  {  
    "defaultStatus": "unaffected",  
    "product": "Apache Tomcat",  
    "vendor": "Apache Software Foundation",  
    "versions": [  
      {  
        "lessThanOrEqual": "8.5.63",  
        "status": "affected",  
        "version": "8.5.7",  
        "versionType": "semver"  
      },  
      {  
        "lessThanOrEqual": "9.0.43",  
        "status": "affected",  
        "version": "9.0.0-M11",  
        "versionType": "semver"  
      }  
    ]  
  }  
],
```


Bright spots: CVE-2024-1250

```
"affected": [  
  {  
    "vendor": "GitLab",  
    "product": "GitLab",  
    "repo": "git://git@gitlab.com:gitlab-org/gitlab.git",  
    "versions": [  
      {  
        "version": "16.8",  
        "status": "affected",  
        "lessThan": "16.8.2",  
        "versionType": "semver"  
      }  
    ],  
    "defaultStatus": "unaffected"  
  }  
],
```



Special mention: Linux Kernel CVEs (CVE-2024-26634)

```
"descriptions": [  
  {  
    "lang": "en",  
    "value": "In the Linux kernel, the following vulnerability has been  
resolved:\n\nnet: fix removing a namespace with conflicting altnames\n\nMark  
reports a BUG() when a net namespace is removed.\n\nkernel BUG at  
net/core/dev.c:11520!\n\nPhysical interfaces moved outside of init_net get  
\n\"refunded\"\n\nto init_net when that namespace disappears. The main  
interface\n\nname may get overwritten in the process if it would  
have\n\nconflicted. We need to also discard all conflicting altnames.\n\nRecent  
fixes addressed ensuring that altnames get moved\n\nwith the main interface,  
which surfaced this problem."  
  }  
],
```

Special mention: Linux Kernel CVEs

```
"affected": [  
  {  
    "product": "Linux",  
    "vendor": "Linux",  
    "defaultStatus": "unaffected",  
    "repo": "https://git.kernel.org/pub/scm/linux/kernel/git/stable/linux.git",  
    "versions": [  
      {  
        "version": "673edcffa096",  
        "lessThan": "a2232f29bf52",  
        "status": "affected",  
        "versionType": "git"  
      },  
      {  
        "version": "7663d522099e",  
        "lessThan": "e855dded4b70",  
        "status": "affected",  
        "versionType": "git"  
      },  
    ]  
  },  
]
```

Bottom line ~~up-front~~ towards the end

- **CNAs**
 - Think about the CVEs you're authoring and their fitness for purpose, in aggregate
- **CVE Program**
 - Make it easy for CNAs to do the right thing, and harder for them to do the wrong things

OSV.dev's data quality story for 2024

- Shifting left: tooling for record validation at creation time
- Machine-readable feedback for records that fail validation
- Enforcing schema validation at import time
- Surfacing known import failures to users at vulnerability search time
- “Don’t shoot the messenger!”

Collaborating with us

- OpenSSF Vulnerability Disclosures Working Group
 - github.com/ossf/wg-vulnerability-disclosures
 - #wg_vulnerability_disclosures_wg in <https://slack.openssf.org/>
- Mailing list
 - groups.google.com/g/osv-discuss
 - osv-discuss@googlegroups.com
- github.com/openssf/osv-schema
- github.com/google/osv.dev
- github.com/google/osv-scanner