

Deep dive in VulnerableCode v31 Q1 2023



- ▷ Databases of known FOSS software vulnerabilities are mostly proprietary and/or privately maintained.
- ▷ **Why not open data? Open source likes open!**
- ▷ Find how we are working to build new FOSS tools to:
 - Aggregate and publish software component vulnerability data from multiple sources and
 - Automate the search for FOSS component security vulnerabilities.
 - With open code and open data.
- ▷ The benefit will be improved security of software applications with **open tools and open data for everyone.**

Agenda

- ▷ Software Supply Chain Security
- ▷ The Problem(s)
- ▷ The VulnerableCode Solution
- ▷ New VulnerableCode Features in v31
- ▷ Demo
- ▷ VulnerableCode Next Steps
- ▷ References

Software Supply Chain Security

- Obviously, a huge topic
- Linux Foundation / OSSF has identified 3 focus areas:
 - Securing OSS Production
 - **Improving Vulnerability Discovery & Remediation**
 - Shorten Ecosystem Patching Response Time
- nexB focus is on using SCA to **identify software vulnerabilities, their impact and remediation**

The Problem

- There is no open code AND open data **QUALITY** vulnerability DB
- Security vendors **DO NOT WANT TO SOLVE** the problem
- Race to the **BIGGEST** database of @#\$!*
- Expensive **and** poor quality
- Existing "free" database are problematic with opaque review process
 - And poor data quality

The Quality Problem

- DBs are making up packages that do not exist
 - No big deal, none uses them... but this shatters confidence
- DBs invent vulnerable ranges
- They do not agree on vulnerable ranges
- "Cry wolf" with dubious vulnerabilities
- Claim to have "secret" "premium" vulnerabilities
- Do not publish their findings upstream at the NVD to share back

Meanwhile at Security (1)

What I would hope for: a tidy and organized security gate



Credit: <https://www.flickr.com/photos/oddharmonic/4756905580> " oddharmonic Security at Denver International Airport

License: CC-BY 2.0 <https://creativecommons.org/licenses/by/2.0/>

Meanwhile at Security (2)

The reality:



Credit: Arne Mueseler <http://arne-mueseler.com/> https://commons.wikimedia.org/wiki/File:Loveparade_2010_duisburg_tunnel_ramp.jpg

License: CC-BY-SA-3.0 https://en.wikipedia.org/wiki/en:Creative_Commons

From https://en.wikipedia.org/wiki/Love_Parade_disaster

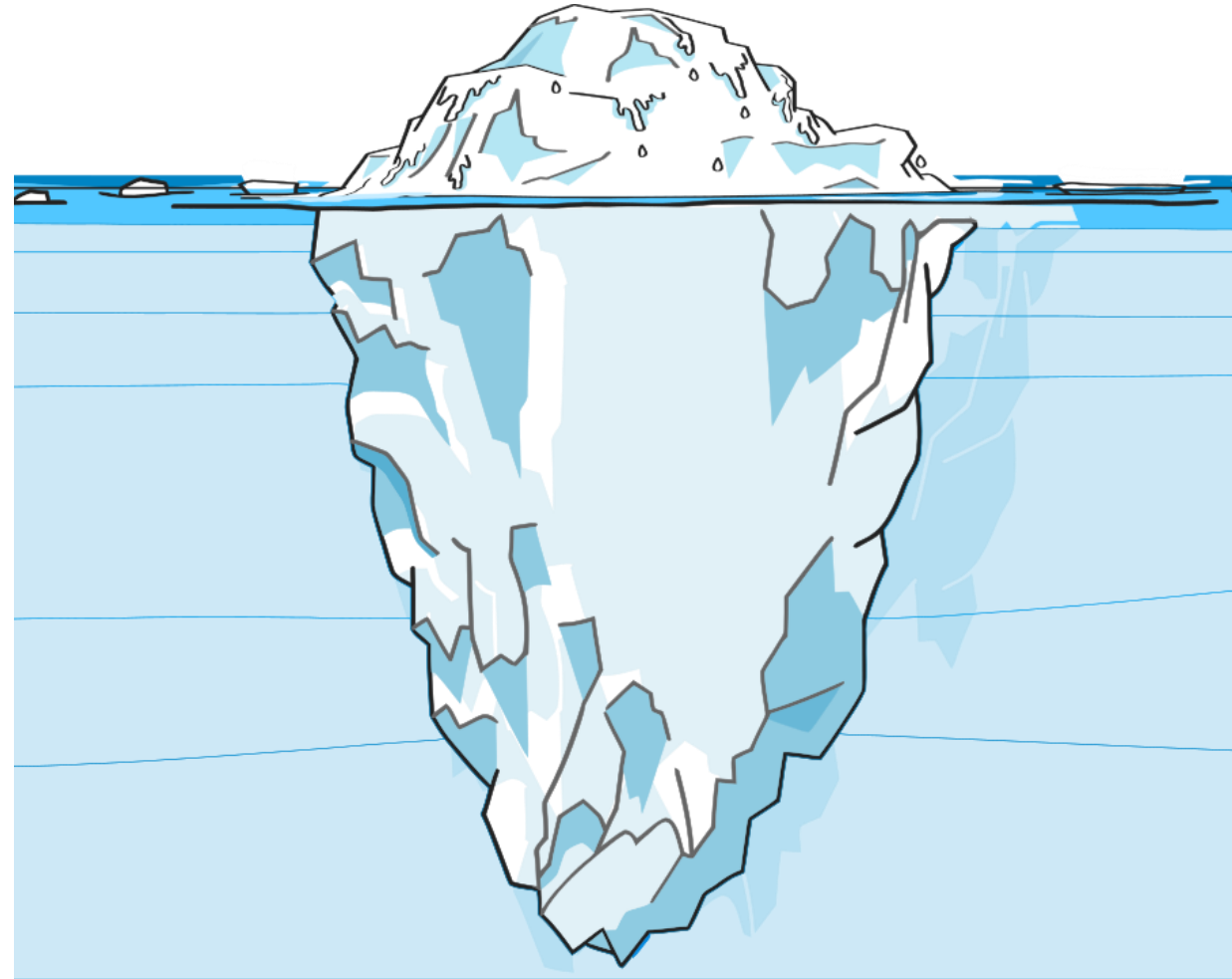
On 24 July 2010, a crowd disaster at the 2010 Love Parade electronic dance music festival in Duisburg, North Rhine-Westphalia, Germany, caused the deaths of 21 people from suffocation as attendees sought to escape a ramp leading to the festival area.[1] 652 people were injured.

Our Solution

VulnerableCode DB with GUI, API and VulnTotal

And below the water line:

- Many data sources
- Import from upstream!
- PURL, package first!
- Cleaned, combined and merged
- All data licenses are tracked
- Smart vulnerable version ranges (Also used in OSV)
- Bots to improve the data



Credits: MoteOo <https://pixabay.com/illustrations/iceberg-above-water-white-cold-3273216/> and openclipart by <https://www.openclipart.org/detail/299871/tip-of-the-iceberg>
License: <https://scancode-licensedb.aboutcode.org/pixabay-content.html>

VulnerableCode v31 new things (1)

- A Web UI!
- Public instance with enhanced API access
- Faster API for bulk search
- New batch query capabilities
- Documentation!
- VulnTotal, a vulnerability DB comparison shopping
- Total rework of internal models and importers code
 - No more big batches
 - Support for version ranges "vers" spec which we designed as part of PURL

VulnerableCode v31 new things (2)

- Re-enabled, and refined many data importers
 - postgresql
 - pypa
 - archlinux
 - ubuntu
 - debian
 - npm
 - retiredotnet
 - apache_httpd

VulnerableCode v32 is coming right after

- Re-enabled many importers, completing migration
 - mozilla, gentoo, istio, Eclipse steady, suse, elixir, apache tomcat
- Added support for CWE.
- Improved handling of CVSS scores
- License documentation for all importers

Can you prove this
is better?

W. Edwards Deming said:

“In God we trust.
All others must
bring data.”

VulnTotal: Comparison shopping for VDB

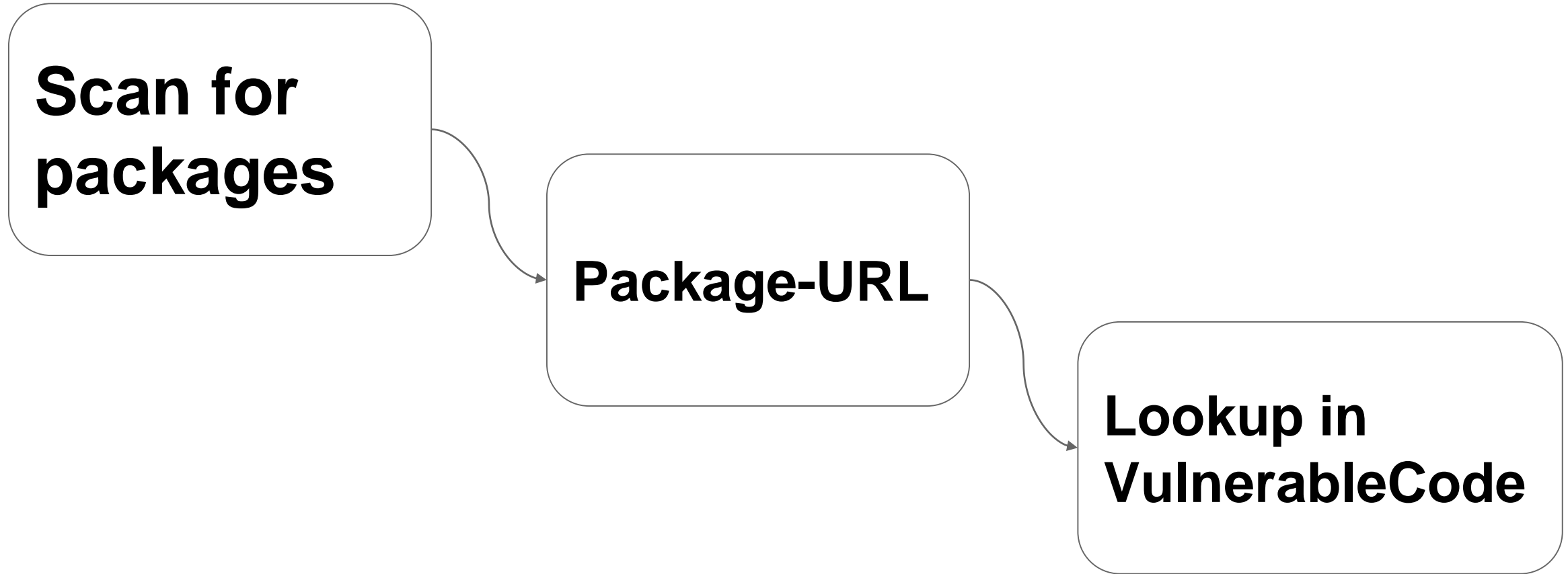
- By VulnerableCode team member, Keshav Priyadarshi

The screenshot shows a Google Shopping search for "Vulnerability Database". The search results are sorted by relevance and show four items, all of which are USB cables. Each item is associated with a specific CVE. The items are:

- Item 1:** UGREEN USB-A to USB-B cable. Price: \$5.99. CVE: CVE-2021-44228. Seller: Amazon.com - Seller. Delivery: \$5.99 delivery. Trusted store.
- Item 2:** Micro USB cable. Price: \$0.51. CVE: CVE-2021-44228. Seller: DHgate online store. Delivery: Free delivery.
- Item 3:** USB-A to Micro USB cable. Price: \$1.47. CVE: CVE-2021-45046. Seller: AliExpress.com. Delivery: Free delivery.
- Item 4:** Micro USB cable. Price: \$1.99. CVE: CVE-2020-9488. Seller: DHgate online store. Delivery: Free delivery.

The left sidebar contains filters for "Show only" (On sale, Buy on Google), "Price" (Up to \$3, \$3 - \$6, \$6 - \$10, Over \$10), "Color" (Black), and "Connection Type" (USB-C, Micro USB).

The workflow



VulnerableCode.io Demo

- Packages
 - Affected by
 - Fixing
- Vulnerabilities
- VulnTotal

- I created Package URL for VulnerableCode and ScanCode

The critical GLUE between all the software supply chain tools

- Package URL project: <https://github.com/package-url>
 - Spec is at: <https://github.com/package-url/purl-spec>
 - Implementations for .NET, Go, Java, JavaScript, PHP, Python, Ruby and Rust
- Recent proposal to add purl to NVD:
 - <https://owasp.org/blog/2022/09/13/sbom-forum-recommends-improvements-to-nvd.html>

But wait!

- If I alone control this data, I am adopting the same flawed ways as others before us.
- We need to make this a **shared resource for everyone**, that's under the shared control of EVERYONE
- **You can already self-host VulnerableCode**
- **Working next on a new way to federate and decentralize this**

- Collect the **fix commits** for call graph or dynamic analysis
- Design an **Advisory clarity score** and of their providers
- Make VulnTotal work as a browser extension with zero install
- **More, smarter improvers**
 - Check if package exists, validate all ranges are correct
 - Mine the graph to establish correlations
- Natural language parsing of vulnerability descriptions and advisories
- Extract unpublished vulnerabilities from commit histories and trackers
- **Federated and crowdsourced** vulnerability curation
- **NVDR**: Universal non-vulnerable dependency resolvers

NVDR: keep the barbarians at the gate!

- By VulnerableCode maintainer, Tushar Goel
- If you could blend
 - **Functional dependency constraints**
 - **Knownvulnerable ranges**
- And inject these in a package dependency resolver
- You get

Non Vulnerable Dependency Resolution!

- Working PoC implemented in **python-inspector** tool
and paper https://www.tdcommons.org/dpubs_series/5224/

- ▷ Help us make this work for everyone and for YOU

You MUST fund this project to build security commons

- ▷ We received grants from the European Union through the NGI-0 program and NLnet



Contact me at pombredanne@nexb.com

References

- VulnerableCode
- ScanCode
- Package-URL (purl)
- AboutCode

- Collect and aggregate vulnerability data from many public sources
 - Projects, GitHub, Linux Distros, NVD, Package managers and others
 - Focus on upstream projects (source of the source)
- Apply confidence-based system
 - not all data are equally trusted and of equivalent quality
- Find anomalies using correlations and comparisons
- Discover relations between vulnerabilities and packages from mining the graph
- Public.VulnerableCode.io database
- See <https://nexb.com/vulnerablecode/> for more information

- o Package URL project: <https://github.com/package-url>
 - Spec is at: <https://github.com/package-url/purl-spec>
 - Implementations for .NET, Go, Java, JavaScript, PHP, Python, Ruby and Rust
- o Recent proposal to add purl to NVD:
 - <https://owasp.org/blog/2022/09/13/sbom-forum-recommends-improvements-to-nvd.html>

- Identify FOSS and other third-party components & packages
- Detect licenses, copyrights and dependencies
- ScanCode Projects include:
 - ScanCode.io: Server system with customizable pipelines
 - ScanCode Toolkit: Scanning engine - use as CLI or library
 - LicenseDB: 2000+ licenses detected by ScanCode
 - ScanCode Workbench: Desktop app to review Scans
- See <https://nexb.com/scancode/> for more information

- o AboutCode is a virtual org for our collection of FOSS SCA tools
 - Also, the home for our GSoC projects
 - And we are on OpenCollective at:
<https://opencollective.com/aboutcode>
- o Our projects are at: <https://github.com/nexB>
- o Documentation for each project is at ReadTheDocs.org
- o AboutCode home: <https://www.aboutcode.org/>

Special thanks to all the people who made and released these excellent free resources:

- Presentation template by SlidesCarnival at <https://www.slidescarnival.com/> licensed under CC-BY-4.0 <https://www.slidescarnival.com/terms-of-use#templates-license>
- Photographs by Unsplash <https://unsplash.com/license> licensed under the unsplash license <https://scancode-licensedb.aboutcode.org/unsplash.html>

All the open source software authors that made VulnerableCode, ScanCode and other AboutCode FOSS projects possible