# Scan code to reuse code safely, with ScanCode

# Agenda

AboutCode

- About me, nexB and AboutCode
- (SCA) Software Composition Analysis
  - Why scan code? Package identification and standards
  - Why is software license, quality and versions important?
- Code Scanning concepts, problems and solutions
  - Package types, identification and dependency resolvers
  - SBOMs, Automation and SCA, why use FOSS?
- AboutCode Stack:
  - ScanCode, VulnerableCode, Dejacode, PurlDB
  - Who is using these tools?
- Demo and Questions

# About Ayan

AboutCode

- Core maintainer of ScanCode
  - also contributes to and helps maintain other AboutCode tools: license-expression licenseDB scancode-workbench PURLdb
- Working primarily on License detection and Package identification, data summarization and visualization
- Google Summer of Code Mentor at AboutCode
  - participant in GSoC2020 and GSoD2019
- Software developer and Analyst at nexB, Inc.
  - asmahapatra@nexb.com
  - GitHub: https://github.com/AyanSinhaMahapatra/
  - LinkedIn: https://www.linkedin.com/in/ayansinhaju/

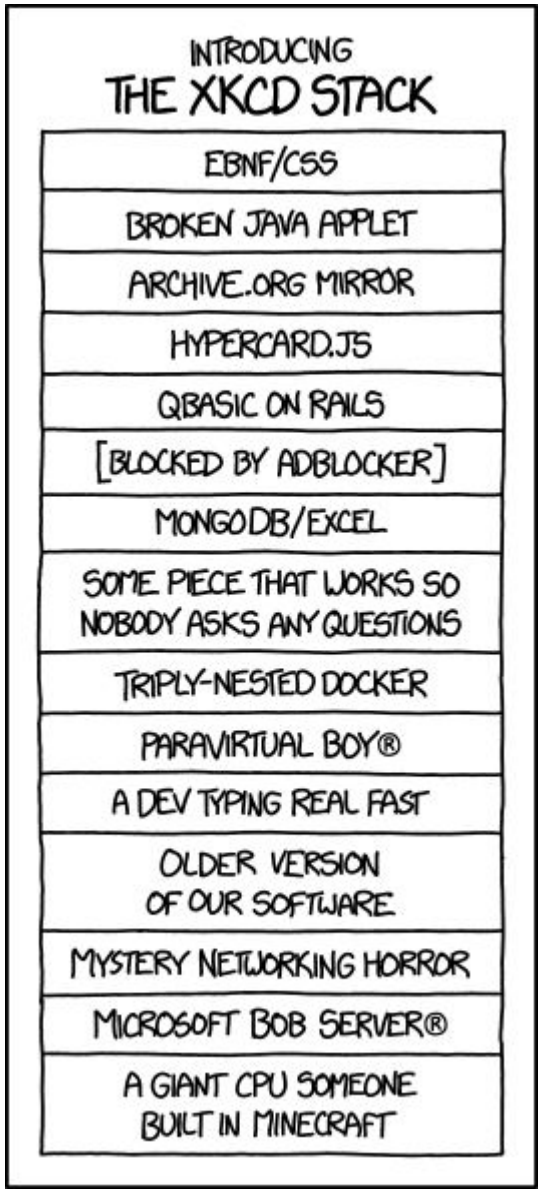
# AboutCode and nexB

AboutCode

- AboutCode's FOSS-first mission: FOSS for FOSS
  - Open source tools and open knowledge base  (AboutCode stack)
  - Simple and practical standards (Package-URL)
  - Applications for Legal Business users (DejaCode, also FOSS) with APIs
- Trusted experts on Software Composition Analysis (SCA) since 2007
  - Creator of Package-URL: https://github.com/package-url
  - Co-founders of SPDX: https://spdx.org
  - Contributors to CycloneDX: https://cyclonedx.org
  - Co-founders of ClearlyDefined: https://clearlydefined.io
- nexB: professional services for SCA
  - 800+ SCA projects completed to-date
  - Sponsored development for AboutCode projects
  - Technical support and advisory for SCA process, and deployments

# The problem with modern software

- Ever more FOSS software packages are reused
  - small apps routinely embed 500 FOSS packages
  - large apps: 10,000!
- Everyday you have new vulnerabilities, license problems and package updates in your package dependency trees
  - Impossible to check this manually!
- Goal: Discover the problems and help alleviate the pain




Source: https://xkcd.com/1636/

# Why is Software License important?

- FOSS: Freedom
- Freedom and Responsibilities
  - Can we use the software in different scenarios?
  - Can we modify and redistribute freely, under my choice of terms?
  - Give credit, generate attribution
- See License categories for more details
- Copyrights:
  - Copyright notices often have to be included and redistributed
- History of Litigation

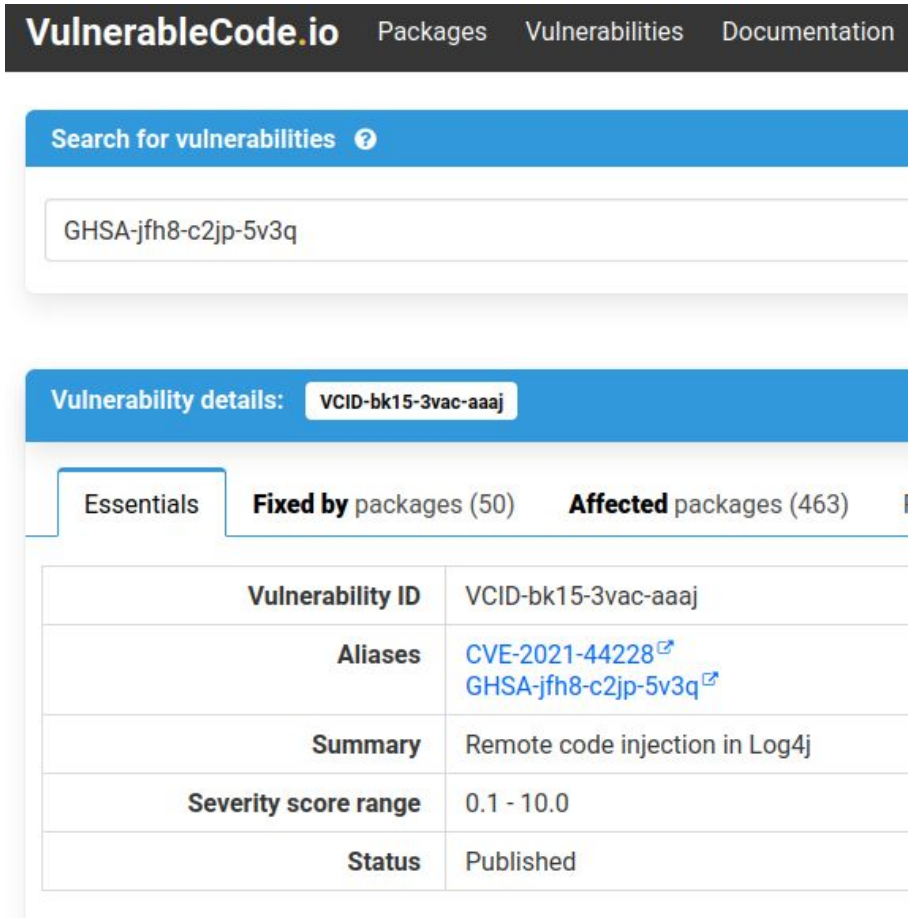
# Why is identification important?

- Modifications can be released under different terms
- License could change between versions
  - packages/products often decide to change their license
    - https://redis.com/blog/redis-adopts-dual-source-available-licensing/
    - https://www.elastic.co/blog/elastic-license-update


AboutCode

# Why is identification important?

- Vulnerabilities are introduced and fixed by versions (or not!)
- False positives!

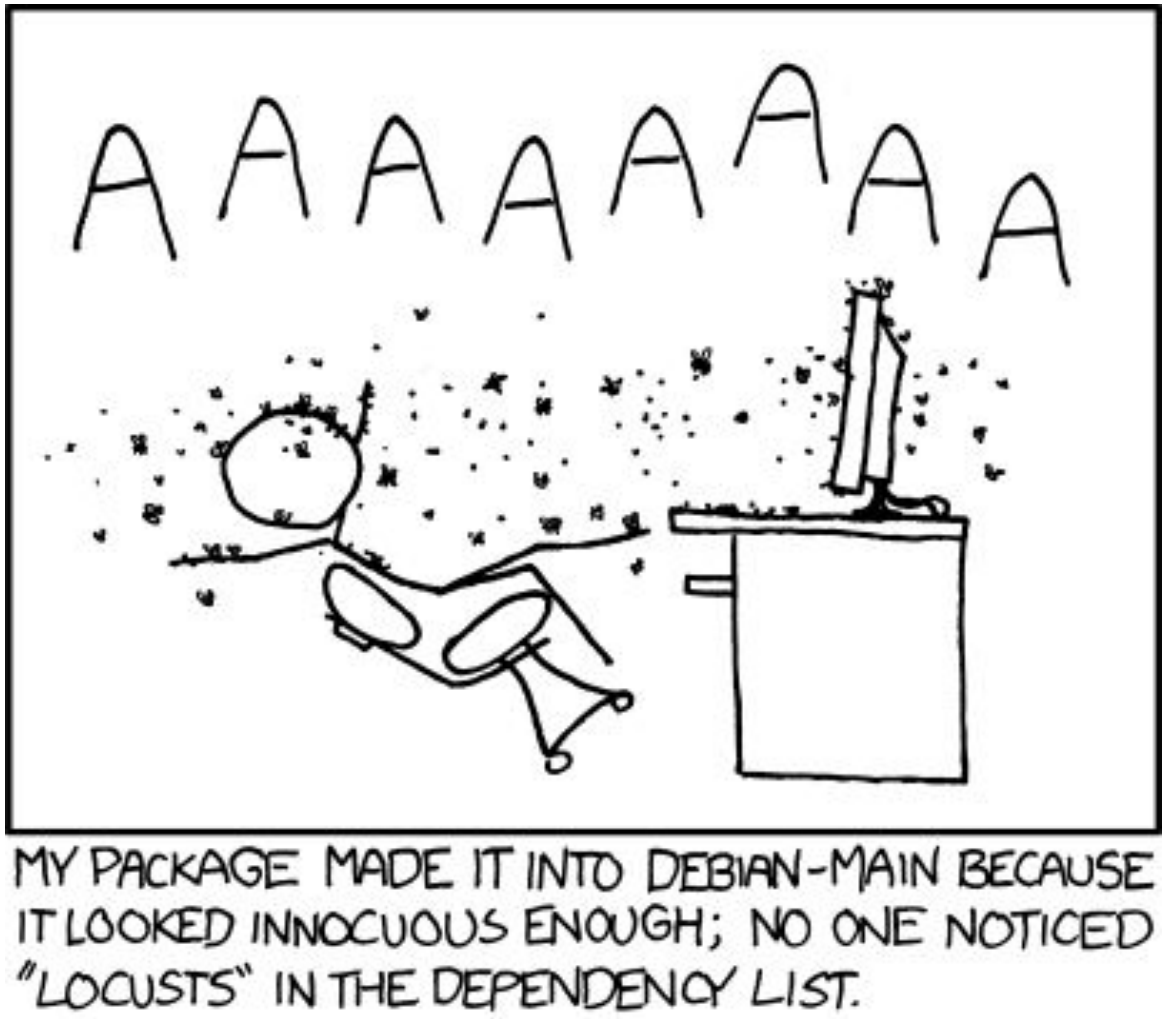




Sources:
https://public.vulnerablecode.io/vulnerabilities/VCID-bk15-3vac-aaaj?search=GHSA-jfh8-c2jp-5v3q
https://github.com/advisories/GHSA-jfh8-c2jp-5v3q

# Why is Software Quality important?

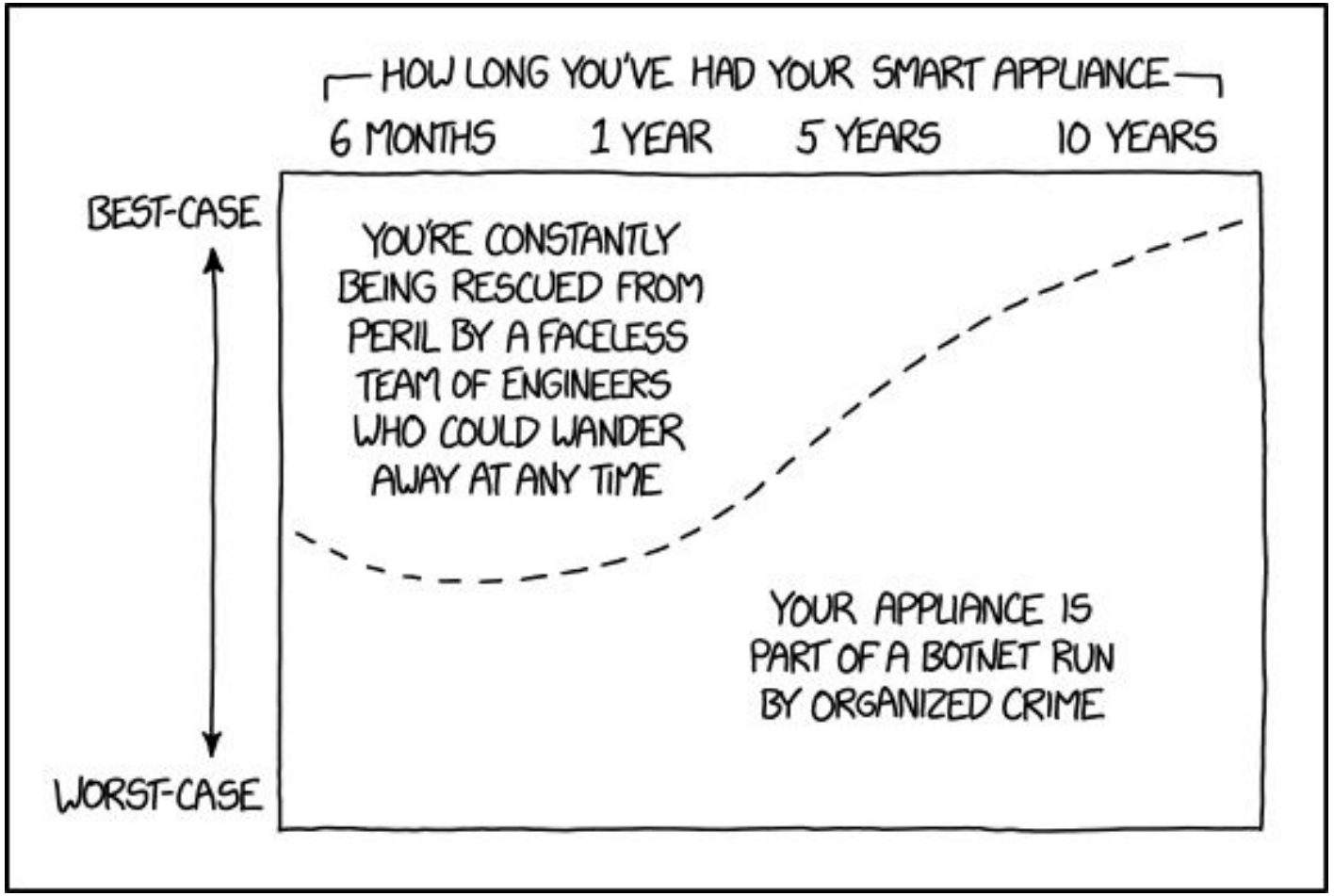
- Better maintained: more secure
- dependencies can be yanked from package archives and replaced by malicious code
- code review, branch protection and other quality checks are important
- Great FOSS projects with open data on quality:
  - OpenSSF Scorecard
  - endoflife.date




Source: https://xkcd.com/797/

# How to communicate? SBOMs

- How to disclose security vulnerabilities in my software?
  - lots of legacy software being used all around us
- What are the software licenses for all the packages used?
- Disclose to direct users, but also to other packages using this




Source: https://xkcd.com/1966/

# And really why?

AboutCode

In the US and in Europe, it’s the law.

- US presidents executive order 14028 mandates SBOM for any software business with the federal government.
- In Europe the CRA (Cyber Resilience Act) was voted this year and mandates SBOM, vulnerability disclosures both downstream to customers and upstream to FOSS projects.
- Software Inventory often looked at by companies before using a software products/acquiring other companies
- Similar legislation/requirements likely in everywhere else

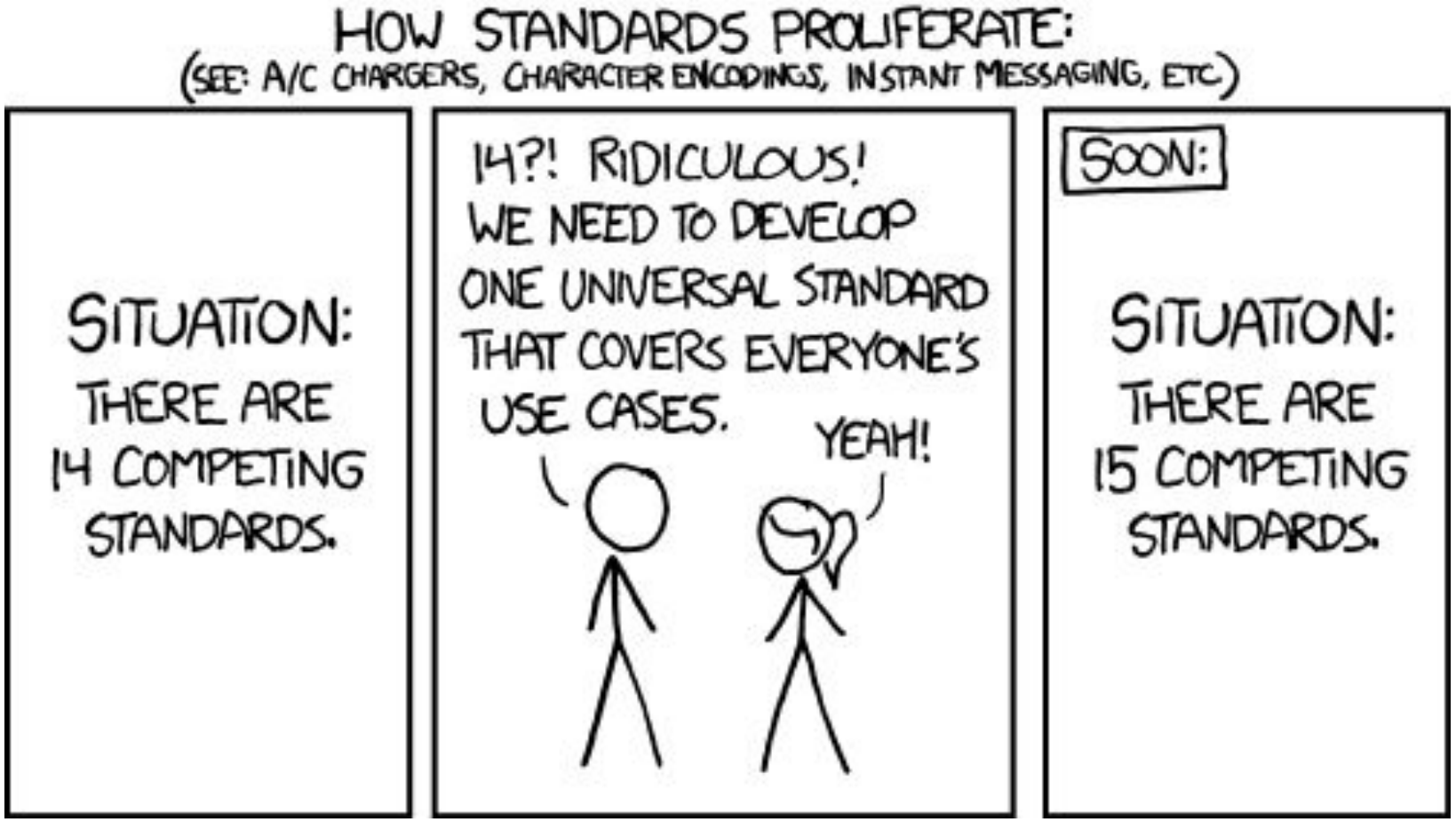
# Are these just more standards?

PackageURL (PURL):

An identifier to uniquely identify and download packages

Vers:

Version range specification for package requirements




Source: https://xkcd.com/927/

# Who is using PackageURL and Vers?

Everyone!

- GitHub Dependency Submission API
- OWASP Dependency-Track
- Two major SBOM standards: CycloneDX and SPDX
- OSS Index
- OSV Schema and OSV.dev (Google)
- AboutCode tools: Scancode Toolkit scancode.io dejacode vulnerablecode
- ORT: OSS Review Toolkit, Osselot
- Anchore, Trivy, Microsoft, Chainguard, Snyk
- cve.org, NVD (soon, maybe)
- Vers is used at vulnerablecode, Google OSV, AppThreat vulnerability-db

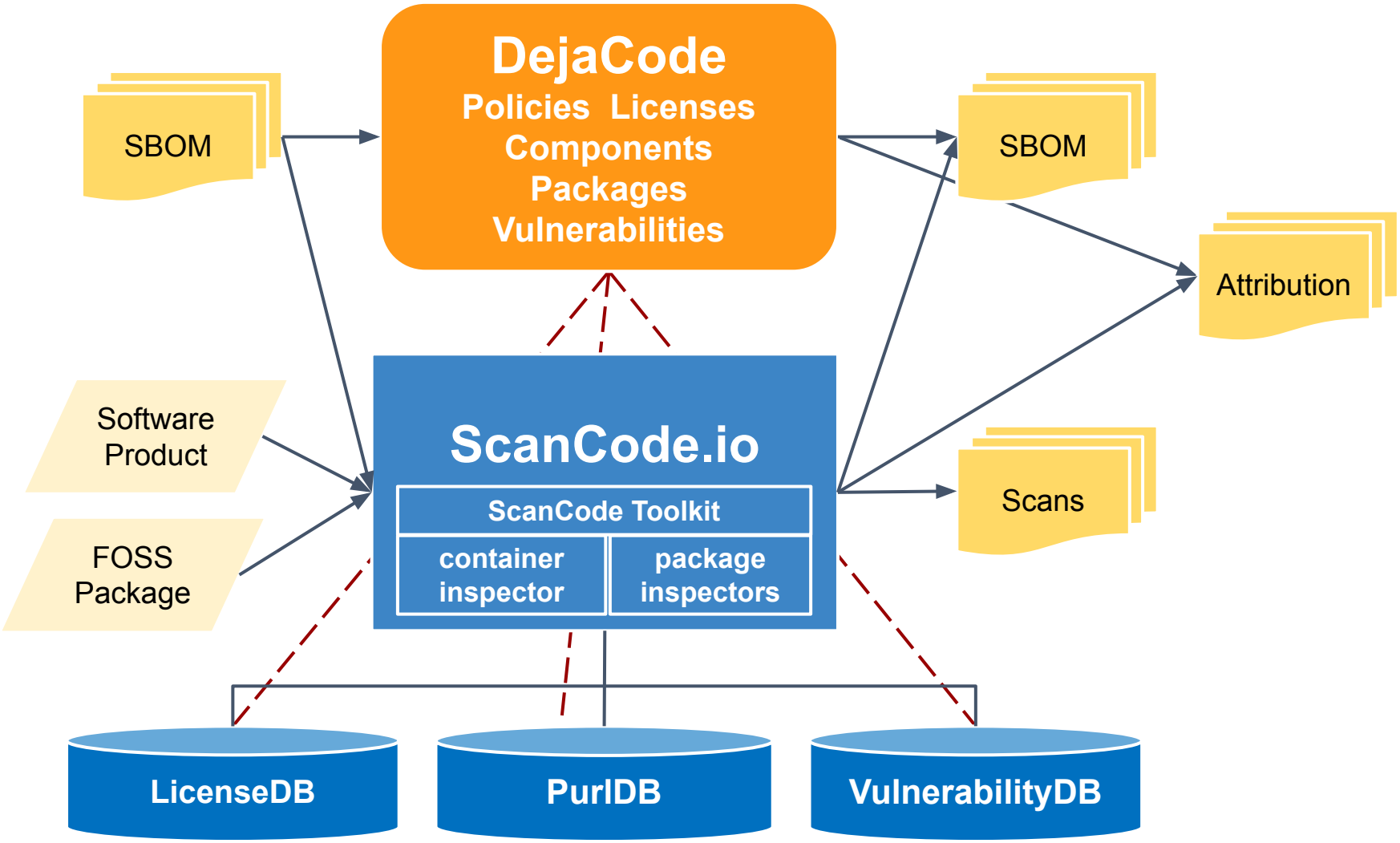
# PackageURL

**Started in ScanCode to uniquely identify packages.**

- pkg:type/namespace/name@version?qualifiers#subpath
  - Specification: https://github.com/package-url/purl-spec
- PURL examples:
  - pkg:deb/debian/curl@7.50.3-1?arch=i386&distro=jessie
  - pkg:github/package-url/purl-spec
  - pkg:pypi/django@1.11.1
  - pkg:rpm/fedora/curl@7.50.3-1.fc25
  - pkg:golang/google.golang.org/genproto#googleapis/api/annotations
- Vers: https://github.com/nexB/univers/

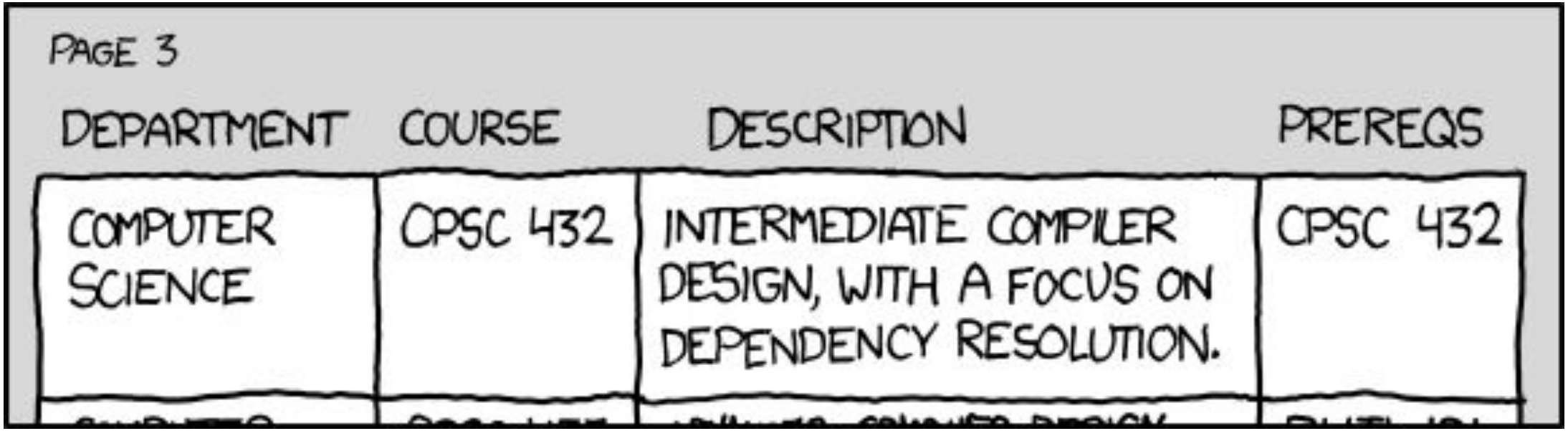
# The AboutCode stack:

AboutCode



# Dependency resolution issues

- Different package versions for the same requirements
- Different results across algorithm/time/environment
- could be useful! Non-vulnerable dependency resolution




Source: https://xkcd.com/754/

# Package identification can be hard

- Code included from different origins
  - vendored (copied partially/fully)
  - distributed with binaries (maven uberjars, jars inside jars)
  - Code matching (MatchCode and PurlDB)
    - Exact archive and file matching
    - Exact and approximate file tree and subtrees matching
- Finding source repo is not trivial:
  - metadata on source repo missing/incorrect
  - many binaries are compiled from the same source package/monorepo
- Customized build systems + metadata formats together

# Where are we now?

AboutCode

- Proprietary solutions getting expensive with the surge of interest in SBOMs
  - may not even work for large companies
- Lots of messy areas in identification.
- Is the vulnerability actually applicable?
  - dependency updates are not always possible
- more automation and FOSS tooling -> more accessible
- Open data as important as open tools!
  - conclude data for packages
  - avoid re-scanning: peer-reviewed, analyzed and curated data
  - initiatives to fix the problem at source

# Other FOSS SCA tools

- ORT: OSS Review Toolkit (Uses ScanCode)
- FOSSology (Uses ScanCode)
- TERN (Uses ScanCode)
- OWASP DependencyTrack
- DepScan (and other AppThreat projects)
- CycloneDx cdxgen

# AboutCode: Who is using it?

(based on public data)

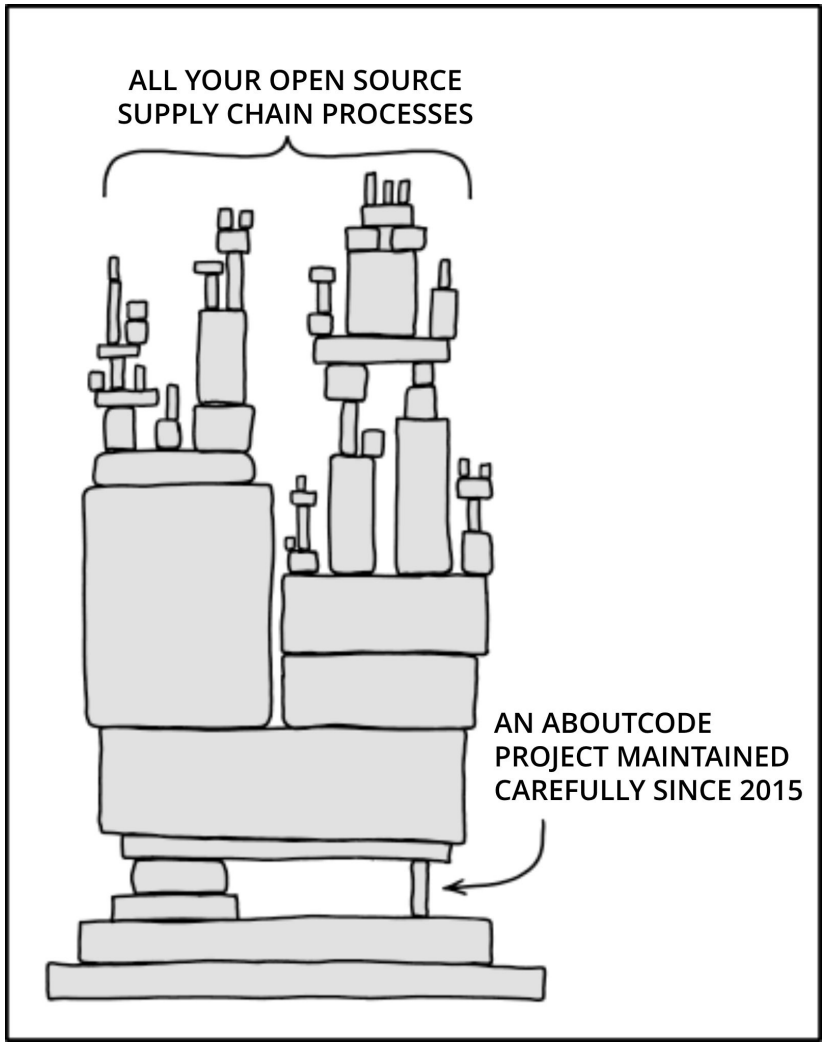
**Most FOSS Orgs, many commercial and open source SCA providers use our libraries or standards**

- Most FOSS Foundations.
- Five of the top big tech companies
- A leading database company, a leading Linux company
- European and US government agencies
- All major European car manufacturers and most of their vendors
- Major US chip and microprocessor providers
- All SBOM and VEX standards
- Used to create a database of permissive code to train an open code LLM
- See https://huggingface.co/blog/starcoder2

# AboutCode also needs your help!

AboutCode

- Contribute to an AboutCode project with code, documentation, use cases, bug reports
  - https://github.com/nexB
- Sponsor AboutCode project maintainers
  - Accelerate development of new features and fund contributors
    - https://github.com/sponsors/nexB
- Buy support, implementation, and advisory services from nexB to pay the maintainers
- Join the community:
  - https://www.aboutcode.org/
  - https://gitter.im/aboutcode-org/discuss




"Dependency" by xkcd, Modified text from original

# Demo

# Questions?

# Credits

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

- All the open source software authors that make AboutCode possible
- xkcd comics under cc-by-nc-2.5
- Presentation template by SlidesCarnival