ScanCode Overview
Fall 2019
Topics

▷ Why ScanCode
▷ ScanCode Community
▷ What is ScanCode?
▷ ScanCode Roadmap
Why ScanCode Toolkit

▷ Easy and simple to install and run, self contained
▷ Best in class license & copyright scan accuracy
  ○ Based on natural language processing
▷ Runs on Linux, Mac OSX and Windows
▷ Find structured package manifest and dependencies
▷ Modern codebase, easy to grok and evolve
Easy to install and run

ScanCode is easy to install and run. It runs on Windows, macOS and Linux. It runs on your laptop.
- It can be used by your development teams
- It can be used by your supply chain partners, big and small

So easy my mother can install and run it!
Best in class license detection

- ScanCode has the most accurate license detection engine
  - Less review needed
- Also collects the full matched notice texts
  - Can be used to automate attribution notice creation
- Other scanning tools miss detecting or misdetect.
  - Used by Linux kernel maintainers to clean kernel licensing
  - Selected by Here.com for ORT as best in class
  - Selected by the Eclipse Foundation for ip due diligence
  - Used in ClearlyDefined to scan 6 Million+ FOSS packages
Easy to integrate in your process

▷ Command line tool, minimal dependencies
  ○ Easy to add to your CI/CD process pipeline
▷ Multiple OS support: Windows, macOS and Linux.
▷ SPDX, JSON, CSV outputs
ScanCode community

▷ 750+ stars on GitHub, 200+ forks, 70+ contributors
▷ Reactive bug fixing
▷ Heavily tested with 10,000+ unit and integration tests
▷ Apache-licensed
▷ ScanCode is part of the OpenChain compliance automation tooling group
▷ ScanCode is the referenced scanner in ClearlyDefined, ORT and Quartermaster.
Who uses ScanCode

▷ Used at top FOSS orgs and projects
  ○ ClearlyDefined, Debian, Eclipse, FSF, Linux kernel, Object Web, OpenEmbedded.org, Openshift analytics, ORT, Quartermaster, CHAOSS and others.

▷ Used at major companies
  ○ Amazon, Comcast, Facebook, Google, Here.com, Microsoft, Red Hat, VMware and others
What is ScanCode?

Open source tools for open source compliance

▷ Toolkit - Identify software origin and license from the code
  https://github.com/nexB/scancode-toolkit

▷ Workbench - Review scans and conclude licensing
  https://github.com/nexB/scancode-workbench

▷ Licenses
  ○ Software - Apache 2.0 (SPDX id: Apache-2.0)
  ○ License Data - Creative Commons Public Domain (SPDX id: CC0-1.0)
Other companions tools and projects

- DeltaCode - compare two scans
- AboutCode toolkit - generate attribution notices
- TraceCode - trace your build: what code is used and how
- VulnerableCode - The free correlated vulnerabilities DB
- conan - Analyze Docker images packages
Next steps

▷ New version 3.x about to be released
▷ We are starting monthly community requirement planning online meeting to plan for 4.x
ScanCode Roadmap

▷ Deduction and inference for scan conclusions
  ○ Traceable Machine Learning
▷ Open source scancode.io - server for scanning
▷ License detection in other languages, beyond English
▷ VulnerableCode - the free vulnerabilities DB is now funded by the EU and NLnet
▷ SPDX Lite support?
ScanCode Toolkit

Detect provenance (origin and license) data from files, packages or package manifests

- Copyright detection based on natural language processing
- License detection based on automatons, inverted indexes and multi-diffs
  - Public repository of license rules and samples
  - Add/correct detections by adding/correcting rule or samples - not code
- JSON, CSV, SPDX and other output formats
ScanCode Toolkit [2]

Other features

▷ Detect authors, URLs and email addresses
▷ Report copyright holders to summarize copyright notices
▷ Plugin architecture for “pre” or “post” extensions - good for filters, summarization or other......
▷ DeltaCode to compare Scans
▷ "Universal" archive extractor
ScanCode Workbench

- Visualize Scan data
- Document license conclusions
- Electron-based desktop application
- Linux, Mac OS X and Windows
ScanCode Workbench

▷ Tree View - see and navigate codebase hierarchy
▷ Dashboard View - visualizations showing the number of Files Scanned and Licenses, Copyrights and Programming Languages detected
▷ BarChart View - bar charts showing summary data for Copyrights, Licenses and other file data
▷ Table View - DataTable for all Scan data
  ○ Configure columns displayed by set or individually
  ○ Set filters on any column
ScanCode Workbench [2]

▷ Conclusions View
  ○ Option to record your concluded license and copyright holder or other fields - very useful for summarisation
  ○ Most fields are pre-filled from Scan data
  ○ Export as draft Inventory to other systems

▷ Other
  ○ WB converts JSON file to SQLite database for use within WB
  ○ Some users use a SQL-DB tool to query the data separately from WB
ScanCode Community

▷ 70+ contributors, 750+ stars, 200+ forks
▷ Used at major tech companies - Amazon, Facebook, Google, Red Hat and others
▷ Used at top FOSS orgs - ClearlyDefined, Debian, Eclipse, FSF, Linux kernel, ORT, Quartermaster, Bitergia/CHAOSS and others.
▷ Google Summer of Code organization: Three students completed projects in GSoC 2019
About nexB

▷ Our mission is to make it easier to reuse FOSS
  ○ Open source solutions for open source compliance
  ○ ScanCode, AboutCode, TraceCode and other projects
▷ Bootstrap company based in Silicon Valley
  ○ DejaCode enterprise compliance system (commercial)
  ○ Acquisition and product audit/analysis services
  ○ Working on FOSS compliance since 2007
Credits

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

▷ Presentation template by SlidesCarnival
▷ Photographs by Unsplash
▷ And all the software authors that made ScanCode possible