AboutCode

The AboutCode stack: End-to-end open source compliance automation with ScanCode and beyond

Agenda

AboutCode

- About me, nexB and AboutCode
- Software Composition Analysis
  - Vulnerabilities AND licensing
  - Proprietary problems
- The AboutCode stack
  - ScanCode
  - VulnerableCode
  - DejaCode
  - Other projects
- Demo
- Questions?

About Philippe

AboutCode

- On a mission to enable easier and safer to reuse FOSS code with best-in-class open source Software Composition Analysis (SCA) tools, data, and standards for open source discovery, license & security compliance
- Lead maintainer of ScanCode, purlDB, VulnerableCode and other AboutCode projects
- CTO and co-founder of nexB, Inc.
  - pombredanne@nexb.com
  - GitHub: https://github.com/pombredanne
  - LinkedIn: https://www.linkedin.com/in/philippeombredanne

About nexB

AboutCode

- Trusted experts on Software Composition Analysis (SCA) since 2007
  - nexB team members are thought leaders:
    - Creator of Package-URL: https://github.com/package-url
    - Co-founders of SPDX: https://spdx.org
    - Co-founders of ClearlyDefined: https://clearlydefined.io
- FOSS-first mission: FOSS for FOSS
  - Open source tools (AboutCode) and open data (licenseDB, vulnerabilityDB, purlDB)
  - Simple and practical standards (Package-URL)
  - Applications for Legal Business users (DejaCode) with APIs for everything
- Professional services for SCA
  - 800+ SCA projects completed to-date with 100% customer satisfaction
  - Sponsored development for AboutCode projects
  - Technical and advisory support for SCA tools implementations

Software Composition Analysis

Software Composition Analysis (SCA)

AboutCode

- Identification – Identify distinct “units” of third-party software used in a product or project and their provenance
- Licensing – Determine the licensing for each software unit
- Security – Identify known security vulnerabilities for each software unit
- Quality – Evaluate the quality of a software unit based on software development data, such as number of bugs, fixes, etc.
  - Read SCA the FOSS Way for more information: https://www.nexb.com/software-composition-analysis/

- SCA needs to be a core competency for any software development organization
  - Embed in the software development workflow from design through release - as it is in manufacturing
  - The choice of SCA tools will depend on your platform, stack and product

SCA: Vulnerabilities AND licensing

AboutCode

- Most SCA tools focus on either vulnerabilities OR licensing
  - Focused on security vulnerabilities because of perceived higher risk
  - The communities of interest are separate - security vs legal
  - License data may be complex, but generally stable over time
  - Vulnerability data is also complex, but extremely dynamic - if included directly in an SBOM, it may be wrong by the time you receive an SBOM

- But you need SCA coverage for both - plus quality

Proprietary SCA Tools

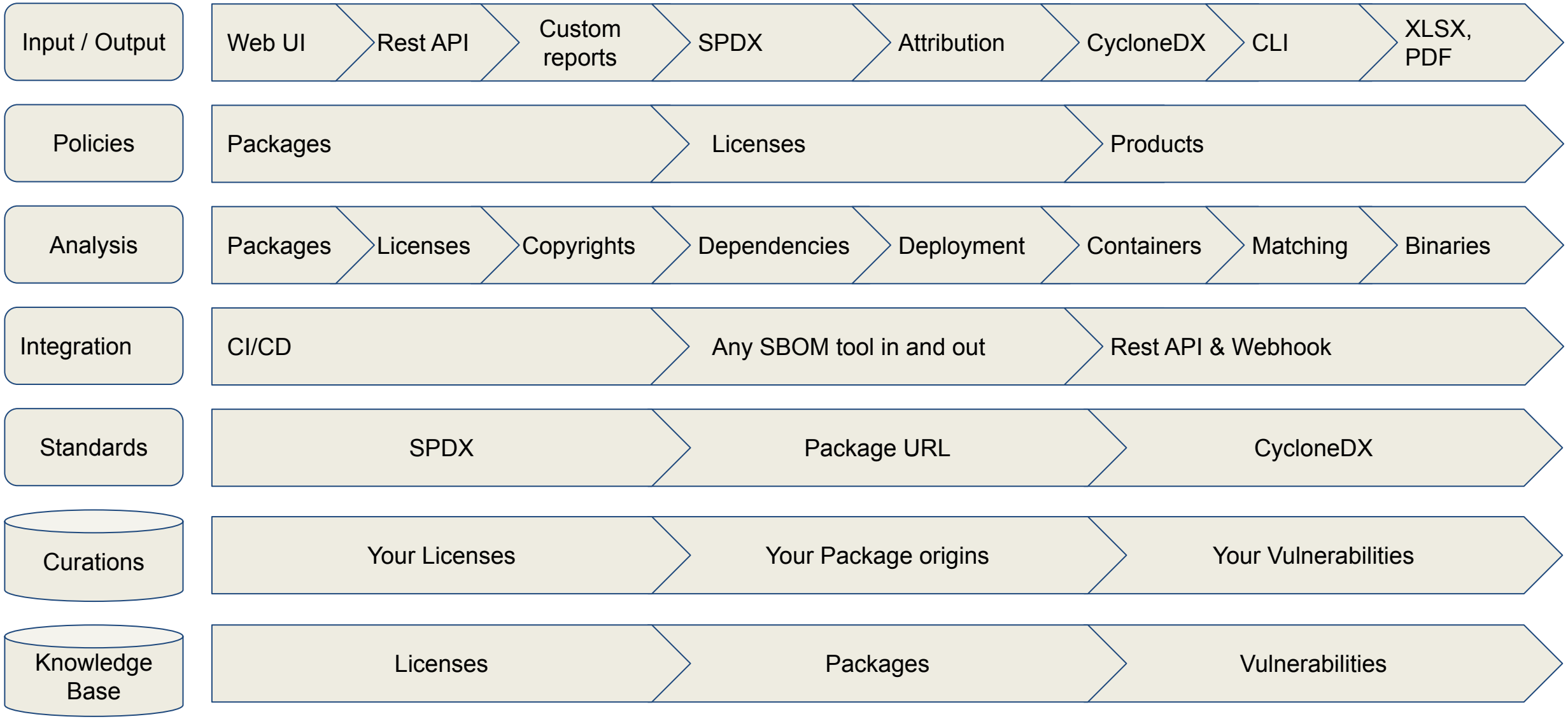
AboutCode

- Increasingly expensive with the surge of interest in SBOMs
  - May work for large companies, but not across the FOSS supply chain
- Most current data about FOSS vulnerabilities is proprietary
  - Barrier to community access and analysis
  - No data -> No SCA!
- We need open reference data and knowledge base
  - License reference data
  - Package origin and details data
  - Vulnerability data
  - Curated, peer reviewed data

Introducing the AboutCode stack

The AboutCode stack: feature areas

AboutCode



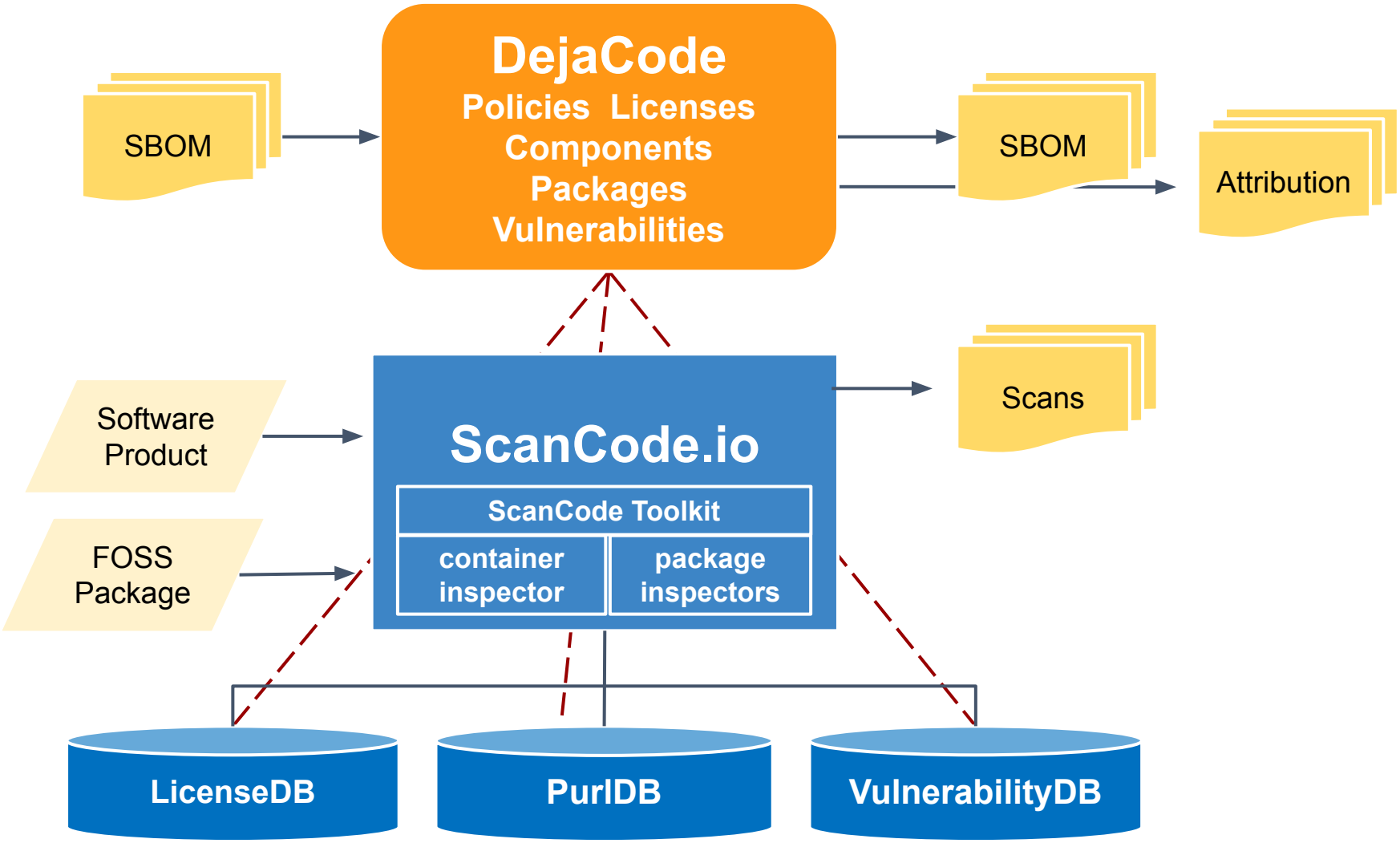
The AboutCode stack for SCA

AboutCode

- Modular and integrated best-in-class SCA tools for developers
  - Free and open source software AND free and open data
- **ScanCode** is the leading code scanner for software component, package and dependency identification and license detection
- **VulnerableCode** for aggregated, comprehensive vulnerability reporting
- **MatchCode** for package and file matching
- **DejaCode** is the SCA and compliance management application
- Many other supporting projects such as
  - Package-URL: specification and tools for identifying packages
  - container-inspector: analysis tool for Docker & other images
- See https://aboutcode.org for an overview of our projects
- See https://github.com/nexB for the code

The AboutCode stack: Architecture

AboutCode



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 free software and open source Foundations.
- Fix 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
- Four leading European industrial companies
- All SBOM and VEX standards
- All open source SCA and SBOM tools
- Most proprietary SCA, SBOM and code hosting tools

ScanCode: Find open source with open source

AboutCode

- Identify FOSS and other third-party components & packages
- Detect licenses, copyrights and dependencies
- ScanCode Projects include:
  - **ScanCode.io:** Server system with customizable pipelines and UI
  - **ScanCode Toolkit**: Scanning engine - use it in SCIO or as a separate CLI or library
  - **LicenseDB**: 35,000 licenses and notices detected by ScanCode
  - **ScanCode Workbench**: Desktop app to review Toolkit Scans
  - scancode-analyzer: Analyze and improve license detection accuracy
  - DeltaCode: Compare scans

- See https://nexb.com/scancode/ for more information

VulnerableCode: Find vulnerabilities, improve security

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
- Discover relations (and inconsistencies) between vulnerabilities and packages from mining the graph
- Public VulnerableCode database is available at: https://public.vulnerablecode.io/
- Also tools to build your own database and working on data sharing and curation

- See https://nexb.com/vulnerablecode/ for more information

DejaCode: Automate continuous compliance

AboutCode

- Open source enterprise application and system of record for:
  - Managing Product Inventory and BOM data
  - Defining and applying license policies
  - Identifying and addressing package vulnerabilities
  - Consuming SBOMs (CycloneDX and SPDX)
  - Generating FOSS compliance documents, such as Product Attribution Notices and SBOMs (CycloneDX and SPDX)
- Built-in integration with ScanCode.io, VulnerableCode.io and PurlDB
- SaaS or on-premises

- See https://nexb.com/dejacode/

Other AboutCode projects

AboutCode

- container-inspector: Analyze Docker and other VM images
- debian-inspector: Parse Debian copyright files and manifests
- elf-inspector: Parse ELF binary files and DWARF symbols
- nuget-inspector: Resolve C# and NuGet dependencies
- python-inspector: Resolve Python dependencies
- aboutcode-toolkit: Generate Attribution Notices
- package-url (purl): URL string to identify and locate a software package across programing languages, package managers, packaging conventions, tools, APIs and databases: https://github.com/package-url
  - Adopted by ORT, CycloneDX and many other major projects
- source-inspector: Parse source files, extract symbols
- univers: parse and compare package versions and version ranges
- license-expression: parse and compare License expressions

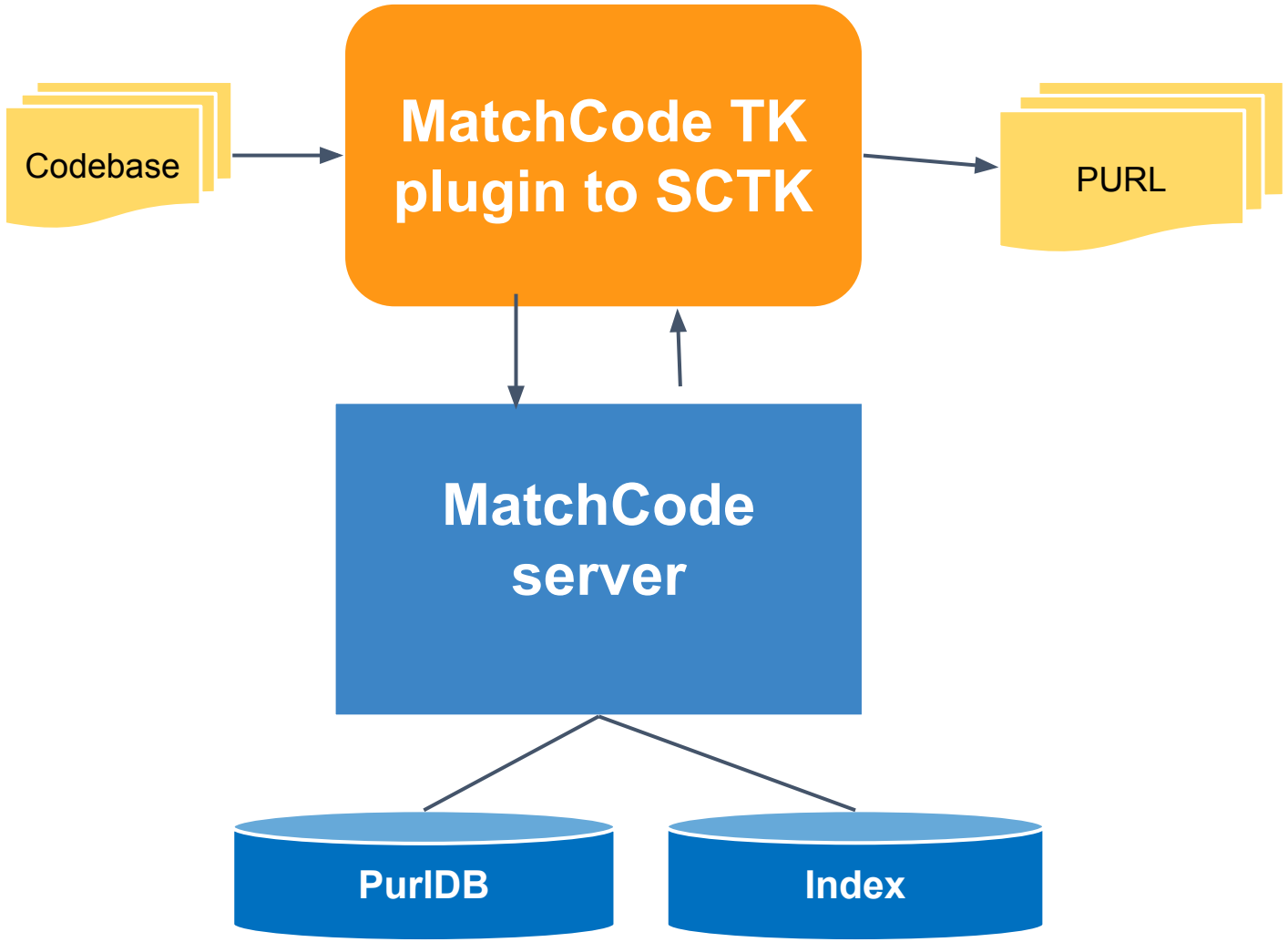
MatchCode

AboutCode

- New matching engine
  - Match to the purlDB
- Different matching approach
  - Exact matching demands a forever growing index
  - Approximate matching can match software that is NOT indexed
  - Top down rather than bottom up
- Initially exact and approximate whole package matching
- To be followed by approximate file and later snippet matching

MatchCode Overview

AboutCode



DEMO

Key roadmap directions

AboutCode

- More code matching
  - Match to the true origin of code
- More shared scans and shared data
  - Never rescan!
- Curation and peer review of scans
  - Never renalyze!
- More binary analysis and deployment tracing workflows
  - Find the exact subset of the code that is deployed
- More supply chain package verification
  - Check but verify all the packages
- More analysis automation
  - End to end automated pipelines for embedded devices

Roadmap items of interest

AboutCode

- DeltaCode
  - Compare scans to propagate curations and to focus review work
- purl2all
  - Many mini tools and services to query things by PURL
- purl2sym
  - Collect code symbols
- back2source
  - Map deployed binaries with the corresponding source code
- FederatedCode
  - Decentralized, federated sharing of scans and vulnerability data
  - To speed up scans, possibly support the social review of package scans
  - See https://www.tdcommons.org/dpubs_series/5632/
- MineCode to build the purlDB

Benefits of the AboutCode stack

AboutCode

- Support safe and compliant use of FOSS
  - Recognized worldwide as best-in-class tools
  - Modular design for adaptation to development team processes, tools and environment
  - Lightweight installation - desktop or server
  - Coverage for all languages and frameworks
  - Code AND data licensed under open source licenses
- Reduce licensing and vulnerability risks from using FOSS or other third-party software components
  - Share risk management responsibilities among business, legal, engineering and security teams
  - Provide a comprehensive view of open source and other third-party components used in your software
- Active community
- Technical support, implementation, advisory services available from nexB

AboutCode also needs your help!

AboutCode

- Contribute to an AboutCode project:
  - Development
  - Documentation
  - Use cases
    - https://github.com/nexB
- Sponsor AboutCode projects
  - Accelerate development of new features and fund community contributors
    - https://github.com/sponsors/nexB
    - https://opencollective.com/aboutcode
- Technical support, implementation, advisory services available from nexB
- Join the community:
    - https://www.aboutcode.org/
    - https://gitter.im/aboutcode-org/discuss

Credits

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

- Presentation template by SlidesCarnival
- Photographs by Unsplash
- All the open source software authors that make AboutCode possible