AboutCode

FOSS for FOSS: DejaCode is your new FOSS control center for SBOMs

Agenda

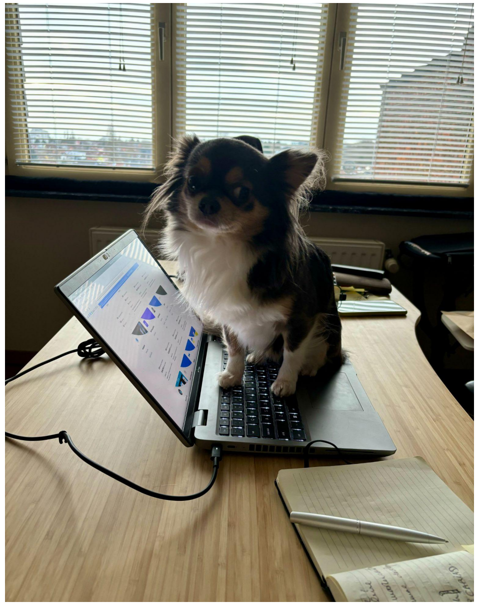
- About me, about AboutCode
- DejaCode
    - Now open source!
    - Demo
    - Other projects in AboutCode stack
    - Roadmap

- Questions

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

About me

- 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 AboutCode projects (ScanCode, DejaCode, VulnerableCode and others)
- CTO and co-founder of nexB, Inc.
  - pombredanne@nexb.com
  - GitHub: https://github.com/pombredanne
  - LinkedIn: https://www.linkedin.com/in/philippeombredanne
  - Often assisted by Chihuahua Technical Advisor



About AboutCode

- **AboutCode's FOSS-first mission: FOSS for FOSS**
  - **Open source tools and open knowledge base (AboutCode stack)**
  - Simple and practical standards (Package-URL / PURL https://github.com/package-url)
  - Applications for Legal & Business users (DejaCode) with APIs for everything
  - Co-founders of SPDX: https://spdx.org
  - Contributors to CycloneDX: https://cyclonedx.org
  - Co-founders of ClearlyDefined: https://clearlydefined.io
  - Anchors for a community of SCA tools user and developers
  - Supported by contributors, nexB and others generous sponsors and supporters!
    - nexB provides professional services and support for SCA to sustain FOSS tools development

The problem

We can assemble code like Lego™.

And it is easy to forget how and where we got the code from.

Yet, this is important for:

- **License?**
- **Security?**
- **But also quality, sustainability and more!**

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

Software Composition Analysis

SCA: proprietary tools and data problem

- Increasingly expensive with the surge of interest in SBOMs and pricing based on number of developers
- Large companies may be able to “afford” proprietary SCA scanning tools, but they do not scale across the FOSS supply chain
  - The cost of scan curation is prohibitive with high false positive rates and poor origin and license detection accuracy
- Most current data about FOSS packages and vulnerabilities is proprietary
  - Vendors may offer some free or open source tools but you must pay for access to their data
  - Vulnerability databases data quality and accuracy is abysmally low
  - Barrier to community access and analysis
- Openwashing/Fauxpen: Many vendors use open source for marketing only

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

The AboutCode stack

SCA Tools

Management Apps

Open Knowledge Base

SCA Tools

ScanCode

- Scan
- Match
- Analysis pipelines
- Binary analysis
- Dependency analysis

Management Apps

DejaCode

- Policies
- Curations
- Software inventory
- Workflows
- SBOMs
- Custom reports

Open Knowledge Base

- Licenses
- Packages
- Vulnerabilities

The AboutCode stack for SCA

- **Web-based "enterprise" management application**
  - DejaCode for ensuring license and security compliance, today's focus
- SCA tools for identifying third-party code license and origin
  - Scan for package, dependencies, license and copyright
  - Match for code origin
- Open knowledge base with open data
  - all the licenses,
  - all the packages
  - all the known vulnerabilities
- Standards: Package-URL, vers: Package version ranges
  - Common identifiers across SCA and vulnerability management

SCA Tools

Management Apps

Open Knowledge Base

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

AboutCode: Who is using it?

Many organizations, and most SCA providers use AboutCode tools, libraries or standards:

- Most free software and open source foundations
- Five of the top big tech companies
- A leading database company and 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 or code hosting tools

SCA Tools

Management Apps

Open Knowledge Base

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

The AboutCode stack: SCA Tools

SCA Tools

- ScanCode, industry-leading scanning engine
  - Scripted scan pipelines for large codebase, containers, VMs, and deployed binary-to-source analysis
- Code matching integrated with the open knowledge base
- Many other libraries and tools
  - ABOUT files for curations/corrections stored in the codebase
  - Inspectors for packages and dependencies
  - univers: parse and compare package versions and version ranges
  - license-expression: parse and compare License expressions
- package-url (PURL) adopted by CycloneDX, CSAF, SPDX and the whole SCA ecosystem

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

The AboutCode stack: Open Data

- **Licenses**: 2,000+ licenses and 35,000 rules
  - No known alternative with comparable depth and breadth
- **Packages**: 21M+ package, their files and fingerprints
  - All Package-URL / PURL-based, public and open data
  - All major ecosystems and distributions - sources AND binaries
  - Metadata, scans, and index of all the packages sources, binaries and VCS repos
  - Index with code fingerprints used for code matching
- **Vulnerabilities**: 760K+ packages and 240K+ vulnerabilities
  - All Package-URL / PURL-based, public and open data
  - All major ecosystems and vulnerability DBs aggregated and correlated
  - Surface conflicting data for vulnerable ranges, fixed versions or affected packages

Open Knowledge Base

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

The AboutCode stack: DejaCode [1]

Integrate all tools and data in one web-based application for SCA and compliance management

- Consume and enrich SBOMs (CycloneDX or SPDX)
- Generate FOSS compliance documents, such as product Attribution Notices and SBOMs (CycloneDX or SPDX)
- Manage product and component inventories
- Curate code origin and licenses
- Identify package vulnerabilities
- Launch scans and access the Knowledge Base
- Define and apply license policies

SCA Tools

Management Apps

Open Knowledge Base

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

The AboutCode stack: DejaCode [2]

Integrate all tools and data in one web-based application for SCA and compliance management

- Standard and custom reports
- JSON API and webhooks
- Built-in basic workflows
- Integrated with AboutCode SCA Tools and open knowledge Base

SCA Tools

Management Apps

Open Knowledge Base

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

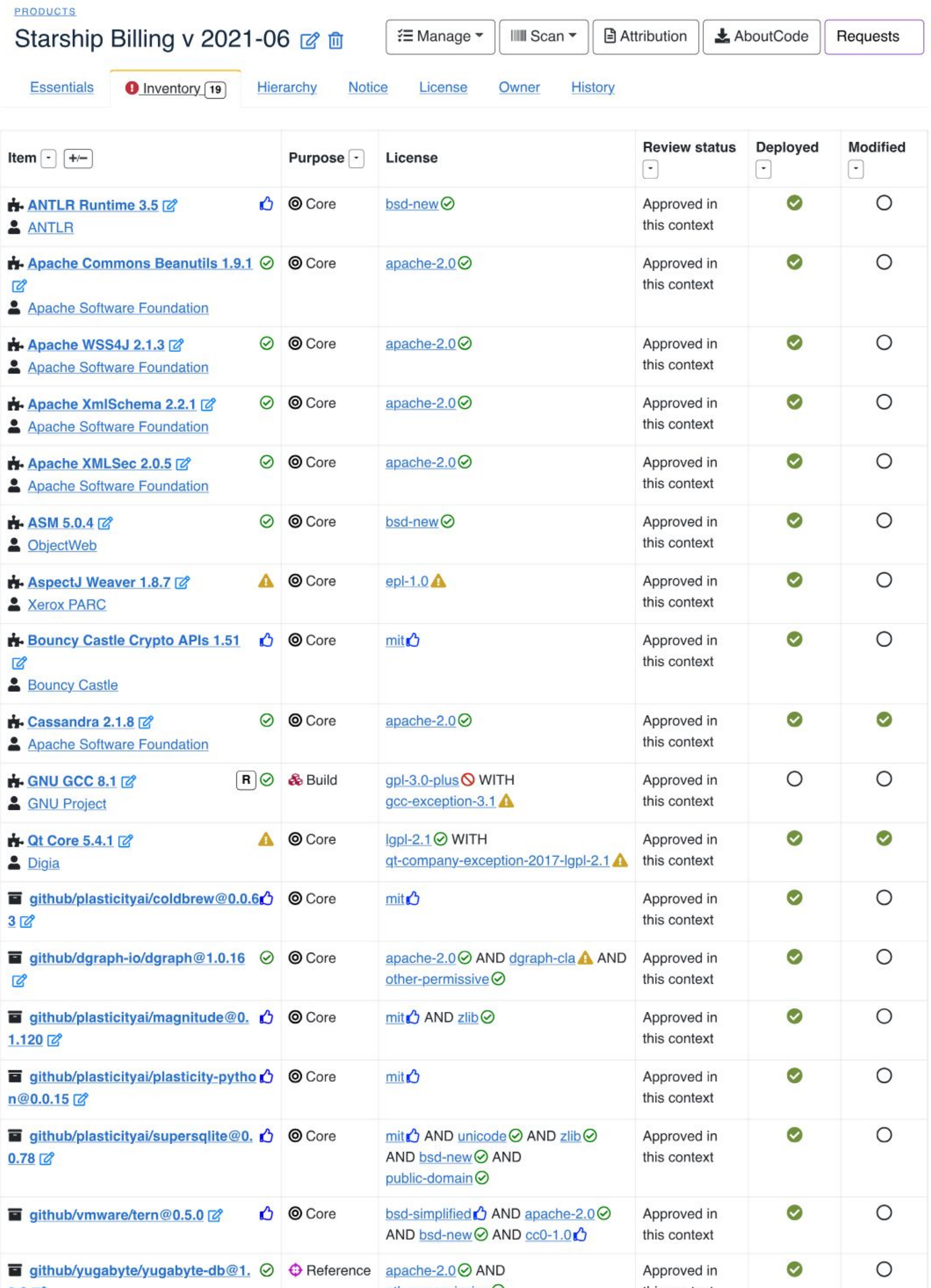
DejaCode Benefits

- **FOSS on FOSS, and open data, free as in puppy**
- Establish code provenance as core to compliance
- Reduce potential licensing and vulnerability risks for using FOSS or other third-party software components responsibly
- Share risk management responsibilities among business, legal, engineering and security concerns and teams
- Provide a comprehensive view of open source and other third-party components used in your software
- Support safe and compliant use of FOSS

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

Get started with DejaCode

- Download and run DejaCode
  - https://github.com/nexb/dejacode
- Or sign up for a free demo hosted account to use DejaCode in a public dataspace
  - Explore, create, and modify components, packages, licenses and assign usage policies to them.
  - Create your own test products and generate attribution.
  - Exercise the DejaCode API and DejaCode integrations with open source tools, such as ScanCode.io.
  - Create SBOMs, run reports and use workflow requests.
  - https://public.dejacode.com/account/register



Starship Billing v 2021-06

Item	Purpose	License	Review status	Deployed	Modified
ANTLR Runtime 3.5 👍 ANTLR	Core	bsd-new ✓	Approved in this context	✓	○
Apache Commons Beanutils 1.9.1 ✓ Apache Software Foundation	Core	apache-2.0 ✓	Approved in this context	✓	○
Apache WSS4J 2.1.3 ✓ Apache Software Foundation	Core	apache-2.0 ✓	Approved in this context	✓	○
Apache XmlSchema 2.2.1 ✓ Apache Software Foundation	Core	apache-2.0 ✓	Approved in this context	✓	○
Apache XMLSec 2.0.5 ✓ Apache Software Foundation	Core	apache-2.0 ✓	Approved in this context	✓	○
ASM 5.0.4 ✓ ObjectWeb	Core	bsd-new ✓	Approved in this context	✓	○
AspectJ Weaver 1.8.7 ⚠ Xerox PARC	Core	epl-1.0 ⚠	Approved in this context	✓	○
Bouncy Castle Crypto APIs 1.51 👍 Bouncy Castle	Core	mit 👍	Approved in this context	✓	○
Cassandra 2.1.8 ✓ Apache Software Foundation	Core	apache-2.0 ✓	Approved in this context	✓	✓
GNU GCC 8.1 R ✓ GNU Project	Build	gpl-3.0-plus 🚫 WITH gcc-exception-3.1 ⚠	Approved in this context	○	○
Qt Core 5.4.1 ⚠ Digia	Core	lgpl-2.1 ✓ WITH qt-company-exception-2017-lgpl-2.1 ⚠	Approved in this context	✓	✓
github/plasticityai/coldbrew@0.0.63 👍	Core	mit 👍	Approved in this context	✓	○
github/dgraph-io/dgraph@1.0.16 ✓	Core	apache-2.0 ✓ AND dgraph-cla ⚠ AND other-permissive ✓	Approved in this context	✓	○
github/plasticityai/magnitude@0.1.120 👍	Core	mit 👍 AND zlib ✓	Approved in this context	✓	○
github/plasticityai/plasticity-python@0.0.15 👍	Core	mit 👍	Approved in this context	✓	○
github/plasticityai/supersqlite@0.0.78 👍	Core	mit 👍 AND unicode ✓ AND zlib ✓ AND bsd-new ✓ AND public-domain ✓	Approved in this context	✓	○
github/vmware/tern@0.5.0 👍	Core	bsd-simplified 👍 AND apache-2.0 ✓ AND bsd-new ✓ AND cc0-1.0 👍	Approved in this context	✓	○
github/yugabyte/yugabyte-db@1. ✓	Reference	apache-2.0 ✓ AND	Approved in	✓	○

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

DejaCode DEMO

AboutCode

Why AboutCode?

- Free and open source software AND free and open data
  - FOSS for FOSS
  - Open knowledgebase with open data for licenses, packages and vulnerabilities
- Modular and integrated best-in-class SCA tools for developers
  - Tackling the harder code analysis problems so you do not have to
  - PURL-based for easier integration in/out
- Bespoke pipelines enable true end-to-end automation
  - Working towards management by exception to focus on the complex cases of origin and license
  - Decentralized analysis, close to the developers
- Management web app for centralized policies, curations and compliance workflows and data
  - Supports engineering, business and legal stakeholders with features tailored for each using common/shared information

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

Benefits of the AboutCode stack

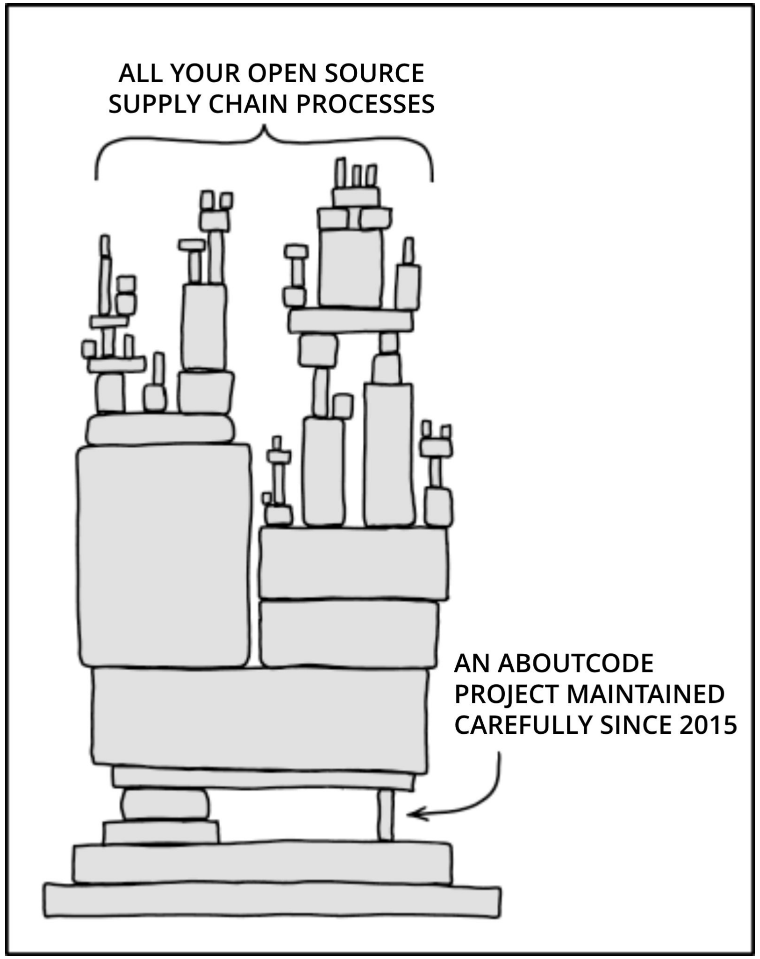
- Supports safe and compliant use of FOSS, with FOSS
  - Recognized worldwide as best-in-class tools
  - Modular design for adaptation to development team processes, tools and environment
  - Coverage for all languages and frameworks
  - Package URL (PURL) used throughout as the package identifier
  - Code AND data licensed under open source licenses, no gimmicks
- 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 of contributors and users, including many FOSS tools
- Technical support, implementation, advisory services available from nexB

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

AboutCode also needs your help!

AboutCode

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



"Dependency" by xkcd, used under CC BY-NC 2.5 / Modified text from original

Roadmap

Roadmap for AboutCode: ScanCode Toolkit

SCA Tools

- Build single exe standalone apps for ScanCode for easier deployment in Ci/CD
- Improve copyright and license detection speed
- Build smaller single-purpose tools and libraries from "mono repo"
- Improve data models for Packages and Dependencies/Requirements
- Parse more package manifests and lock files
- Improve support for license exceptions (WITH)
- Move inconclusive, unknown license detection to clues
- Add post-processing to rematch using SPDX matching guidelines

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

Roadmap for AboutCode: SCA Tools

SCA Tools

- Integrate with CI and other tools
  - Create Ci/CD pre-configured integrations with main CI (GitHub, GitLab, Jenkins)

- Extend binary analysis and deployment tracing workflows
  - Support ELF/Native, Go, Ruby, Android in addition to Java and JS
  - Find the exact subset of the code that is deployed and used in production

- Automate analysis review in ScanCode.io
  - End to end automated pipelines for embedded devices, Android and C/C++
  - Multi-stack deployment analysis for Java, JS, C/C++
  - Report TODO items to review only "by exception"

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

Roadmap for AboutCode: Code matching

SCA Tools

- Code match smart ranking and disambiguation
  - Avoid false positives
- Accurately match to the correct package version
- Match code snippets approximately
  - Using our new approximate fingerprinting
  - Integrate other code matching schemes from SWH and SCANOSS
- Match source symbols and binary symbols to sources and binaries
- New matching pipelines
- Decentralized curation and corrections using in-codebase ABOUT files

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

Roadmap for AboutCode: Other SCA Tools

SCA Tools

- Compare scans to focus review work on changes only (DeltaCode)
- APIs and CLI to query all the things by PURL from the KB (purl2all)
- More code inspectors
  - Lightweight package dependency resolution
  - Dedicated ecosystem-focused libraries
- New lightweight package-inspector
  - Single executable to find packages and dependencies
- Trace build execution to find the exact subset of source code that is deployed and used (TraceCode)

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

Roadmap for AboutCode: Management Apps

- Add support for CycloneDX 1.5 and 1.6 and SPDX 3.0
- Create new review automation apps:
  - License detection review
  - Code match review
  - Vulnerability review
- Overall goal is to reduce review and curation work
  - Extend license clarity scoring to code matches with origin clarity scoring
  - "Auto conclude" matches that are conclusive
- New app for advanced Vulnerability management and support for CRA (Cyber Resiliency Act) compliance
  - Automated triage of vulnerabilities and workflow triggers
  - VEX creation, VEX import and export (Vulnerability Exploitability Exchange) with CSAF and CycloneDX

Management Apps

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

Roadmap for AboutCode: Licenses

- Extend License data with compatibility matrix
- Add new license aliases dataset
- Add more extensive tagging and categorization
- Extend License data with improved exception details
  - To disambiguate license detections of L/GPL with/without exceptions
- Extend License data with improved "or later" details
  - To disambiguate detection of "or later" notices with their primary texts
- Add "key phrases" to all license detection rules
- Add variable text segments to license rules
- Add Fedora alternative SPDX identifiers
- Work with CycloneDX to become their license reference

Open Knowledge Base

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

Roadmap for AboutCode: Vulnerabilities

- Extend Non-vulnerable dependency resolution
    - Beyond Python - add Java and JS
- Extend vulnerability data with new upstream data sources
- Add fix commit details and support for vulnerability reachability
- Mine the graph to surface related package fixes
- Mine git logs, issues and forums to enrich vulnerability data
- Surface inconsistencies and conflicts between different advisory data sources (VulnTotal throughout)
- Add source/binary discrepancy data (from back2source)

Open Knowledge Base

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org

Roadmap for AboutCode: Packages

- Confirm the true origin of code to avoid ambiguous matches
- Supply chain package verification
  - Map deployed binary packages to their corresponding source code
  - Find suspicious code drift between package versions
- Mine extensive list of "off registry" packages
  - Common native C/C++ code and libraries for embedded
  - Glibc, Busybox, zlib, etc. that are not published on ecosystem package registries
- Collect code symbols from source and binaries (for matching)
- On demand, just in time code mining to build your KB on the fly
- Federated, decentralized shared KB data with Git and ActivityPub
  - Share scans, vulnerabilities, origin facts and curations
  - Scan once, analyze once and collaborate on reviews to clear out the junk!

Open Knowledge Base

© AboutCode - License: CC-BY-SA-4.0 - https://www.aboutcode.org