

# FROM LICENSE COMPATIBILITY CHECKS TO SOFTWARE QUALITY ASSURANCE WITH SQA

eScience Center - GFZ

19.09.2022

## Publishing software with good quality

- there can be many criteria for a software to consider it as having good quality:
  - no license or author issues
  - comply to the REUSE Specification
  - good code quality (linting, best practices, static code analysis ...)
  - ...
- why not detect flaws automatically by analyzing the software repository?

*Goal: implement a framework that offers tools to check the quality of a software project*

→ therefore the SQA project was started

## What it does

- command-line tool
- get information about a given software
- software should be contained within a Git repository
- list used programming languages, authors and some metadata
- list licenses (only Python and JavaScript for now)
- check compliance to REUSE Specification
- add missing license headers to files if desired
- check for passwords/secrets committed to the repository
- integration in CI pipelines possible (via Docker image)
- generates a quality report

## Reuse Specification

- SQA uses the reuse-tool to check the software
- SPDX license identifiers need to be used
- each file needs to have license information

```
# SPDX-FileCopyrightText: 2022 John Doe  
#  
# SPDX-License-Identifier: Apache-2.0
```

Example project

GIT REPOSITORY

TC

GeoPeril

Project information

Repository

Issues36

Merge requests0

CI/CD

Security & Compliance

Deployments

Packages & Registries

Infrastructure

Monitor

Analytics

Wiki

Settings

id2 > geoperil > TC GeoPeril

TC

GeoPeril

Project ID: 204

Star

0

Fork

0

485 Commits 7 Branches 44 Tags 316.3 MB Project Storage

tridec-cloud: [Production](#) | [Staging](#), master = Open source version on GitHub, open-source-old-master = GitHub: \_old/master

master

GeoPeril /

+

Find file

Web IDE

Download

Clone

Ignore some Python imports when checking with SQA

Matthias Rüster authored 1 hour ago

43bf1d05

README

CI/CD configuration

Add LICENSE

Add CHANGELOG

Add CONTRIBUTING

Add Kubernetes cluster

Configure Integrations

Name	Last commit	Last update
.sqa	Ignore some Python imports when checking ...	1 hour ago
LICENSES	Add license for world water body data	10 months ago
backend	Mongo: Add index for faster sorting of seale...	3 months ago
docker	Migrate missing features to new architecture	10 months ago
frontend	Frontend: Fix import of vuex types	5 months ago
.gitignore	Add license headers to files	10 months ago
.gitlab-ci.yml	Migrate missing features to new architecture	10 months ago
QUALITY.md	Fix license headers	10 months ago
README.md	Update README.md	10 months ago

## CONFIGURATION FOR LICENSE CHECK

```
description: 'A platform for the computation and web-mapping of hazard specific geospatial data, as well as for serving funct

ignore paths:
--- '*.git/*'
--- '*.reuse/*'

comment in file:
--- '*.sh$'
--- '*.py$'
--- '*.yaml$'
--- '*.toml$'
--- '*.md$'
--- '*.css$'
--- '*.sh$'
--- '*.js$'
--- '*.ts$'
--- '*.xml$'
--- '*.csv$'
--- '*.gitignore$'
--- '*.gitmodules$'
--- '*.Dockerfile$'
--- '*.Makefile.$$'

licenses:
--default:
---copyright: 'Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences, Germany (https://www.gfz-potsdam.de)'
---spx-identifier: 'Apache-2.0'
--Software:
---copyright: 'Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences, Germany (https://www.gfz-potsdam.de)'
---spx-identifier: 'Apache-2.0'
---files:
--- -- '*.txt$'
--- -- '*.md$'
--- -- '*.yaml$'
--- -- '*.gitignore'
--- -- '*.gitmodules'
--- -- '*.py$'
--- -- '*.sh$'
--- -- '*.Dockerfile$'
--- -- '*.css$'
--- -- '*.js$'
--- -- '*.ts$'
--- -- '*.vue$'
```

## CONFIGURATION FOR PYTHON DEPENDENCIES

```
exclude:
  - '.*\\.gitlab-ci.yml$'
  - 'meta/.*'

manual dependency config:
  - Python:
    - import name: 'base'
    - pkg name: 'base'
    - ignore: True
    - import name: 'basesrv'
    - pkg name: 'basesrv'
    - ignore: True
    - import name: 'msgsrv'
    - pkg name: 'msgsrv'
    - ignore: True
    - import name: 'processes'
    - pkg name: 'processes'
    - ignore: True
    - import name: 'surfer'
    - pkg name: 'surfer'
    - ignore: True
    - import name: 'data_products'
    - pkg name: 'data_products'
    - ignore: True
    - import name: 'auto_extent'
    - pkg name: 'auto_extent'
    - ignore: True
    - import name: 'build_cpt_file'
    - pkg name: 'build_cpt_file'
    - ignore: True
    - import name: 'build_legend'
    - pkg name: 'build_legend'
    - ignore: True
    - import name: 'tsunami_config'
    - pkg name: 'tsunami_config'
    - ignore: True
    - import name: 'bson'
    - pkg name: 'pymongo'
    - import name: 'jsonschema'
    - pkg name: 'jsonschema'
    - licenses:
      - MIT
    - import name: 'zipp'
    - pkg name: 'zipp'
    - licenses:
      - MIT
```

## RUN SQA WITH COMMAND LINE

```
docker run -v /home/mruester/git/geoperil.git:/repo \  
-u $(id -u ${USER}):$(id -g ${USER}) \  
git.gfz-potsdam.de:5000/id2/software/services/fair/software-quality-assurance/software-quality-  
assurance:latest \  
sqa run \  
  --reuse_lint \  
  --list_used_licenses \  
  --list_used_languages \  
  --list_authors \  
  --check_credentials
```

# QUALITY REPORT

## Software Quality Report

**Note:** This report is automatically generated by [Software Quality Assurance](#)

### Table of Content

- [Meta Data](#)
- [Completed Services](#)
  - [List Authors](#)
  - [List Used Languages](#)
  - [List Used Licenses](#)
  - [Reuse Lint](#)

### Meta Data

**Software Name:** GeoPeril

**Software Location:** [git](#)

**Last Commit:** 08df94eacb7406c66311c7b8da8090e1864540ee

**Report Time:** 30/05/2022 12:07:44 UTC

**Report Version:** 0.7.1

### Completed Services

#### List Authors

**Description:** Lists all people that committed to this config.

- [Hannes Fuchs](#)
- [Johannes Spazier](#)
- [Martin Hammitzsch](#)
- [Matthias Ruester](#)
- [Sebastian Jüngling](#)
- [Sven Reißland](#)
- [seth0r](#)

#### List Used Languages

**Description:** Detects used programming languages with [linguist](#)

Language	Percentage
Python	54.36
Vue	26.48
Shell	8.41
TypeScript	6.46

# QUALITY REPORT

## List Used Licenses

**Description:** Generates a list of dependencies and their license.

### found licenses

license	libraries
Apache 2.0	Python / requests / 2.22.0 Python / requests / 2.24.0
Apache License, Version 2.0	Python / pymongo / 3.2
Apache Software License	Python / importlib-metadata / 2.0.0
BSD	Python / cherrypy / 3.8.2 Python / geopandas / 0.8.1 Python / itsdangerous / 1.1.0 Python / lxml / 4.6.5 Python / numpy / 1.21.5 Python / owslib / 0.20.0 Python / pandas / 1.1.2 Python / shapely / 1.7.1 Python / weasyprint / 51.0
BSD-2-Clause	Node / ol / 5.3.3
BSD-3-Clause	Python / click / 7.1.2 Python / flask / 1.1.2 Python / jinja2 / 2.11.2 Python / markupsafe / 1.1.1 Python / werkzeug / 1.0.1
BSD-like	Python / idna / 2.10
Dual License	Python / python-dateutil / 2.8.1
LGPL	Python / chardet / 3.0.4
LGPLv3	Python / flufl.enum / 4.1.1
MIT	Python / attrs / 20.2.0 Python / jsonschema / user config Python / pyproj / 2.6.1.post1 Python / pyrsistent / 0.17.3 Python / pytz / 2020.1 Python / pywps / 4.5.1 Python / pyyaml / 5.3.1 Python / six / 1.15.0 Python / sqlalchemy / 1.3.20 Python / urllib3 / 1.25.11 Python / zipp / user config
MIT License	Python / pathlib / 1.0.1
MPL-2.0	Python / certifi / 2020.6.20

# INTEGRATION IN GITLAB CI

<div><div>passed</div><div>00:04:47</div><div>3 months ago</div></div>	<div>Add license information to cronjob file</div> <div>#42847 master f03709a9</div>	<div><div></div><div><div>✓</div><div>✓</div></div></div>	<div><div>Download</div></div>
<div><div>canceled</div><div>00:13:06</div><div>3 months ago</div></div>	<div>Add script for cleaning up old sealevel data</div> <div>#42846 tridec-cloud 70d29d84</div>	<div><div><div>Stage: sqa</div><div>✓ sqa</div></div></div>	<div><div>Next</div><div>Refresh</div><div>Download</div></div>

# INTEGRATION IN GITLAB CI

Search job log

```
1 Running with gitlab-runner 14.4.0 (4b9e985a)
2   on GitLab ID2 group runner on rz-vm8 WGVgZEFR
3 Preparing the "docker" executor 00:16
4 Using Docker executor with image git.gfz-potsdam.de:5000/id2/software/services/fair/software-quality-assurance/software-quality-assurance:latest ...
5 Authenticating with credentials from job payload (GitLab Registry)
6 Pulling docker image git.gfz-potsdam.de:5000/id2/software/services/fair/software-quality-assurance/software-quality-assurance:latest ...
7 Using docker image sha256:de0066ef3e9dce49a5714c19ab21d2ad1e8f06221d4d3c46ee208bc0f1e066a7 for git.gfz-potsdam.de:5000/id2/software/services/fair/software-quality-assurance/software-quality-assurance:latest with digest git.gfz-potsdam.de:5000/id2/software/services/fair/software-quality-assurance/software-quality-assurance@sha256:8b8ec5e0892161db0a0ad24ebf2ea4e019479941815d94c43290dd07b11dcd60 ...
8
9 Preparing environment 00:03
10 Running on runner-wgvgezfr-project-204-concurrent-0 via 027facdeee95...
11
12 Getting source from Git repository 00:01
13 Fetching changes...
14 Reinitialized existing Git repository in /builds/id2/geoperil/GeoPeril/.git/
15 Checking out 43bf1d05 as fix-sqa...
16 Skipping Git submodules setup
17
18 Executing "step_script" stage of the job script 00:11
19 Using docker image sha256:de0066ef3e9dce49a5714c19ab21d2ad1e8f06221d4d3c46ee208bc0f1e066a7 for git.gfz-potsdam.de:5000/id2/software/services/fair/software-quality-assurance/software-quality-assurance:latest with digest git.gfz-potsdam.de:5000/id2/software/services/fair/software-quality-assurance/software-quality-assurance@sha256:8b8ec5e0892161db0a0ad24ebf2ea4e019479941815d94c43290dd07b11dcd60 ...
20 $ sqa --repository_dir "$CI_PROJECT_DIR" run -r -lli -lla -la || true
21 Starting List Authors service...
22 List Authors service done...
23 Starting List Used Languages service...
24 List Used Languages service done...
25 Starting List Used Licenses service...
26 List Used Licenses service done...
27 Starting Reuse Lint service...
28 Reuse Lint service done...
29 Service 'List Used Licenses' exited with 1
30
31 Uploading artifacts for successful job 00:03
32 Uploading artifacts...
33 QUALITY.md: found 1 matching files and directories
34 Uploading artifacts as "archive" to coordinator... ok id=166917 responseStatus=201 Created token=ZcyCsn2j
35
36 Cleaning up project directory and file based variables 00:00
37
38 Job succeeded
```

**sqa**

Duration: 33 seconds

Finished: 18 minutes ago

Queued: 0 seconds

Timeout: 1h (from project)

Runner: #597 (WGVgZEFR) GitLab ID2 group runner on rz-vm8

Job artifacts

The artifacts will be removed in 4 weeks

Keep

Download

Browse

Commit 43bf1d05

Ignore some Python imports when checking with SQA

Pipeline #49581 for fix-sqa

sqa

→ sqa

## Future plans

- support more programming languages (for licenses, linting ...)
- offer a self-service for scientist to check their repository via web browser
- check for duplicate code
- check if tests exist
- static code analysis (vulnerabilities, best practices ...)

Interested?

- SQA repository: <https://git.gfz-potsdam.de/id2/software/services/fair/software-quality-assurance>
- contact us: [software-services@gfz-potsdam.de](mailto:software-services@gfz-potsdam.de)
- helpful links:
  - REUSE specification: <https://reuse.software>
  - SPDX license list: <https://spdx.org/licenses>