
# OSSR metadata implementation

Enrique GARCIA & Thomas VUILLAUME

FG2 call – 03/12/2020

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

# Outline

- Metadata
  - Metadata Contexts
  - CodeMeta Project
- OSSR metadata implementation.
  - Zenodo metadata.


Funded by the European Union's Horizon 2020 - Grant N° 824064

# Metadata

- *"data about data"*
  - Title, authors, description of a source code *is already* metadata.

- Main metadata problem:
  - Each platform/service/program uses their own 'syntax'
  - Lack of ***crosswalk tables*** to 'translate' metadata makes that most of it is lost.

    Ex: `softwareRequirements` vs `install_requires` defining the same concept.

- Adding metadata file starts to become an common practice
  - Full day metadata session @ ADASS XXX meeting.


Funded by the European Union's Horizon 2020 - Grant N° 824064

# Metadata II

- A metadata context:
  - Defines the schema to be followed by the metadata
    - DataCite, CodeMeta, Schema.org, DCMI, figshare, ORCID…
  - Schema: the structure and organization of the metadata.
    - Syntax, key and values of the entries to be written in the file.
    - Sometimes called Terms, Properties and Values, but let's keep it simple.

- Metadata encodings:
  - JSON-LD, XML, URIs (like URLs)…
    - Nice explanation of JSON-LD expansion and compaction algorithms; https://codemeta.github.io/jsonld/


Funded by the European Union's Horizon 2020 - Grant N° 824064

# Metadata II

- A metadata context:
  - Defines the schema to be followed by
    - DataCite, CodeMeta, Schema.org, DCMI
  - Schema: the structure and organizati
    - Syntax, key and values of the entries to
    - Sometimes called Terms, Properties and
- Metadata encodings:
  - JSON-LD, XML, URIs (like URLs)…
    - Nice explanation of JSON-LD expansion
      https://codemeta.github.io/jsonld/

codemeta.json

```
{
  "@context": "https://doi.org/10.5063/schema/codemeta-2.0",
  "@type": "SoftwareSourceCode",
  "name": "ESCAPE template project",
  "description": "A template repository for the ESCAPE project",
  "keywords": "template",
  "license": "https://spdx.org/licenses/GPL-3.0+",
  "identifier": "10.5281/zenodo.3884963",
  "softwareVersion": "v2.0",
  "developmentStatus": "active",
  "codeRepository": "https://gitlab.in2p3.fr/escape2020/wp3/template_project_es
  "runtimePlatform": "Python >3.6",
  "downloadUrl": "https://gitlab.in2p3.fr/escape2020/wp3/template_project_escap
  "fileSize": "39.3 kB",
  "installUrl": "https://gitlab.in2p3.fr/escape2020/wp3/template_project_escape
  "releaseNotes": "Documentation and implementation of the last version for CI/
  "dateCreated": "2019-11-05",
  "datePublished": "2019-12-12",
  "dateModified": "2020-06-08",
  "isAccessibleForFree": true,
  "isPartOf": [
    "https://gitlab.in2p3.fr/escape2020",
    "https://projectescape.eu/"
  ],
  "contIntegration": "https://gitlab.in2p3.fr/escape2020/wp3/template_project_e
  "buildInstructions": "https://gitlab.in2p3.fr/escape2020/wp3/template_project
  "issueTracker": "https://gitlab.in2p3.fr/escape2020/wp3/template_project_esca
  "readme": "https://gitlab.in2p3.fr/escape2020/wp3/template_project_escape/-/b
  "programmingLanguage": [
    {
      "@type": "ComputerLanguage",
      "name": "Python",
      "url": "https://www.python.org/"
```


Funded by the European Union's Horizon 2020 - Grant N° 824064

# CodeMeta project

- Focused on
  - *minimal metadata schema for science software and code* (JSON and XML).

- Based on *Schema.org* schema and extended by the project.
  - *Schema.org* has a *`SoftwareSourceCode`* class
  - *`SoftwareSourceCode`* is composed of `Properties`.
  - Each `Property` is the valid `key` or term within the CodeMeta context.

- Why CodeMeta ?
  - Supported by Software Heritage, ASCL…
    - In the near future by Zenodo ?
  - Product of Mozilla Science Lab
  - Can be easily extended with Schema.org properties (SoftwareApplication, DataSet, Thing…)


Funded by the European Union's Horizon 2020 - Grant N° 824064

# CodeMeta project

- Satisfactory metadata solution for code (OSSR), the question now is

***is this context enough for the ESCAPE environment and the connection of all the ESCAPE services ?***

- OK, let's start easy; OSSR – DIOS connection ?


Funded by the European Union's Horizon 2020 - Grant N° 824064

# codemeta implementation into the OSSR

- By incorporating a `codemeta.json` file into a project
  - OSSR will work as a test case for the connection between the ESCAPE services
    - Can the Analysis platform read this metadata ?
    - Is it enough ? Need more terms ? Describing what exactly ?
    - What happen with containers ? New codemeta file ? Can codemeta correctly describe it ?
    - Same questions for Jupyter-Notebooks.
  - If the project evolves, the metadata evolves with it.
  - The OSSR-CI pipeline (Gitlab-Zenodo connection) won't need to read different metadata sources.
    - Standardise publication into Zenodo.
  - Any ESCAPE external service that reads/crosswalks codemeta can can access the publication/project/library/container.


Funded by the European Union's Horizon 2020 - Grant N° 824064

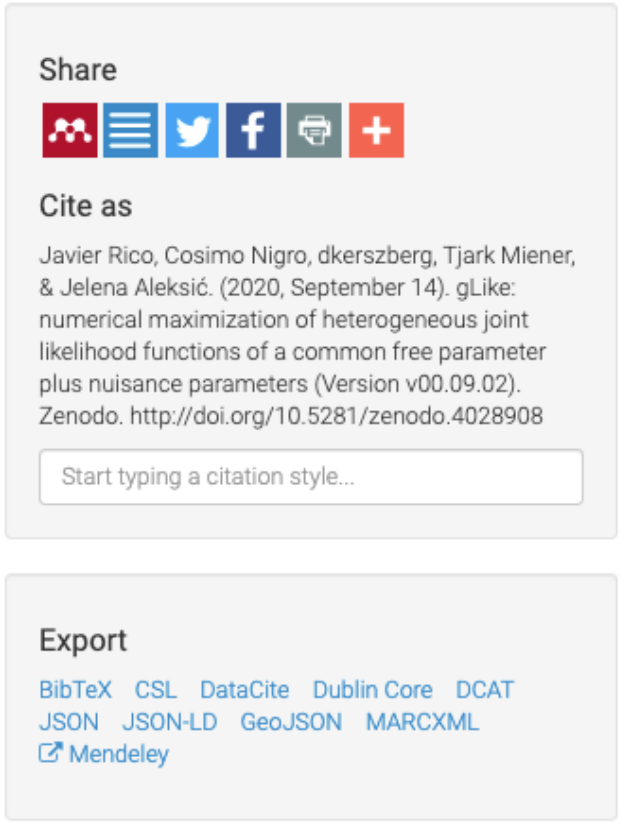
# codemeta implementation into the OSSR

- How to create a `codemeta.json` file
  - https://codemeta.github.io/codemeta-generator/ (5-10 min !)
  - Add it into the root directory of your project.
    - GitlabCI pipeline will make use of the file to provide metadata to Zenodo.


Funded by the European Union's Horizon 2020 - Grant N° 824064

# Zenodo metadata; `.zenodo.json`

- Zenodo follows their own schema
  - However it automatically exports any entry to different contexts and schemas
  - The OSSR-CI (GitLab-Zenodo pipeline) creates this file from a provided codemeta.json file.





Funded by the European Union's Horizon 2020 - Grant N° 824064

# OSSR-CI (GitLab-Zenodo pipeline)

- Starts to get very complicated.

- Lot of development going on, current refactoring the whole service to really simplified it…

…to be continued.



Funded by the European Union's Horizon 2020 - Grant N° 824064