ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

How to harvest and retrieve entries from the OSSR

Enrique GARCIA & Thomas VUILLAUME

FG2 call – 26/05/2021

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

- Harvest metadata from a repository
  - OAI-PMH protocol
  - Repository metadata representations Vs. entry metadata

- Query entry metadata
  - Zenodo REST API
  - OAI-PMH Vs. REST API

- Demo

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

Harvesting metadata from a repository

- The provider is responsible of the service.
  - It chooses the harvest protocol and metadata representation of the records.

- All metadata records are licenced under CC0 license, besides the license applying to the data files of the digital entry.

- Zenodo:
  - OAI-PMH protocol
    - https://developers.zenodo.org/#oai-pmh
  - Metadata supported schemas:
    - DataCite (various versions)
    - Dublin Core
    - MARC21

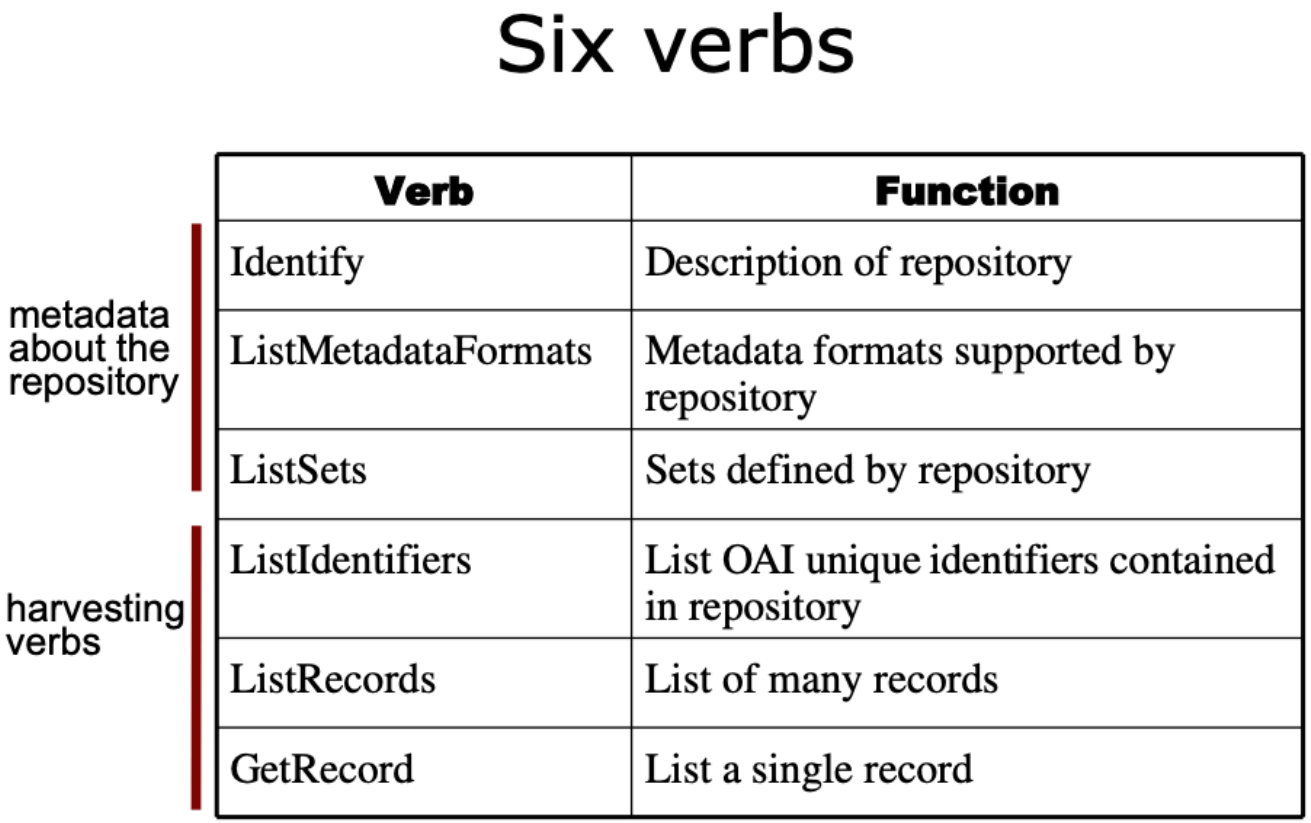
Funded by the European Union's
Horizon 2020 - Grant N° 824064

OAI-PMH Protocol

- Open Archives Initiative Protocol for Metadata Harvesting
  - Developed for harvesting records in an archive/repository.

- Uses a base URL to which you can add different *”verbs”* to reduce the query/search.
  - OAI-PMH (exhaustive!) tutorial
    https://indico.cern.ch/event/5710/sessions/108048/attachments/988151/1405129/Simeon_tutorial.pdf

- Various ’harvester’-libraries in various programming languages
  - Python: oai-harvest, pyoai…



Six verbs

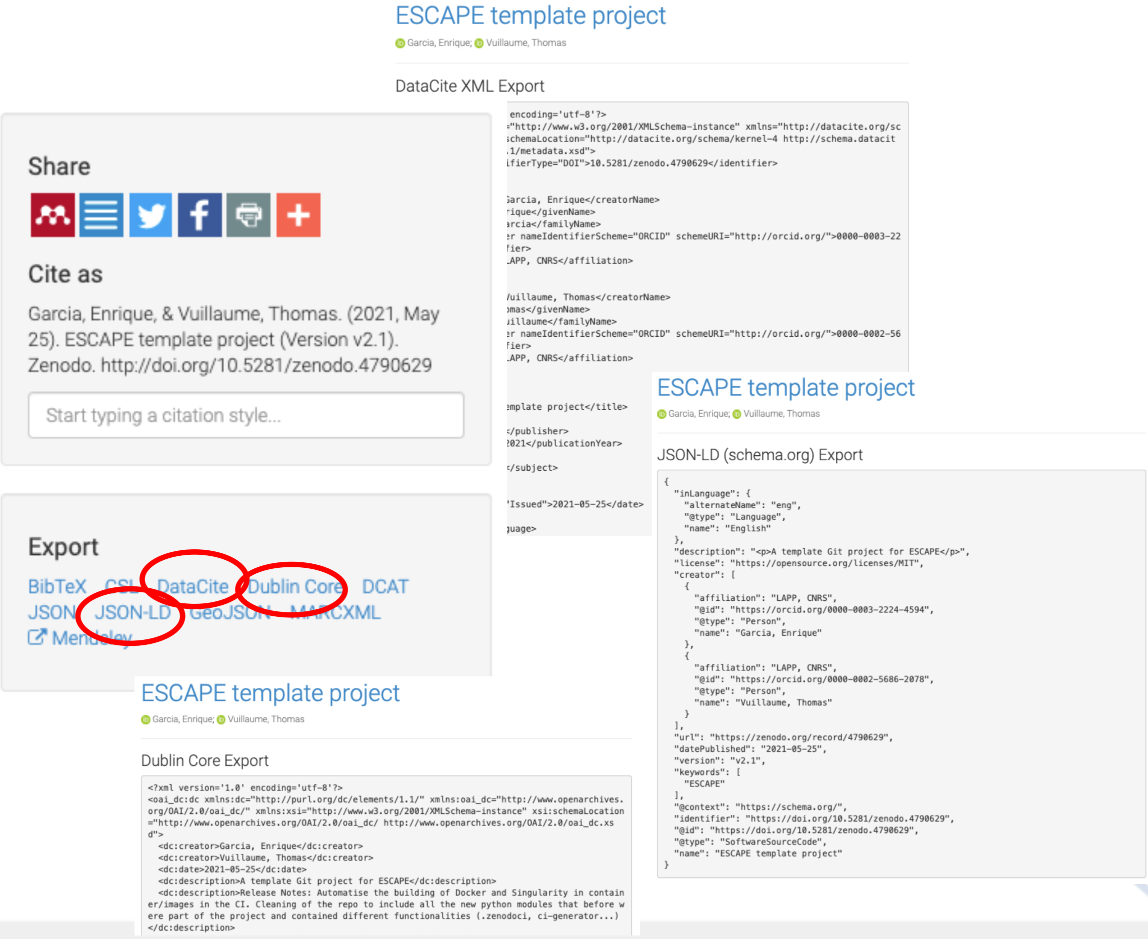
metadata about the repository

harvesting verbs

Verb	Function
Identify	Description of repository
ListMetadataFormats	Metadata formats supported by repository
ListSets	Sets defined by repository
ListIdentifiers	List OAI unique identifiers contained in repository
ListRecords	List of many records
GetRecord	List a single record

Repository metadata representation Vs. resource metadata

- Not the same !



ESCAPE template project

Garcia, Enrique; Vuillaume, Thomas

DataCite XML Export

```
 encoding='utf-8'?>
="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://datacite.org/sc
schemaLocation="http://datacite.org/schema/kernel-4 http://schema.datacit
.1/metadata.xsd">
ifierType="DOI">10.5281/zenodo.4790629</identifier>


Garcia, Enrique</creatorName>
rique</givenName>
arcia</familyName>
er nameIdentifierScheme="ORCID" schemeURI="http://orcid.org/">0000-0003-22
fier>
LAPP, CNRS</affiliation>


/uillaume, Thomas</creatorName>
omas</givenName>
uillaume</familyName>
er nameIdentifierScheme="ORCID" schemeURI="http://orcid.org/">0000-0002-56
fier>
LAPP, CNRS</affiliation>



emplate project</title>

</publisher>
2021</publicationYear>

</subject>


'Issued">2021-05-25</date>

guage>
```

Share

Cite as

Garcia, Enrique, & Vuillaume, Thomas. (2021, May 25). ESCAPE template project (Version v2.1). Zenodo. http://doi.org/10.5281/zenodo.4790629

Start typing a citation style...

Export

BibTeX CSL DataCite Dublin Core DCAT JSON JSON-LD GeoJSON MARCXML Mendeley

ESCAPE template project

Garcia, Enrique; Vuillaume, Thomas

JSON-LD (schema.org) Export

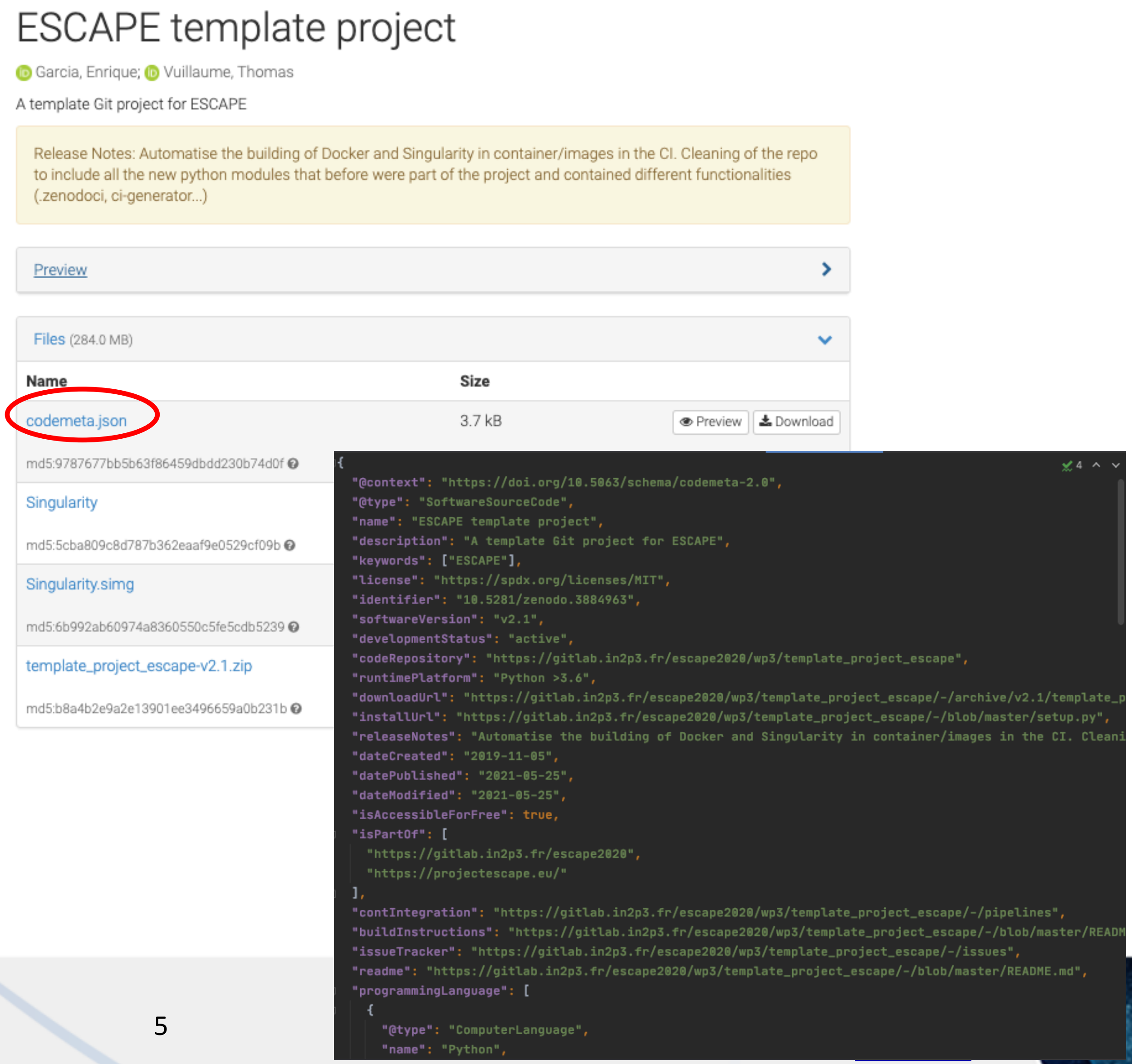
```
{
  "inLanguage": {
    "alternateName": "eng",
    "@type": "Language",
    "name": "English"
  },
  "description": "<p>A template Git project for ESCAPE</p>",
  "license": "https://opensource.org/licenses/MIT",
  "creator": [
    {
      "affiliation": "LAPP, CNRS",
      "@id": "https://orcid.org/0000-0003-2224-4594",
      "@type": "Person",
      "name": "Garcia, Enrique"
    },
    {
      "affiliation": "LAPP, CNRS",
      "@id": "https://orcid.org/0000-0002-5686-2078",
      "@type": "Person",
      "name": "Vuillaume, Thomas"
    }
  ],
  "url": "https://zenodo.org/record/4790629",
  "datePublished": "2021-05-25",
  "version": "v2.1",
  "keywords": [
    "ESCAPE"
  ],
  "@context": "https://schema.org/",
  "identifier": "https://doi.org/10.5281/zenodo.4790629",
  "@id": "https://doi.org/10.5281/zenodo.4790629",
  "@type": "SoftwareSourceCode",
  "name": "ESCAPE template project"
}
```

ESCAPE template project

Garcia, Enrique; Vuillaume, Thomas

Dublin Core Export

```
<?xml version='1.0' encoding='utf-8'?>
<oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:oai_dc="http://www.openarchives.
org/OAI/2.0/oai_dc/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation
="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xs
d">
  <dc:creator>Garcia, Enrique</dc:creator>
  <dc:creator>Vuillaume, Thomas</dc:creator>
  <dc:date>2021-05-25</dc:date>
  <dc:description>A template Git project for ESCAPE</dc:description>
  <dc:description>Release Notes: Automatise the building of Docker and Singularity in contain
er/images in the CI. Cleaning of the repo to include all the new python modules that before w
ere part of the project and contained different functionalities (.zenodoci, ci-generator...)
</dc:description>
```



ESCAPE template project

Garcia, Enrique; Vuillaume, Thomas

A template Git project for ESCAPE

Release Notes: Automatise the building of Docker and Singularity in container/images in the CI. Cleaning of the repo to include all the new python modules that before were part of the project and contained different functionalities (.zenodoci, ci-generator...)

Preview

Files (284.0 MB)

Name	Size	
codemeta.json md5:9787677bb5b63f86459dbdd230b74d0f	3.7 kB	Preview Download
Singularity md5:5cba809c8d787b362eaaf9e0529cf09b		
Singularity.simg md5:6b992ab60974a8360550c5fe5cdb5239		
template_project_escape-v2.1.zip md5:b8a4b2e9a2e13901ee3496659a0b231b		

```
{
  "@context": "https://doi.org/10.5063/schema/codemeta-2.0",
  "@type": "SoftwareSourceCode",
  "name": "ESCAPE template project",
  "description": "A template Git project for ESCAPE",
  "keywords": ["ESCAPE"],
  "license": "https://spdx.org/licenses/MIT",
  "identifier": "10.5281/zenodo.3884963",
  "softwareVersion": "v2.1",
  "developmentStatus": "active",
  "codeRepository": "https://gitlab.in2p3.fr/escape2020/wp3/template_project_escape",
  "runtimePlatform": "Python >3.6",
  "downloadUrl": "https://gitlab.in2p3.fr/escape2020/wp3/template_project_escape/-/archive/v2.1/template_p
  "installUrl": "https://gitlab.in2p3.fr/escape2020/wp3/template_project_escape/-/blob/master/setup.py",
  "releaseNotes": "Automatise the building of Docker and Singularity in container/images in the CI. Cleani
  "dateCreated": "2019-11-05",
  "datePublished": "2021-05-25",
  "dateModified": "2021-05-25",
  "isAccessibleForFree": true,
  "isPartOf": [
    "https://gitlab.in2p3.fr/escape2020",
    "https://projectescape.eu/"
  ],
  "contIntegration": "https://gitlab.in2p3.fr/escape2020/wp3/template_project_escape/-/pipelines",
  "buildInstructions": "https://gitlab.in2p3.fr/escape2020/wp3/template_project_escape/-/blob/master/READM
  "issueTracker": "https://gitlab.in2p3.fr/escape2020/wp3/template_project_escape/-/issues",
  "readme": "https://gitlab.in2p3.fr/escape2020/wp3/template_project_escape/-/blob/master/README.md",
  "programmingLanguage": [
    {
      "@type": "ComputerLanguage",
      "name": "Python",
```

Query metadata within the same record

- Zenodo REST API
  - allows accessing to this metadata, if available.
  - License ?

- OAI-PMH Vs. REST API
  - AOI – PMH
    - + Better harvesting
      - Faster.
      - Thought for large and continuous queries.
    - – Metadata representation provided by the data provider.
  - REST API
    - + Access to the full record information.
    - – Harvesting not optimized for large searches.

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

- Demo

- REST API handlers (python)
  - ZenodoCI: https://gitlab.in2p3.fr/escape2020/wp3/zenodoci
  - PyZenodo: https://github.com/space-physics/pyzenodo3
  - zenodo-python: https://github.com/SiLeBAT/zenodo-python

- OAI-PMH harvesters (python)
  - OAI-PMH Harvest: https://github.com/bloomonkey/oai-harvest
  - pyoai: https://github.com/infrae/pyoai

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