



ESCAPE

European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures

OSSR Implementation concept & status

Enrique GARCIA – LAPP/CNRS

27/07/2020 – ESCAPE WOSSL workshop



- Repository implementation Goals
 - Development platform
 - GitLab in2p3 instance
 - Long term repository
 - Zenodo
 - Connection of services
 - Containerization
 - CI/CD pipeline between Dev platform and repository
 - Metadata

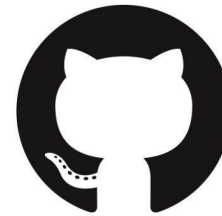


- Converge to common Open Source and FAIR practices that adapt and evolve to the community
 1. Provide an open-source **development platform** available for all partners and beyond;
 2. A long-term **repository** insuring the implementation of the FAIR principles and the **connection** with **other ESCAPE services** (ESAP, EOSC portal/marketplace)



Development Platform

- Environment where to:
 - Develop software.
 - Test, modify, debug, release... software



GitHub



GitLab in2p3 instance

- GitHub is not OS

- Bought by Microsoft 2008

- GitLab is hosted within a ESCAPE institution

- <https://gitlab.in2p3.fr/escape2020>

- No intention of forcing communities to change habits !

- Make available an environment if anybody needs it !
- Feel free to create new projects here !



GitLab



(Git)hub/Lab as a service/software

	+	-
Github or gitlab as a service	<ul style="list-style-type: none"> - Free for open source - No setup - Integration with lot of services - Huge community 	<ul style="list-style-type: none"> - Data stored outside the E.U. - No control over the data & conditions of use - No control of accounts
Gitlab as a software	<ul style="list-style-type: none"> - Private projects - Control - Open source 	<ul style="list-style-type: none"> - Setup, install, maintenance of the service - Costs - Less integration with other services

● Slide from T. Vuillaume (24/07)



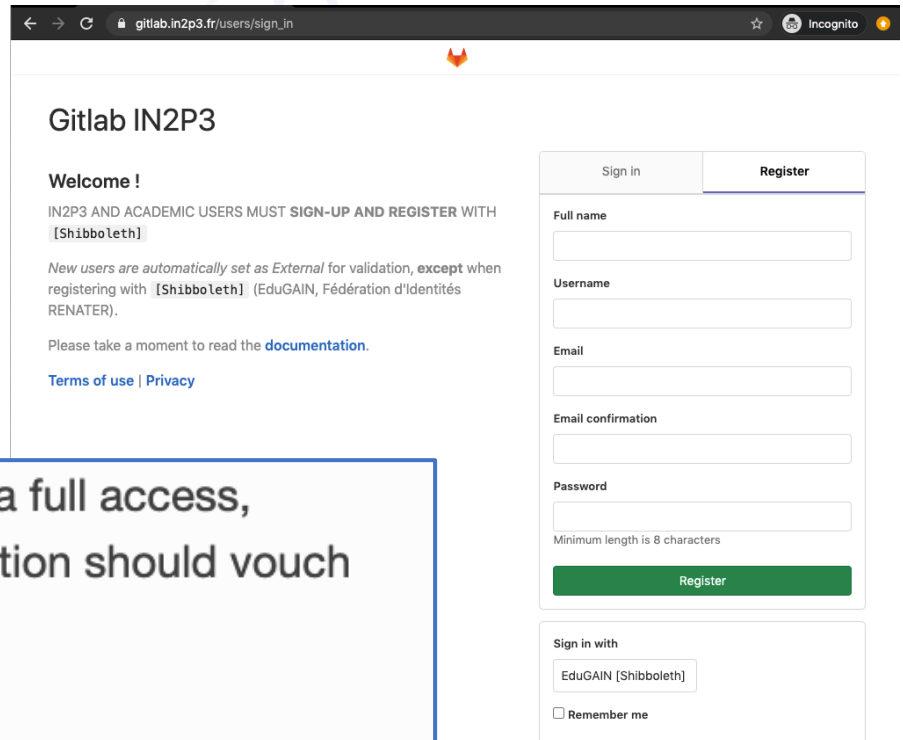
GitLab in2p3 instance - FAQs

☀️ 'I don't have a
EduGAIN/Shibboleth
account'

(thanks Jutta !)

For non-eduGAIN users who wish to get a full access,
someone belonging to a partner organization should vouch
for you to [user support](#).

New non-eduGAIN users will be, by default, set as [external](#).
Validators will unlock the account after verification.



The screenshot shows the GitLab IN2P3 sign-in and registration page. The browser address bar shows 'gitlab.in2p3.fr/users/sign_in'. The page title is 'Gitlab IN2P3'. There are two tabs: 'Sign in' and 'Register'. The 'Register' tab is active. The registration form includes fields for 'Full name', 'Username', 'Email', 'Email confirmation', and 'Password'. A note states 'Minimum length is 8 characters'. A green 'Register' button is at the bottom of the form. Below the form, there is a 'Sign in with' section with a dropdown menu showing 'EduGAIN [Shibboleth]' and a checkbox for 'Remember me'.



Long term repository

- A environment where to store digital artefacts.



Long term repository

- A environment where to store digital artefacts.

Software **Repository** != **Development** Platform

Stable release != Minor changes in a project

- ESCAPE needs → General purpose;
 - Software
 - Software containers
 - Documents, presentations, multimedia...
 - Full projects
 - Data sets *



- General purpose repository

- FAIR principles

- Findable

- Reference/Identification by DOIs + metadata

- Accessible

- Archival + Human/machine readable metadata

- Interoperable

- Accepts different metadata schemas

- Reusable :

- License and provenance



Zenodo as a service/software



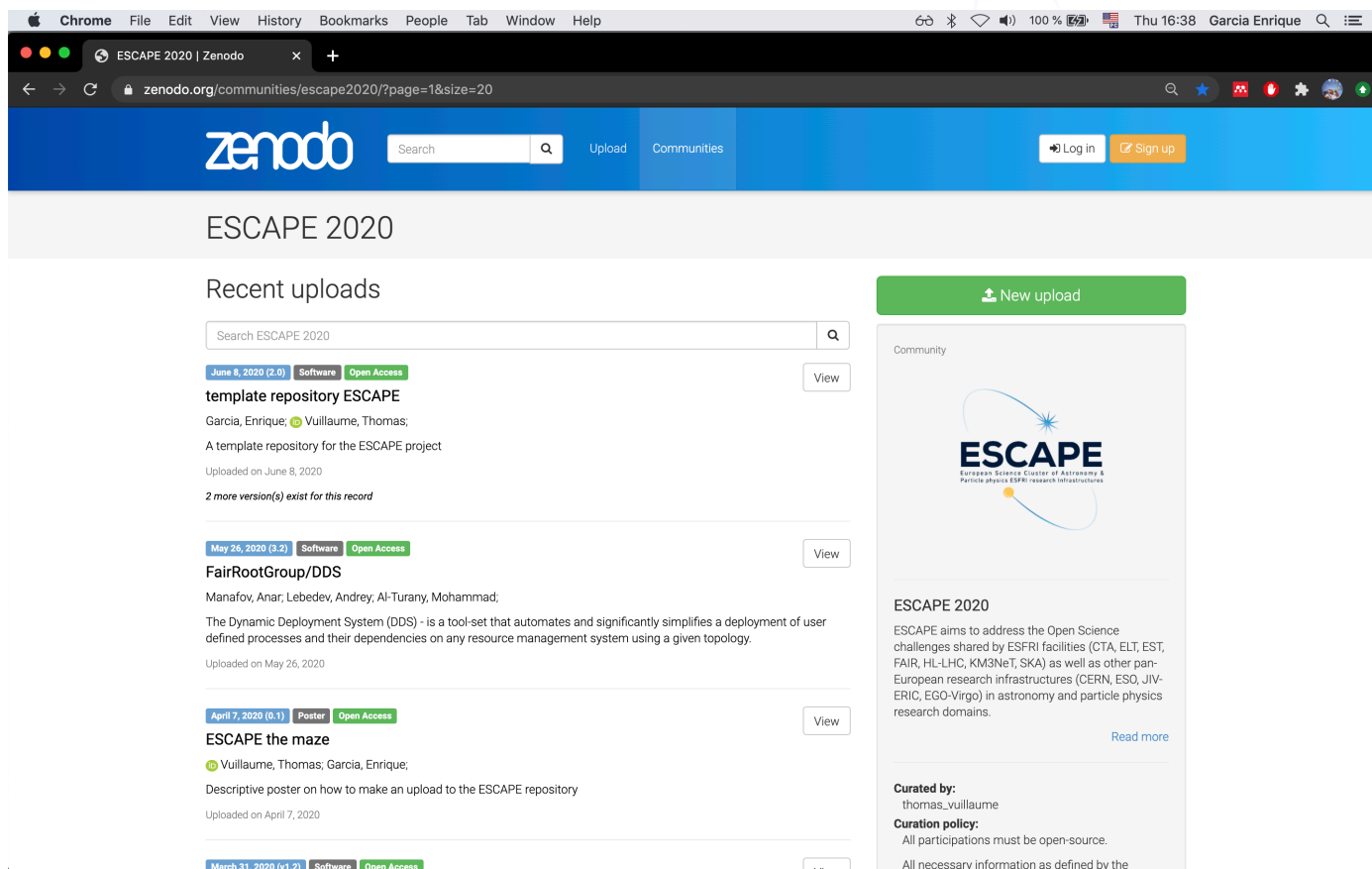
	+	-
Zenodo as a service	<ul style="list-style-type: none"> - Free - No setup - CERN support - Secure, very long time, archive - Always up to date - Integration with OpenAIRE 	<ul style="list-style-type: none"> - Size limitation to 50GB per upload (an agreement is possible for bigger datasets) - No multiple ownership as of today - No multiple curators as of today - Limited customization
Zenodo as a software (later InvenioRDM)	<ul style="list-style-type: none"> - Better branding - Own the data - Custom front-end - Custom A&A 	<ul style="list-style-type: none"> - Need time and money to dev/setup/install - Need computing infrastructure - Need maintenance - Future ?

● Slide from T. Vuillaume (24/07)



ESCAPE 2020 Community

<https://zenodo.org/communities/escape2020/>



Recent uploads

Search ESCAPE 2020

June 8, 2020 (2.0) Software Open Access View

template repository ESCAPE
 Garcia, Enrique; Vuillaume, Thomas;
 A template repository for the ESCAPE project
 Uploaded on June 8, 2020
 2 more version(s) exist for this record

May 26, 2020 (3.2) Software Open Access View

FairRootGroup/DDS
 Manafov, Anar; Lebedev, Andrey; Al-Turany, Mohammad;
 The Dynamic Deployment System (DDS) - is a tool-set that automates and significantly simplifies a deployment of user defined processes and their dependencies on any resource management system using a given topology.
 Uploaded on May 26, 2020

April 7, 2020 (0.1) Poster Open Access View

ESCAPE the maze
 Vuillaume, Thomas; Garcia, Enrique;
 Descriptive poster on how to make an upload to the ESCAPE repository
 Uploaded on April 7, 2020

March 31, 2020 (v1.2) Software Open Access View

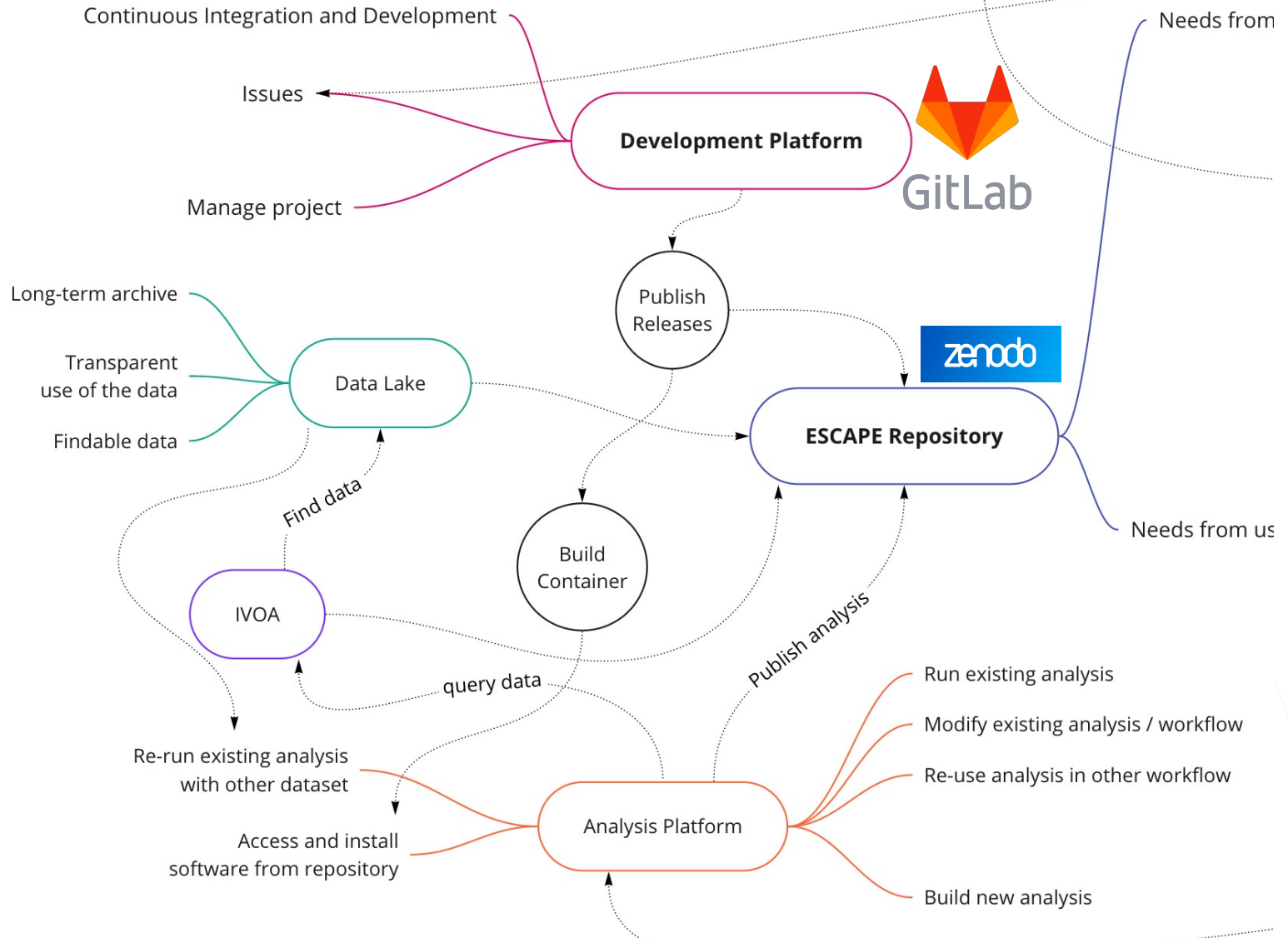
ESCAPE 2020
 ESCAPE aims to address the Open Science challenges shared by ESFRI facilities (CTA, ELI, EST, FAIR, HL-LHC, KM3Net, SKA) as well as other pan-European research infrastructures (CERN, ESO, JIVE, EGO-Virgo) in astronomy and particle physics research domains.
[Read more](#)

Curated by:
 thomas_vuillaume

Curation policy:
 All participations must be open-source.
 All necessary information as defined by the



Connection of services



Connection of services. Containers

- Satisfy all the FAIR principles.
- Allow an easy and fast reproducibility of very complex environments.
- The expected/proposed way of reproducing studies in the analysis platform.



Connection dev platform - repository

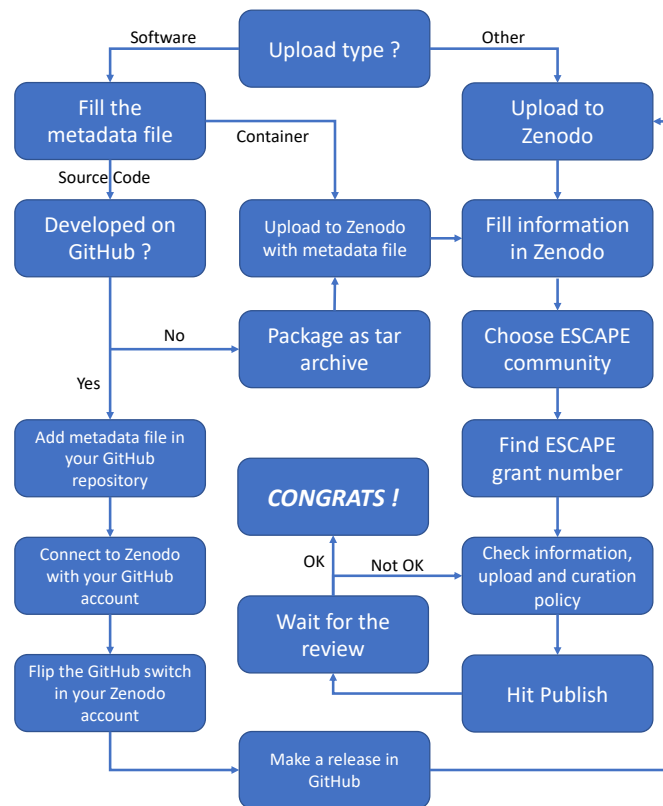
● Upload a project to Zenodo:

● *ESCAPE* the maze

● zenodo.org/record/3885172


ESCAPE the maze

How to upload your project to the repository



ESCAPE maze v0.1, 08/04/2020

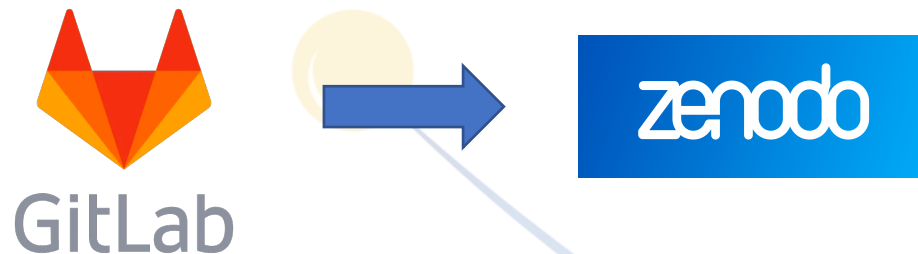


Automatic connection GitLab-Zenodo

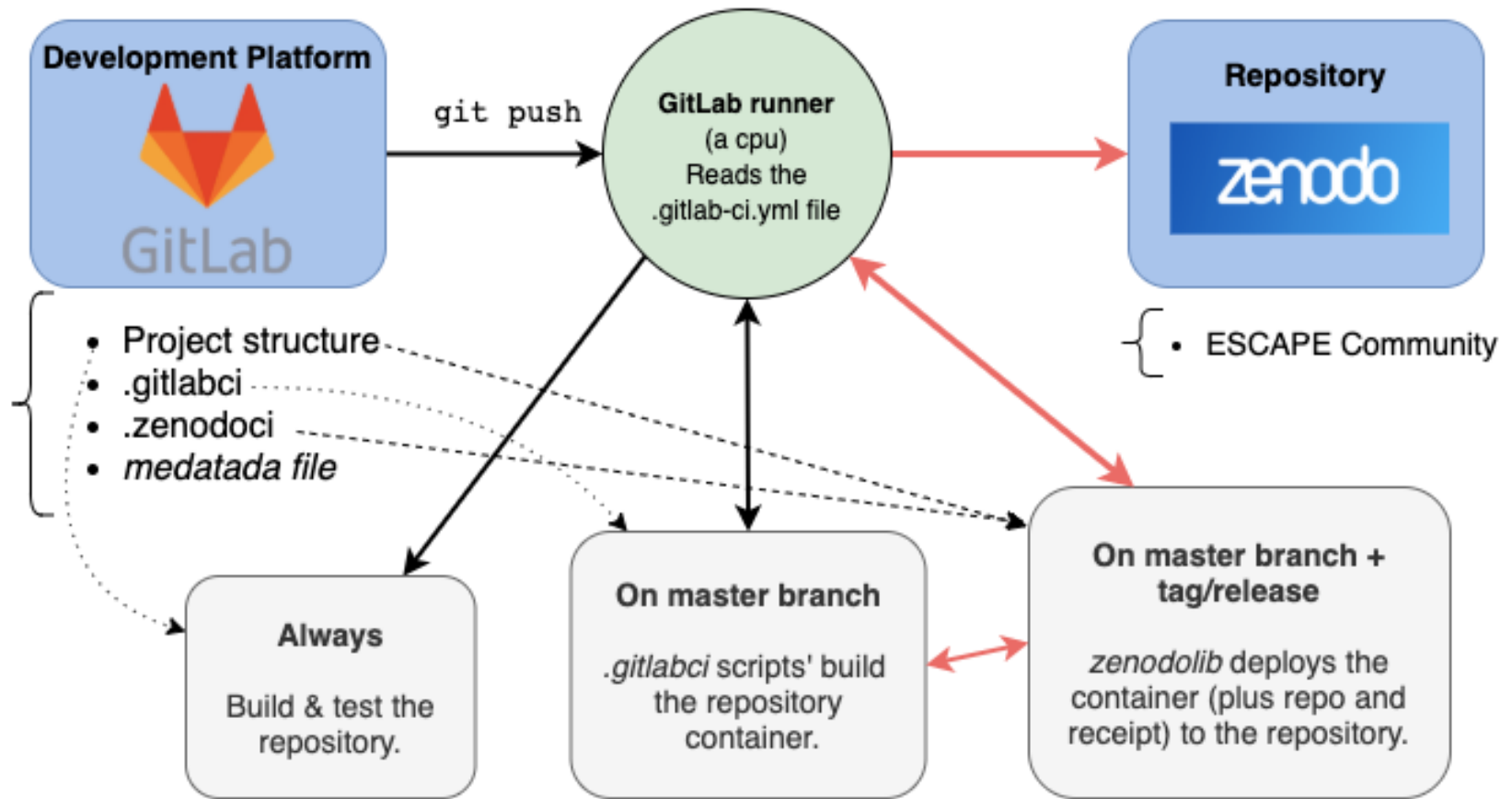
- Development platforms provide very useful tools for continuous integration and continuous delivery/deployment (CI/CD).

https://gitlab.in2p3.fr/escape2020/escape/template_project_escape

- Use the development platform to develop, build and test,
- and deploy a container/project to the repository (or endpoint).



Automatic connection GitLab-Zenodo



Automatic connection GitLab-Zenodo (I)

● Following FAIR principles:

● Singularity-CI builders

1. Invokes DockerHub
2. Creates a Singularity container
3. Uploads the container to an endpoint (+ dev plat)





- <https://github.com/singularityhub/singularity-ci>
- Tutorial : <https://vsoch.github.io/2018/gitlab-singularity-ci/>

Singularity-CI Builders

This is an endpoint to show you how you can build Singularity containers on different Continuous Integration Services. The templates provided are served by Gitlab or Github, depending on where the CI services are commonly used.

General Templates

The following are general builders that you can customize to build on Travis, CircleCI, GitHub, or GitLab.

Service	Status	Template	
GitLab		singularityhub/gitlab-ci	
TravisCI	build passing	singularityhub/travis-ci	
CircleCI	PASSED	singularityhub/circle-ci	
GitHub		singularityhub/github-ci	



Automatic connection GitLab-Zenodo (II)

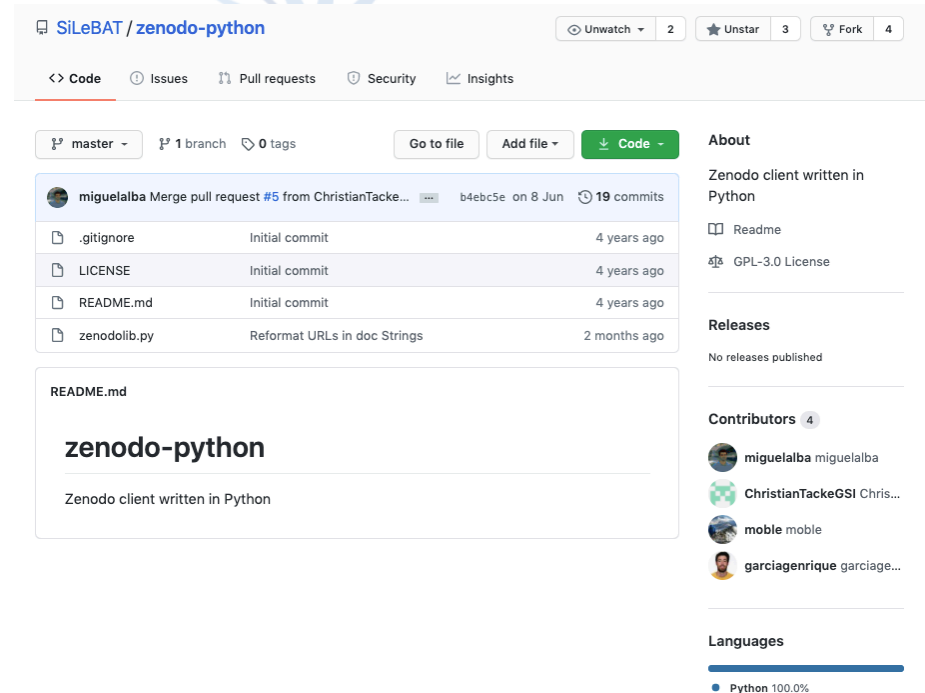
● Following FAIR principles:

● *zenodo-python* repository

- Provides the *zenodolib* library
- Manages upload to Zenodo through their API
- Please use Zenodo sandbox for tests !

<https://sandbox.zenodo.org/deposit>

- <https://github.com/SiLeBAT/zenodo-python/>
- REST API tutorial: <https://developers.zenodo.org>



The screenshot shows the GitLab repository page for `SiLeBAT/zenodo-python`. The repository is on the `master` branch, has 1 branch and 0 tags. It contains 19 commits. The file list includes `.gitignore`, `LICENSE`, `README.md`, and `zenodolib.py`. The `README.md` file is open, showing the title `zenodo-python` and the description "Zenodo client written in Python". The right sidebar shows the repository's license as GPL-3.0, no releases published, and four contributors: miguelalba, ChristianTackeGSI, mobile, and garciagenrique. The language statistics show Python at 100.0%.



Automatic con. GitLab-Zenodo (wrap up)

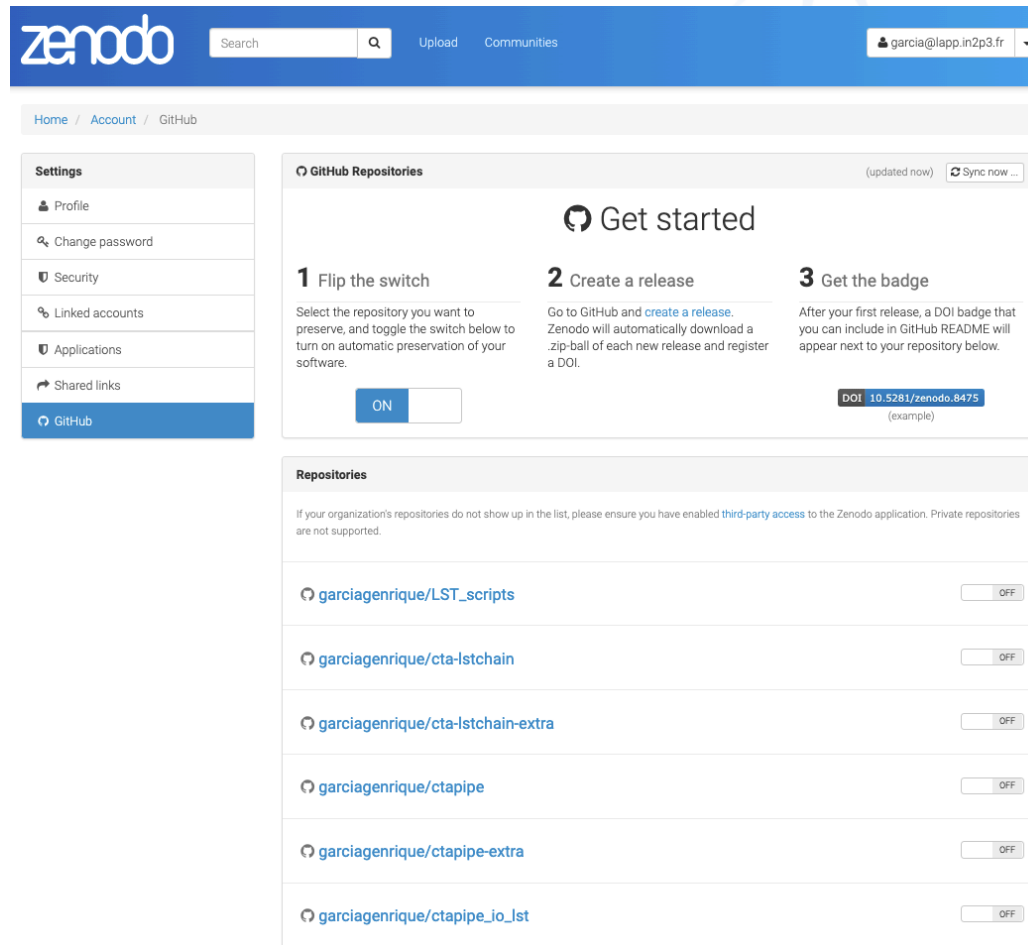
1. Clone/Fork the template repository
https://gitlab.in2p3.fr/escape2020/escape/template_project_escape
2. Create an access token to communicate with the Zenodo API
3. Configure the .gitlab-ci.yml file
4. Configure of the GitLab CI/CD environment variables
5. Adapt/update the uploading/new version scripts
6. Create a release

Follow the **OSSR on-boarding demonstrator**
at 13h15 (27/07/2020)



Automatic connection GitHub-Zenodo

☉ Zenodo → Account → Settings



The screenshot shows the Zenodo user interface. At the top, there is a blue navigation bar with the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. The user's email 'garcia@lapp.in2p3.fr' is visible in the top right. Below the navigation bar, the breadcrumb trail reads 'Home / Account / GitHub'. On the left, a 'Settings' sidebar is open, with 'GitHub' selected. The main content area is titled 'GitHub Repositories' and includes a 'Sync now...' button. A 'Get started' section contains three numbered steps: 1. Flip the switch (with an 'ON' toggle), 2. Create a release, and 3. Get the badge (with an example DOI: 10.5281/zenodo.8475). Below this, a 'Repositories' table lists several repositories with their respective integration status (all are currently 'OFF').



Connection of services. Metadata (I)

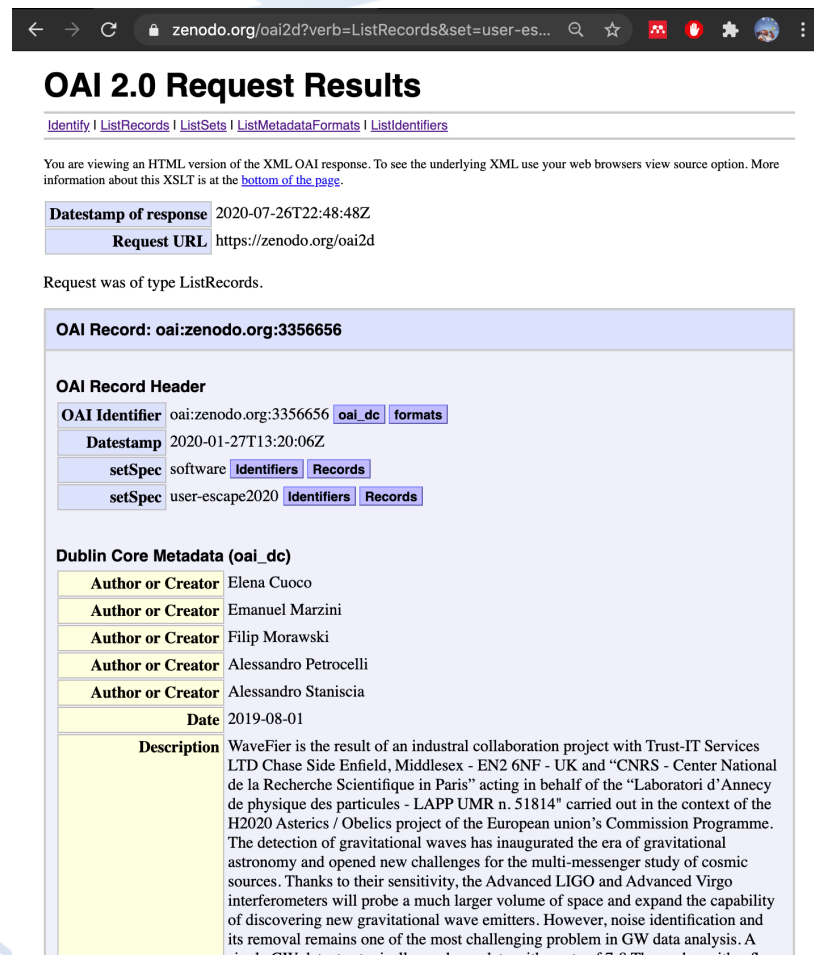
- Allows fulfilling all the FAIR principles in a single standard
 - Findable
 - Accessible
 - Interoperable
 - **Will allow the connection with other services ! (ESAP, Portal)**
 - Reusable :
 - License and provenance



Connection of services. Metadata (II)

Zenodo:

- harvest metadata using the OAI-PMH
 - Open Archive Initiative – Protocol of Metadata Harvesting
- accepts and export different metadata formats:
 - oai_datacite(3, 4,...)
 - marcxml
 - oai_dc
 - dcat
 - marc21



zenodo.org/oai2d?verb=ListRecords&set=user-es...

OAI 2.0 Request Results

[Identify](#) | [ListRecords](#) | [ListSets](#) | [ListMetadataFormats](#) | [ListIdentifiers](#)

You are viewing an HTML version of the XML OAI response. To see the underlying XML use your web browsers view source option. More information about this XSLT is at the [bottom of the page](#).

Datestamp of response 2020-07-26T22:48:48Z

Request URL https://zenodo.org/oai2d

Request was of type ListRecords.

OAI Record: oai:zenodo.org:3356656

OAI Record Header

OAI Identifier	oai:zenodo.org:3356656	oai_dc	formats
Datestamp	2020-01-27T13:20:06Z		
setSpec	software	Identifiers	Records
setSpec	user-escape2020	Identifiers	Records

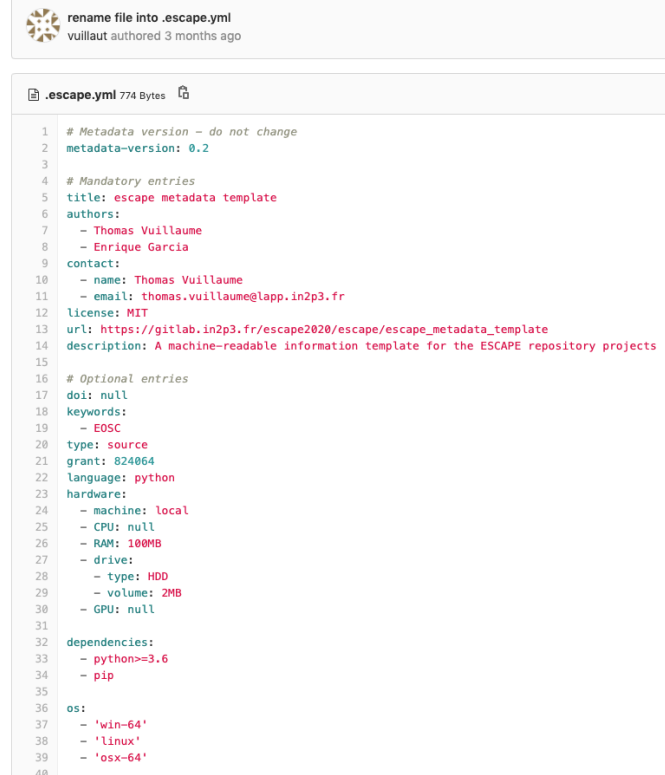
Dublin Core Metadata (oai_dc)

Author or Creator	Elena Cuoco
Author or Creator	Emanuel Marzini
Author or Creator	Filip Morawski
Author or Creator	Alessandro Petrocelli
Author or Creator	Alessandro Staniscia
Date	2019-08-01
Description	WaveFier is the result of an industrial collaboration project with Trust-IT Services LTD Chase Side Enfield, Middlesex - EN2 6NF - UK and "CNRS - Center National de la Recherche Scientifique in Paris" acting in behalf of the "Laboratori d'Annee de physique des particules - LAPP UMR n. 51814" carried out in the context of the H2020 Asterics / Obelics project of the European union's Commission Programme. The detection of gravitational waves has inaugurated the era of gravitational astronomy and opened new challenges for the multi-messenger study of cosmic sources. Thanks to their sensitivity, the Advanced LIGO and Advanced Virgo interferometers will probe a much larger volume of space and expand the capability of discovering new gravitational wave emitters. However, noise identification and its removal remains one of the most challenging problem in GW data analysis. A single GW detector typically produces data with a rate of 7-8 Tb per day with a flux



Connection of services. Metadata (III)

- Started discussion (Mark Kettenis' talk – 24/07)
 - Metadata + I/o info + computing resources + media types
- ESCAPE metadata template **basic proposal**
https://gitlab.in2p3.fr/escape2020/escape/escape_metadata_template
 - Gathers all the mentioned info
- Need to converge to a
 - Personalised schema ?
 - DataCite/CodeMeta – based ?



rename file into .escape.yml
vuillaud authored 3 months ago

.escape.yml 774 Bytes Edit

```
1 # Metadata version – do not change
2 metadata-version: 0.2
3
4 # Mandatory entries
5 title: escape metadata template
6 authors:
7   - Thomas Vuillaume
8   - Enrique Garcia
9 contact:
10  - name: Thomas Vuillaume
11  - email: thomas.vuillaume@lapp.in2p3.fr
12 license: MIT
13 url: https://gitlab.in2p3.fr/escape2020/escape/escape_metadata_template
14 description: A machine-readable information template for the ESCAPE repository projects
15
16 # Optional entries
17 doi: null
18 keywords:
19   - EOSC
20 type: source
21 grant: 824064
22 language: python
23 hardware:
24   - machine: local
25   - CPU: null
26   - RAM: 100MB
27   - drive:
28     - type: HDD
29     - volume: 2MB
30   - GPU: null
31
32 dependencies:
33   - python>=3.6
34   - pip
35
36 os:
37   - 'win-64'
38   - 'linux'
39   - 'osx-64'
40
```


Repository License

- FAIR principles might lead to confusing situations
 - BSD 3-Clause → Singularity-CI builders
 - GPLv3 → zenodo-python repository
- Both licenses are compatible*, however GPLv3 is copyleft
- *It can be combined code released under the other license with code released under the GNU GPL



OSSR ToDo list

- Improve OSSR onboarding / uploads
 - More exhaustive documentation
- Implement ESCAPE ecosystem-like metadata schema
- Aim for an integrated virtual environment

