



ESCAPE

European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures

WP3 OSSR – ESCAPE Repository

Thomas Vuillaume

ESCAPE WOSSL, 24/07/2020



Goals

- Maximise software re-use
- Foster co-developments
- Identify open standards for software releases
- Investigate data mining tools and new analysis techniques
- Expose community-specific data services under the FAIR principles

→ Open-access repository to share scientific software, digital libraries, data analysis, datasets, documentation, tutorials and training activities to be included in the EOSC catalogue



Important note

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



Why is that important

- To not re-invent existing tools, analysis
- **Combine** efforts
- Go further in the **multi-messenger** analysis domain
- Promote and ensure **open, reproducible science**, encouraging the implementation of the FAIR principles



An example of open science project : The Crab bundle

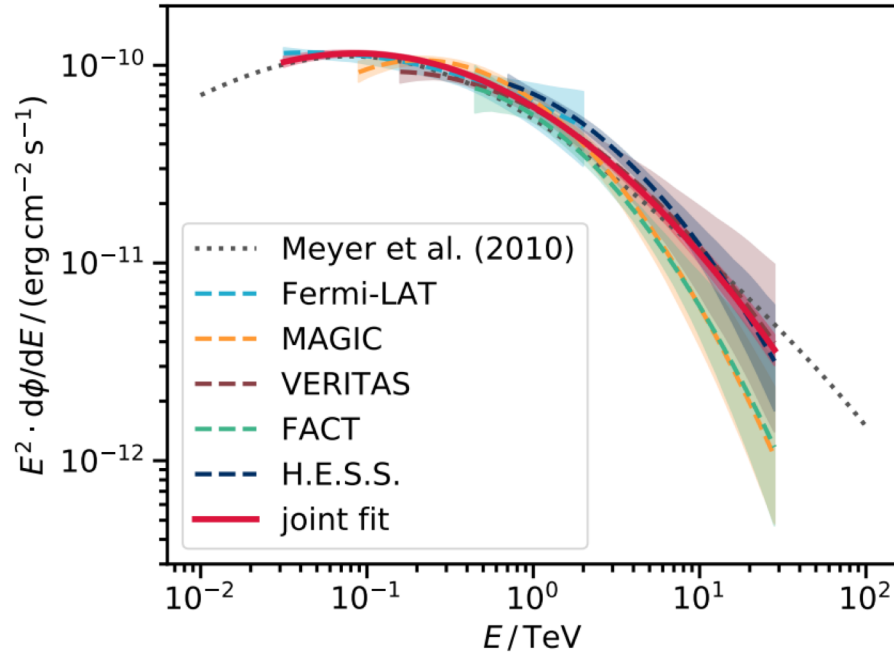


Fig. 2. Crab nebula SED for individual instrument fits and from the joint fit. Single-instrument results are represented with dashed lines, the fit of all the datasets together, labelled as joint, is represented as a thick, solid red line. The shaded areas represent the SED error bands whose calculation is explained in Sect. 3.2. The dotted line shows the model in Meyer et al. (2010).

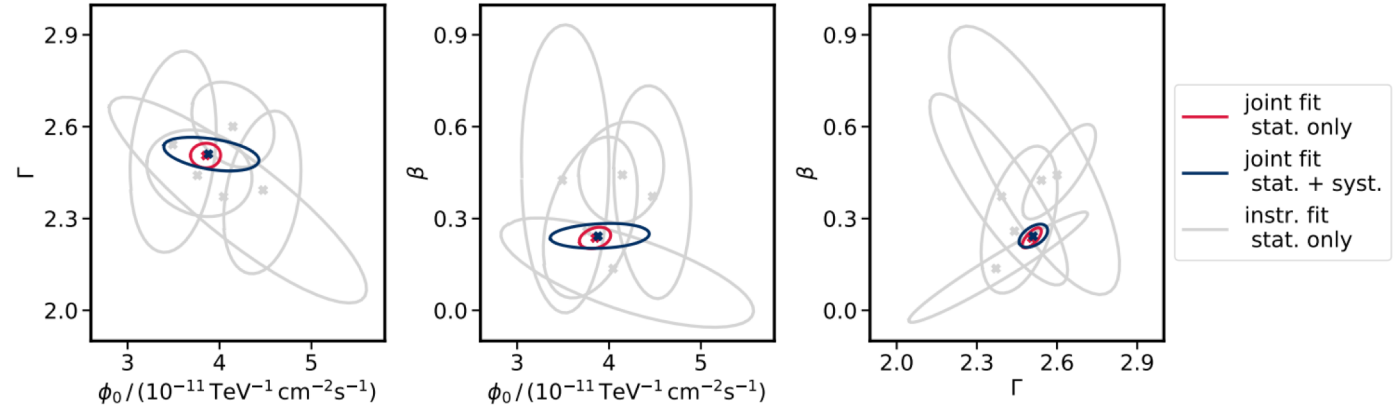


Fig. 5. Likelihood contours corresponding to 68% probability content for the fitted spectral parameters (ϕ_0, Γ, β) , for the likelihood in Eq. (1) (red) and the likelihood in Eq. (5) (blue). Results from the individual instruments with the likelihood in Eq. (1) are shown in grey.

Multi-instrument gamma-ray analysis of the Crab Nebulae with Fermi-LAT, MAGIC, VERITAS, FACT and H.E.S.S.

****Towards open and reproducible multi-instrument analysis in gamma-ray astronomy, C. Nigro et al, in A&A 625 (2019)***



An example of open science project : The Crab bundle

The Crab multi-instrument gamma-ray analysis with MAGIC, VERITAS, FACT and H.E.S.S.

<https://github.com/open-gamma-ray-astro/joint-crab/tree/v0.1>

<https://zenodo.org/record/2381863#.XkxcD5NKhhA>

license

License (for files):
[BSD 3-Clause Clear License](#)

Versions

Version v0.1 Dec 18, 2018
[10.5281/zenodo.2381863](https://doi.org/10.5281/zenodo.2381863)

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.2381862](https://doi.org/10.5281/zenodo.2381862). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

Share



Cite as

C. Nigro, C. Deil, R. Zanin, T. Hassan, J. King, J.E. Ruiz, ... A. Sinha. (2018, December 18). The joint-crab bundle (Version v0.1). Zenodo. <http://doi.org/10.5281/zenodo.2381863>

Start typing a citation style...

Export

BibTeX CSL DataCite Dublin Core DCAT JSON JSON-LD GeoJSON MARCXML [Mendeley](#)

Cited by

Files (10.1 MB)

Name	Size
open-gamma-ray-astro/joint-crab-v0.1.zip	10.1 MB

md5:55cd7fed91c53f54a96e32945e764665

Citations 1

Show only: Literature (1) Dataset (0) Software (0) Unknown (0)

Citations to this version

[Towards open and reproducible multi-instrument analysis in ...](#) 2019
 Nigro, C. et al. (DOI: 10.1051/0004-6361/201834938)

Link to project and article

Cite as

December 18, 2018

Software Open Access

The joint-crab bundle

C. Nigro; C. Deil; R. Zanin; T. Hassan; J. King; J.E. Ruiz; L. Saha; R. Terrier; K. Bruegge; M. Noethe; R. Bird; T. T. Y. Lin; J. Aleksic; C. Boisson; J.L. Contreras; A. Donath; L. Jouvin; N. Kelley-Hoskins; B. Khelifi; K. Kosack; J. Rico; A. Sinha

This **joint-crab** bundle allows for a first reproducible multi-instrument gamma-ray analysis, achieved by using the **prototypical DL3 data format** and the open-source **Gammapy** software package, for a small set of MAGIC, VERITAS, FACT, and H.E.S.S. Crab nebula observations.

Preview

[joint-crab-v0.1.zip](#)

- open-gamma-ray-astro-joint-crab-752a165
 - .gitignore 1.2 kB
 - 1_data.ipynb 3.0 kB
 - 2_results.ipynb 6.0 kB
 - 3_systematics.ipynb 100.8 kB
 - 4_naima.ipynb 72.0 kB
 - 5_crab_pulsar_nebula_sed.ipynb 65.4 kB
 - Dockerfile 604 Bytes
 - LICENSE 1.5 kB
 - README.md 4.3 kB
 - analysis.md 2.9 kB
 - binder
 - environment.yml 443 Bytes
 - data 512 Bytes
 - README.md
 - fact
 - 20131103_103_dl3.fits 25.9 kB
 - 20131103_104_dl3.fits 25.9 kB
 - 20131103_105_dl3.fits 23.0 kB

Source code and data

141 views
 12 downloads

[See more details...](#)

Indexed in

Publication date:
December 18, 2018

DOI:
DOI: [10.5281/zenodo.2381863](https://doi.org/10.5281/zenodo.2381863)

Keyword(s):
Astronomy Gamma-rays Data analysis

Related identifiers:
Referenced by <https://arxiv.org/abs/1903.06621>

Alternate identifiers:
<https://github.com/open-gamma-ray-astro/joint-crab/tree/v0.1>

Communities:
Astronomy-General



An example of open science project : The Crab bundle

The Crab multi-instrument gamma-ray analysis with MAGIC, VERITAS, FACT and H.E.S.S.

<https://github.com/open-gamma-ray-astro/joint-crab/tree/v0.1>

<https://zenodo.org/record/2381863#.XkxcD5NKhhA>

Towards multi-instrument and reproducible gamma-ray analysis

DOI 10.5281/zenodo.1409831

launch binder

This repository contains contere Docker image or via the MyBin

Name	Last Modified
binder	10 days ago
data	10 days ago
joint_crab	10 days ago
presentations	10 days ago
results	10 days ago
1_data.ipynb	10 days ago
2_results.ipynb	10 days ago
3_systematics.ipynb	10 days ago
4_naima.ipynb	10 days ago
5_crab_pulsar_nebula...	10 days ago
analysis.md	10 days ago
docker.md	10 days ago
Dockerfile	10 days ago
LICENSE	10 days ago
make.py	10 days ago
README.md	10 days ago

```

m = results[name].marginal(names)
ellipse = m.to_matplotlib_ellipse(ec=ecolor, fc="none", label=name)
ax.add_patch(ellipse)
ax.plot(m.mean[0], m.mean[1], ".", color=ecolor)
ax.set_xlabel(names[0])
ax.set_ylabel(names[1])
if legend:
    ax.legend(loc=1)

def plot_ellipses(results):
    fig, axes = plt.subplots(nrows=1, ncols=3, figsize=(10, 3))
    plot_ellipses_panel(results, ["phi", "gamma"], axes[0], legend=False)
    plot_ellipses_panel(results, ["phi", "beta"], axes[1], legend=False)
    plot_ellipses_panel(results, ["gamma", "beta"], axes[2], legend=True)
    fig.tight_layout()

plot_ellipses(results)

```

As mentioned at the start, we can see that the joint fit result is approximately given by the product of the multivariate normal approximation for th of each individual dataset in this analysis.

hub.docker.com/r/gammapy/joint-crab

docker hub Search for great content (e.g. mysql) Explore Sign In Pricing Get Started

gammapy/joint-crab ☆

By gammapy · Updated 9 months ago

Reproduce the results published in "Towards multi-instrument and reproducible gamma-ray analysis"

Container

Overview Tags Dockerfile Builds

Towards open and reproducible multi-instrument analysis in gamma-ray astronomy

DOI 10.5281/zenodo.2381863

launch binder

This repository contains material related to the publication Towards multi-instrument and reproducible gamma-ray analysis appeared in A&A 625, A10 (2019)

DOI <https://doi.org/10.1051/0004-6361/201834938>

Docker Pull Command

```
docker pull gammapy/joint-crab
```

Owner

gammapy

Source Repository

- Docker to ensure reproducibility



ESCAPE Virtual Environment

Now imagine this as a **standard**,
in an **integrated environment** - with a single login,
allowing you to (re-)**run** (part of) **any analysis**,
with another **dataset**,
and easily **publish** your new results,
automatically **giving credit** to original analysis, datasets, workflows...



ESCAPE Virtual Environment

How ?



ESCAPE Virtual Environment

1. Promote and implement **FAIR*** principles

- **Findable** → Rich metadata, **unique** and **persistent** identifier
- **Accessible** → Metadata and data are **understandable** to **humans** and **machines**. Data is deposited in a **trusted repository**.
- **Interoperable** → Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- **Reusable** → Data and collections have a **clear** usage **licenses** and provide accurate information on **provenance**.

2. Provide the **infrastructure** and **services** to foster FAIR (effortless) contributions

→ ESCAPE Virtual Environment



ESCAPE Virtual Environment (a biased view)

Jane as a new science idea



Collaborative Platform

Chat
Forum

Development Platform

Continuous Integration and Development

Issues

Manage project

Publish Releases

Data Lake

Long-term archive
Transparent use of the data
Findable data

Find data

Build Container

IVOA

query data

Publish analysis

Analysis Platform

Run existing analysis
Modify existing analysis / workflow
Re-use analysis in other workflow
Build new analysis

Re-run existing analysis with other dataset

Access and install software from repository

Answer questions and provide support to users
Discuss with other developers
Publish software / datasets

Visibility
Citable

Comment / Ask questions

Rate
Re-use
Cite

Discovery of software / services

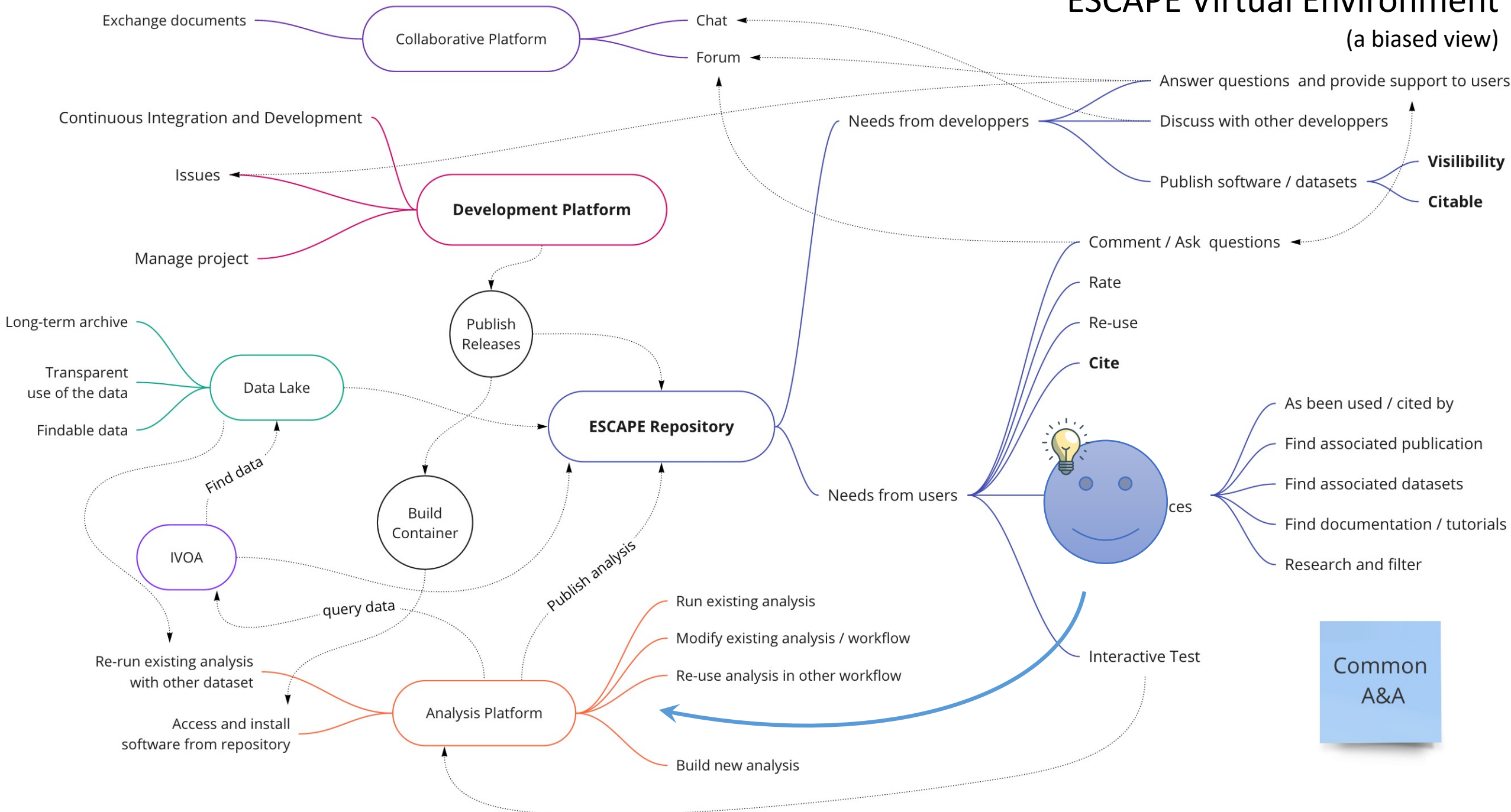
As been used / cited by
Find associated publication
Find associated datasets
Find documentation / tutorials
Research and filter

Interactive Test

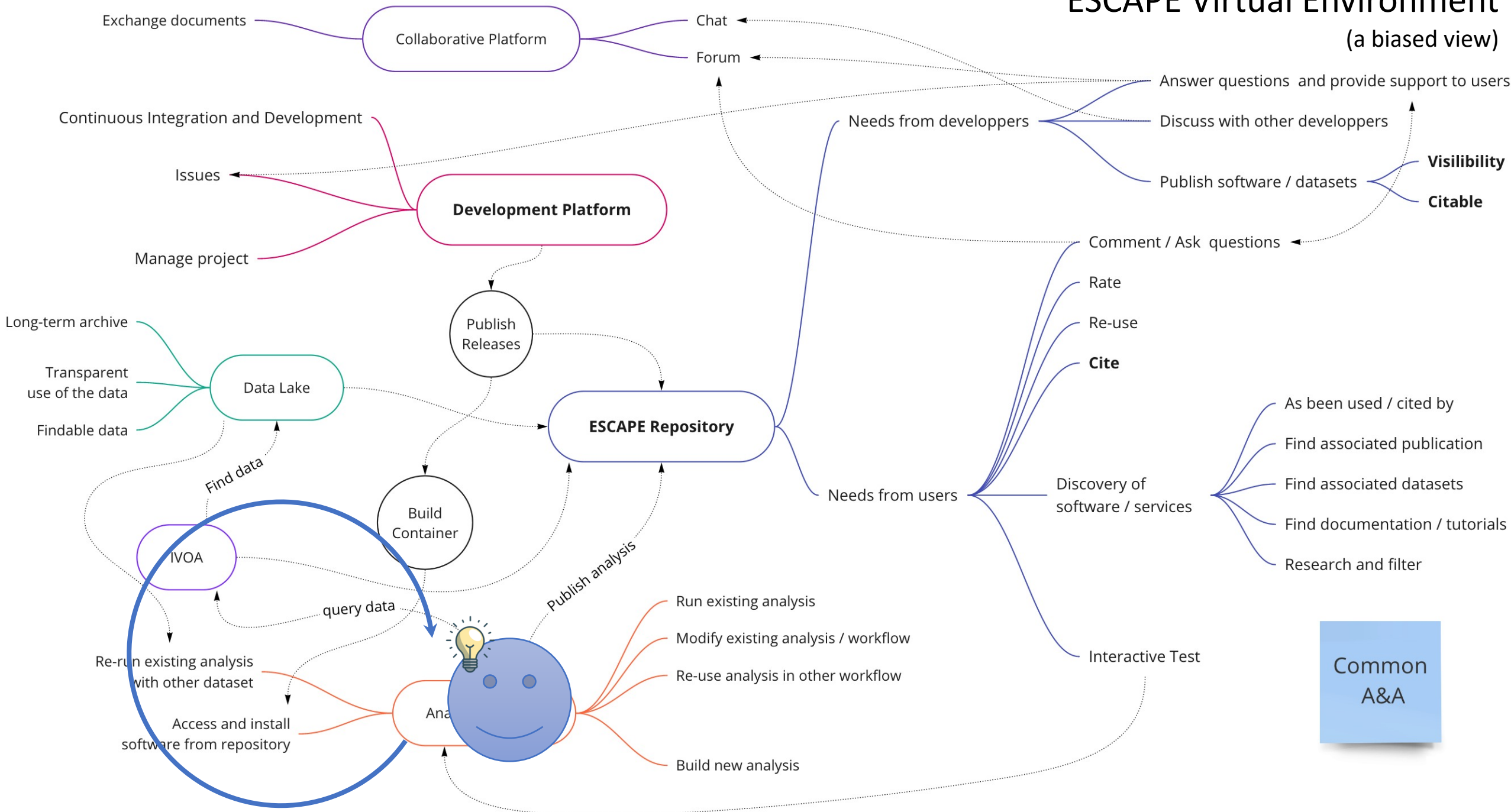
Common A&A

ESCAPE Virtual Environment

(a biased view)

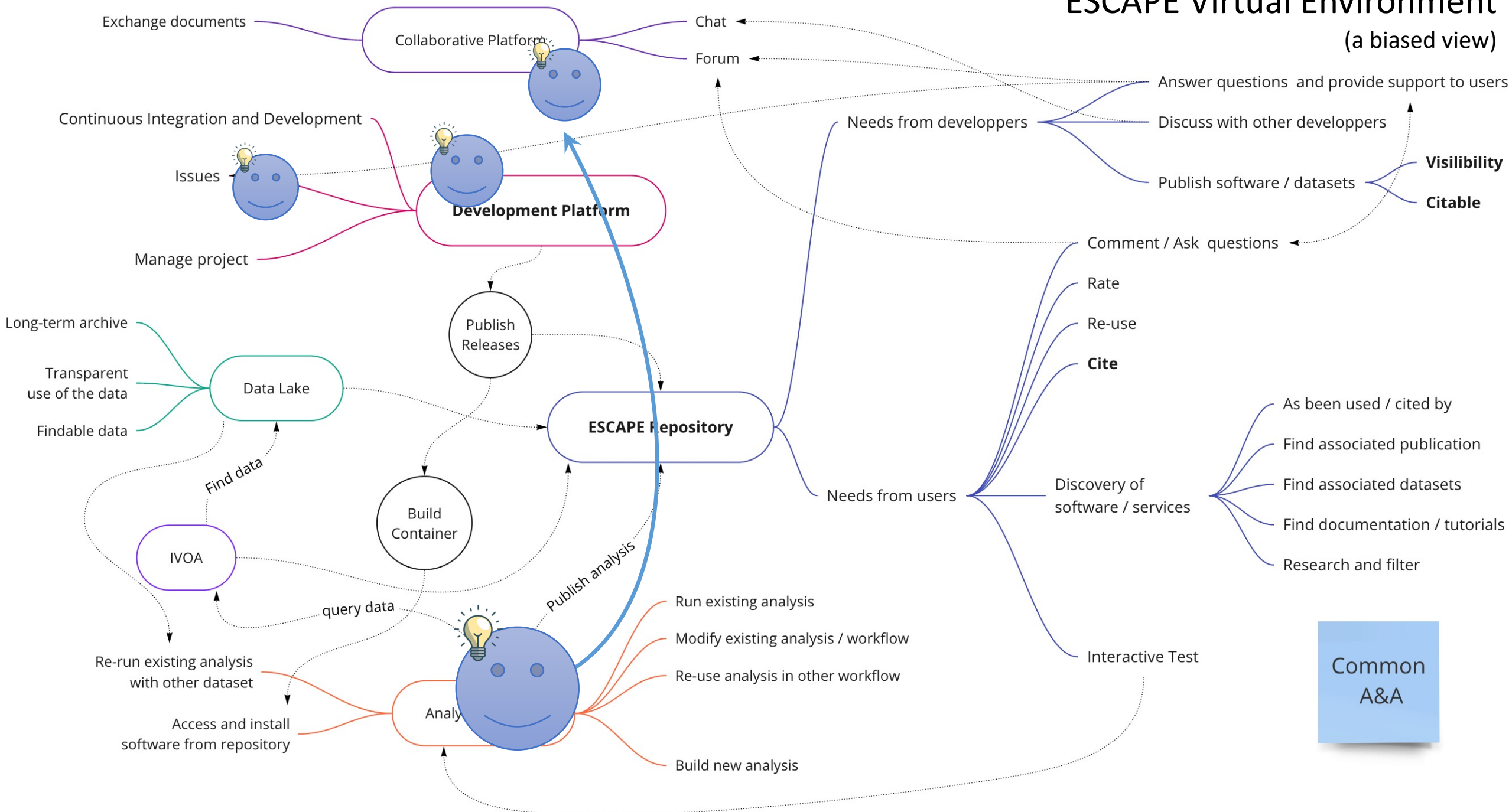


ESCAPE Virtual Environment (a biased view)



ESCAPE Virtual Environment

(a biased view)



Jane's new idea could be quickly tested in a virtual environment,
with the required software pre-installed,
and with access to previous and new datasets.

In the process, she not only made new discoveries,
she became an active developer of the tools,
and published openly her analysis,
allowing other to cross-check, re-use, cite and build on top of it.

→ the repository must be **more than an archive**



Technical implementation ?



Partners survey on the repository

- *Do you have one or several repositories in place within your ESF/RI and how are they implemented?*
- *What features do you expect from the OSSR repository?*
- *How would your ESF/RI transfer the software and services to the OSSR (transferring the full development, adding stable releases or similar?)*
- *What software licenses, versioning and means of preservation do you employ?*
- *How does your ESF/RI implement the FAIR principles to software, services and relevant data to test those, what means are in place for long-term preservation, maintenance and curation?*
- *What features do you expect from the OSSR repository?*
- ...



Development platform: Gitlab

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

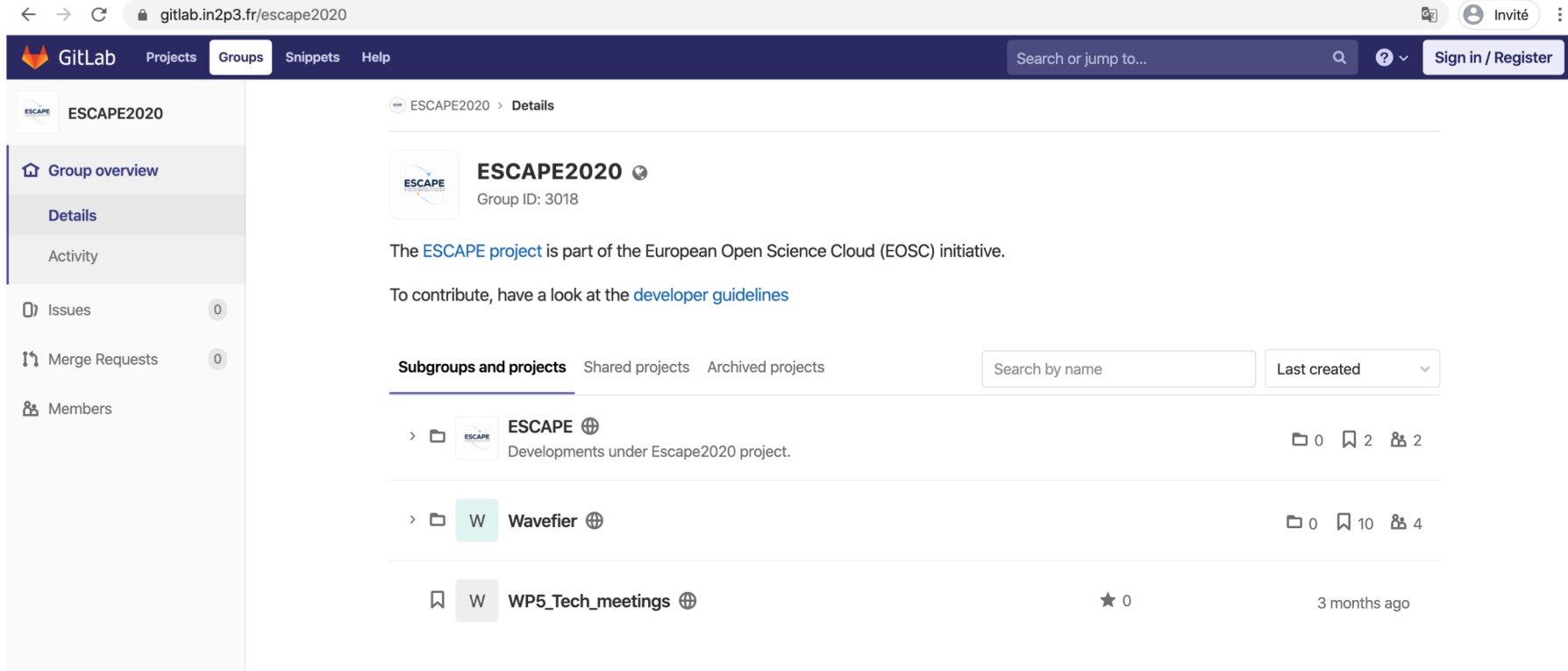
Gitlab IN2P3

Welcome !

IN2P3 AND ACADEMIC USERS MUST **SIGN-UP AND REGISTER** WITH **[Shibboleth]**

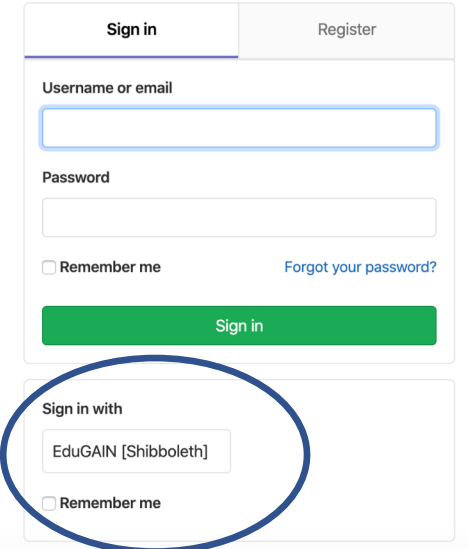
New users are automatically set as *External* for validation, **except** when registering with **[Shibboleth]** (EduGAIN, Fédération d'Identités RENATER).

Please take a moment to read the [documentation](#)



The screenshot shows the GitLab IN2P3 website. The browser address bar displays `gitlab.in2p3.fr/escape2020`. The page title is "ESCAPE2020" and the group ID is 3018. The main content area includes a description: "The ESCAPE project is part of the European Open Science Cloud (EOSC) initiative. To contribute, have a look at the [developer guidelines](#)". Below this, there is a section for "Subgroups and projects" with a search bar and a dropdown menu set to "Last created". The subgroups listed are:

- ESCAPE** (Group ID: 3018): Developments under Escape2020 project. 0 issues, 2 merge requests, 2 members.
- Wavefier**: 0 issues, 10 merge requests, 4 members.
- WP5_Tech_meetings**: 0 stars, created 3 months ago.



The screenshot shows the GitLab login form. It has two tabs: "Sign in" (selected) and "Register". The form includes fields for "Username or email" and "Password", a "Remember me" checkbox, and a "Forgot your password?" link. A green "Sign in" button is at the bottom. Below the main form, there is a "Sign in with" section with a button for "EduGAIN [Shibboleth]" and a "Remember me" checkbox. The "EduGAIN [Shibboleth]" button is circled in blue.

Sign in with EduGain



Development platform: Gitlab

	+	-
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



Development platform: Gitlab

	+	-
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



Repository: concept implementation with Zenodo service

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

ESCAPE 2020

All versions

Access Right

Open (4)

File Type

Zip (2)

Gz (1)

Md (1)

Keywords

CTA (2)

ESCAPE (1)

Machine Learning, Big Data, Apache Kafka, Gravitational Wave (1)

Convolution (1)

Data Format (1)

Machine Learning (1)

Found 4 results.

< 1 >

Sort by:


Most recent

asc.

View

February 7, 2020 (0.1) Software Open Access

ctape_io_mchdf5

 Vuillaume, Thomas; Aubert, Pierre; Garcia, Enrique;

ctape plugin for reading and converting Monte-Carlo files (contains the same information as Simtel files)

Uploaded on February 7, 2020

December 12, 2019 (1.0) Software Open Access

Template for an ESCAPE project

Garcia, Enrique;  Vuillaume, Thomas;

A template to provide software to the ESCAPE project (<https://projectescape.eu/>)

Uploaded on December 12, 2019

View

August 1, 2019 (1.0) Software Open Access

A prototype for a real time pipeline for the detection of transient signals and their automatic classification

 Elena Cuoco; Emanuel Marzini; Filip Morawski; Alessandro Petrocelli; Alessandro Staniscia;

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'Annecy de physique des particules - LA

View

keywords allow to filter by science project or research infrastructure



Zenodo as a service vs Zenodo as an application

	+	-
Zenodo as a service	<ul style="list-style-type: none"> - Free - No setup - CERN support - Secure, very long time, archive - Always up to date 	<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	<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 ?



Zenodo as a service vs Zenodo as an application

	+	-
Zenodo as a service	<ul style="list-style-type: none"> - Free - No setup - CERN support - Secure, very long time, archive - Always up to date 	<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	<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 ?

Note: Some of the limitations from « Zenodo as a service » will be overcome in the future or can be discussed with the Zenodo team through an agreement if needed



Under development or discussion

1. **Connect** the services to create an integrated environment
2. Define **metadata**
 - For humans
 - For machines
3. Define a **template** to upload software (<-> metadata)
4. Propose **license** and **provenance** model (<-> metadata)
5. **Feed** the repository with ESCAPE developments
6. Feed to repository with test **science projects** (see e.g. Dark Matter talk)
7. Provide a single entry point for science projects and ESCAPE virtual environment

See Enrique Garcia's presentations (monday) for technical details



Community Driven Repository, Organised by Science Projects



*Jane as a very big idea
that needs a completely
new development
environment

Objective: ease co-developments and new cross-experiments science projects

- Starting a new Science projet (through validation) sets up a complete virtual environment
- Researchers / Institutes contribute to the science project by publishing software / workflow / data
- The contributions are validated by science project curators
- Users can search the repository or explore it through the science projects



Final word

- Aiming for an integrated virtual environment
- The repository hosts the software
- Central part of the environment
- Only as good as the software it is hosting
 - we need to provide very clear instructions to contribute
 - a lot of points yet to clear and discuss (metadata!)

See Enrique Garcia's presentations on Monday for a status, more technical details, and a demonstration of the repository.

Questions?

