

The logo graphic for ESCAPE features a stylized blue starburst at the top, a thin blue orbital line curving around it, and a small yellow circle at the bottom left of the orbit.

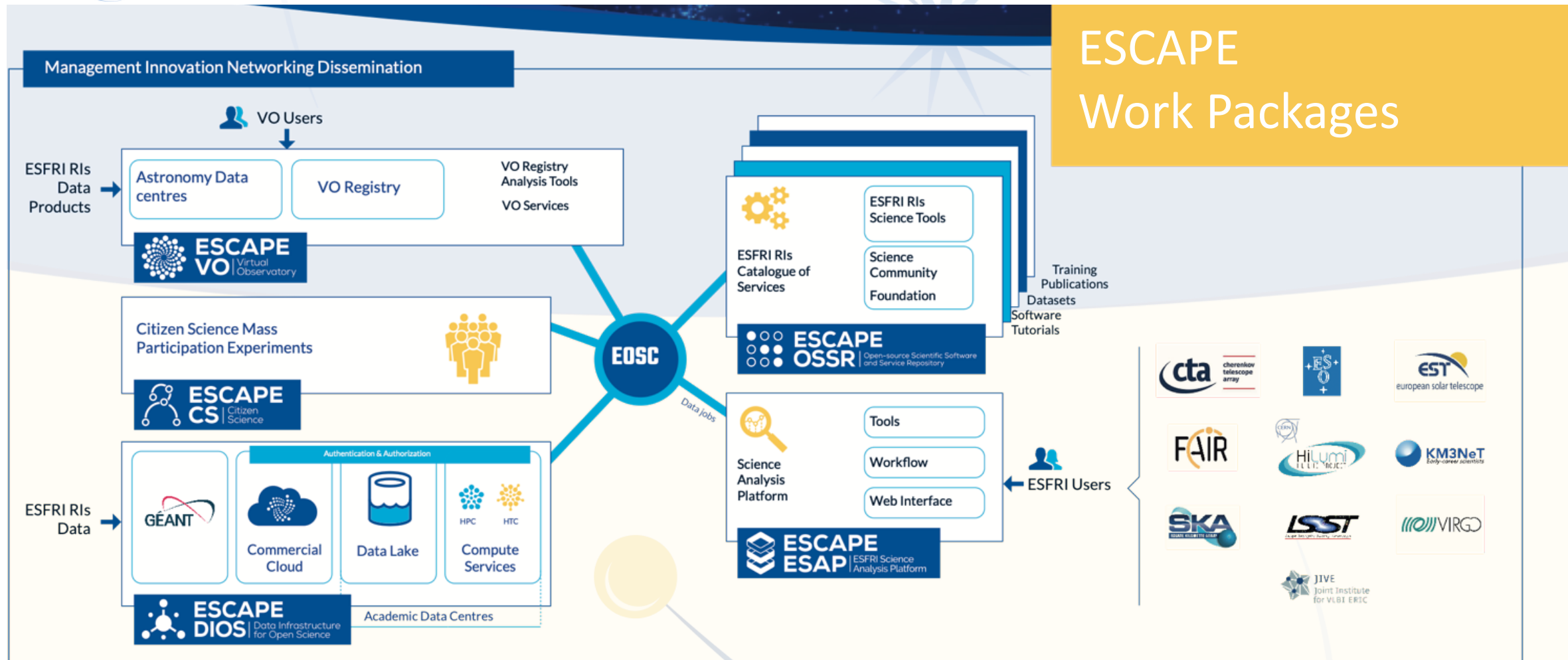
ESCAPE

European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures

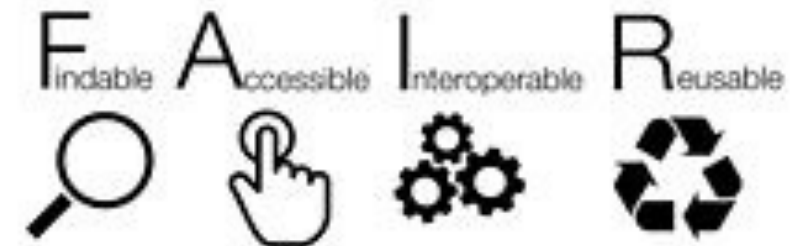
WP3 Integration

Kay Graf, Stelios Voutsinas
Jutta Schnabel, John Swinbank





- **Aim:**
shared open science software and services based on FAIR principles
- **Objectives:**
 - Facilitate and support continuous **development, deployment, exposure and preservation** of partners' software/tools/services
 - Foster **interoperability, software re-use and cross-fertilisation** between ESFRIs (e.g. simulation)
 - Offer an **open innovation environment for open standards** (e.g. workflows, data-formats), **common regulations** and **shared (novel) software** for multi-messenger & multi-probe data
- **All objectives follow:**
 - Paradigm of **enabling open science** – with **software as “first class citizen”**
 - a **community-based, inclusive** approach
 - the **FAIR principles** for open science resources – software and derivatives
 - **Federation** of available resources



Co-Development and Community Engagement



🕒 23 July 2020 to 28 July 2020

ESCAPE Workshop on Open-Source Software Life Cycles

Virtual

Software development is an integral part of modern science, gaining knowledge from data. All ESCAPE partners develop and

[...]



🕒 17 February 2021

Webinar: ESCAPE OSSR | Enhancing science through sharing software - benefits & use cases

Virtual

When: 17th February 3pm CET. In the webinar "ESCAPE OSSR Enhancing science through sharing software - benefits & use cases" we will show the ESCAPE OSSR developments and achievements towards a FAIR multi-messenger data-driven cooperative approach.



🕒 08 March 2021 to 12 March 2021

IWAPP Workshop - Innovative Workflows in Astro and Particle Physics

Online

The objective of this workshop is to bring together the scientists' communities of Astrophysics, Astroparticle Physics and Particle Physics who are leading the development of Innovative Workflows within their domain.



● Currently (trusted ESCAPE partners):

- Checklist for ESCAPE partners at <https://escape2020.pages.in2p3.fr/wp3/ossr-pages/>
- TBD: Sign User Agreement
- Request [an issue](#) in the project [platform](#) to start the onboarding process by filling this registration [form](#), or contacting a repository maintainer.
- A short onboarding presentation should be held during an FG1 call using this [template](#), an example can be found in this [talk](#). Please book your date in this [poll](#).
- A technical report should be filed using this [template](#), an example can be found at this tech [report](#).
- Add the relevant metadata to your repository
- Register to [zenodo](#)

● Later (whole community):

- Based on Rules of Participation
- Exact workflow under discussion



● Software collection deliverable 3.7 is to recommend guidelines Licensing, provenance and metadata for ESCAPE



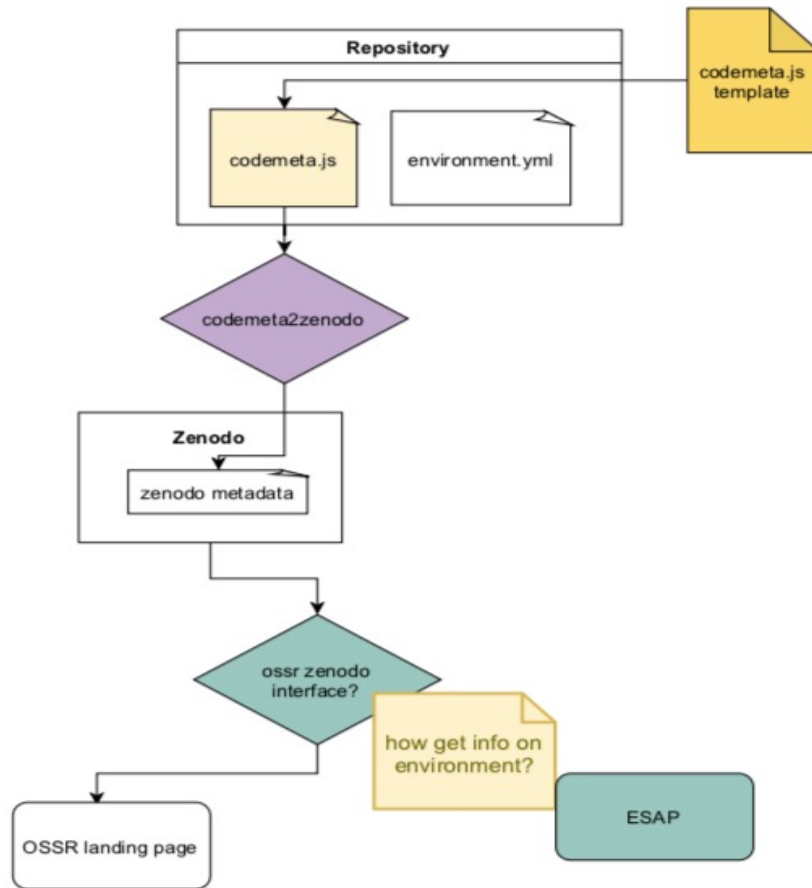
Project Title	European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructure
Project Acronym	ESCAPE
Grant Agreement No	824064
Instrument	Research and Innovation Action (RIA)
Topic	Connecting ESFRI infrastructures through Cluster projects (INFRA-EOSC-4-2018)
Start Date of Project	04.02.2019
Duration of Project	42 Months
Project Website	https://projectescape.eu/

Thomas Vuillaume, LAPP (CNRS)
Enrique Garcia, LAPP (CNRS)

[Link](#)

**D3.7 - License, provenance and metadata guidelines
for the software and service repository**





WP3+WP5 metadata task force

- Based on schema.org
- Supported by Software Heritage, ASCL and mozilla
- Zenodo support soon, but until then there is [converter](#) available
- Zenodo's metadata scheme does not contain everything one would want
- Can easily be extended
- *What else is needed?*

CodeMeta generator

Most fields are optional. Mandatory fields will be highlighted when generating Codemeta.

The software itself Name <input type="text" value="My Software"/> <small>the software title</small> Description <input type="text" value="My Software computes ephemerides and orbit propagation. It has been developed from early '80."/> Creation date <input type="text" value="YYYY-MM-DD"/> First release date <input type="text" value="YYYY-MM-DD"/> License <input type="text" value="from SPDX licence list"/>	Discoverability and citation Unique identifier <input type="text" value="10.151.xxxxx"/> <small>such as ISBNs, GTIN codes, UUIDs etc... http://schema.org/identifier</small> Application category <input type="text" value="Astronomy"/> Keywords <input type="text" value="ephemerides, orbit, astronomy"/> Funding <input type="text" value="PRA_2018_73"/> <small>grant funding software development</small> Funder <input type="text" value="Università di Pisa"/> <small>organization funding software development</small> Authors and contributors can be added below	Development community / tools Code repository <input type="text" value="git+https://github.com/You/RepoName.git"/> Continuous integration <input type="text" value="https://travis-ci.org/You/RepoName"/> Issue tracker <input type="text" value="https://github.com/You/RepoName/issues"/> Related links <input type="text"/>
Run-time environment Programming Language <input type="text" value="C#, Java, Python 3"/>	Current version of the software Version number <input type="text" value="1.0.0"/>	Additional Info Reference Publication <input type="text" value="https://doi.org/10.1000/xyz123"/>





- Zenodo is a general-purpose open-access repository developed under the European OpenAIRE program and operated by CERN. It allows researchers to deposit research papers, data sets, research software, reports, and any other research related digital artifacts
- “Zenodo is derived from [Zenodotus](#), the first librarian of the Ancient Library of Alexandria and father of the first recorded use of metadata, a landmark in library history.”

The screenshot shows the Zenodo Developers page. On the left is a navigation menu with links like Introduction, Quickstart - Upload, Testing, Versioning, Authentication, Requests, Responses, HTTP status codes, Errors, Entities, OAI-PMH, GitHub, Rate Limiting, Privacy policy, Terms of Use, and Contact. The main content area is titled 'REST API' and contains a bullet point explaining the 'new files API' and providing a bucket URL example: `https://zenodo.org/api/files/568377dd-daf8-4235-85e1-a56011ad454b`. On the right, there are two tabs: 'Python' and 'cURL'. The 'Python' tab shows a code snippet for uploading a file using the `requests` library. The 'cURL' tab shows a JSON response from the API, including fields like `key`, `mimetype`, `checksum`, `version_id`, `size`, `created`, `updated`, and `links`.





4) Data product replica prepared for compute on request, interactive session started

Definition of success	Data found in Data Lake, transferred to compute site and access given to user via ESAP
General Description	<p>SKA have delivered to WP3 OSSR a containerised workflow that takes simulated SKA data (images) and undertakes source detection and machine-learning classification. A good testing use case related to this would be to integrate this into the ESAP and prove that the workflow can be run at alternative, on-demand, interactive, compute resources.</p> <p>Ideally we'd want to test a user being given compute access at a site that does not already have the data (but that does have a rucio RSE configured), triggering a rucio-managed(?) data transfer to that site and allowing user to go via the ESAP and start their JupyterHub session.</p> <p>Separate but related use case would be to continue to build interactive prototype (JupyterHub) towards science data challenge capability, by enabling persistent environments with data recovery,</p> <p>Data Products are already stored in the ESCAPE data lake, workflows are already in the OSSR.</p>



Name	5. Analysis of a (simulated) AGN using a combined workflow (gammapy & AGNpy)
ID	CTA005
Goal/Aim	Ability to login to the ESAP and find and analyse data using a combined workflow of two (three) software packages to create an SED
Workflow	<ol style="list-style-type: none"> User logs in to the ESAP <ol style="list-style-type: none"> If public data is accessed: any user can find the data If CTA-only data is accessed: any CTA member can find the data If embargoed CTA data: only a project PI can find the data Search for Data in the <u>datalake</u> <ol style="list-style-type: none"> Search for (simulated) CTA DL3 level data Select data from search results or select all Search for corresponding IRF(instrument response function) for the data selected Search for corresponding metadata, log files etc The data can now be analysed in an interactive session: <ol style="list-style-type: none"> Search for a Jupyterhub with appropriate modules/software Search for and upload appropriate notebook Data and IRFs are available to notebook Run notebook Save data to temporary quicklook location or download final results Possible extension: Include SKA data





<https://project.escape2020.de/issues/78>

Use Case description

The scientist can find and read KM3NeT event data through ESAP and perform analyses on public data sets.

Goal

- Develop interfaces for reading of KM3NeT event data
- explore the use of the Data lake and develop KM3NeT's multisite computing environment

Data & Software

- One week ORCA4 data ([KM3NeT data center](#))
 - simple event table
 - fully reconstructed event files (root)
- Example notebooks
- km3py package

WPs: WP2 (bring data to data lake), WP3 (onboard software), WP5 (access)

Contributions: FAU, Nikhef



- Hickson Compact Groups (HCGs)
- gitlab that can be launched on desktop and mybinder
- Workflow: <https://github.com/AMIGA-IAA/hcg-16>
- Workflow:
 - First download and install runtime dependencies
 - Note: the exception is docker, which is expected to be already available on the target computer
 - Runtime dependencies are provided in a conda environment
 - Download source code for the analysis and input data.
 - Code from GitHub
 - Data from EUDAT's B2SHARE service
 - Run the pipeline
 1. Flagging and calibration with CASA (docker container)
 2. Imaging with CASA (docker container)
 3. Masking with SoFiA (docker container)
 4. Plotting with Jupyter Notebooks (conda environment)




Integration between WP5 & WP3



Current Status (Interactive Data Analysis)

- Proof of concept completed, IDA page can now generate the list of workflows via a call to the OSSR (Zenodo REST API).
- esap-api-ida-zenodo [branch](#)

Archives Interactive Analysis VO-SAMPLog In to Enable Data Selection Login

Select ESCAPE ESFRI Jupyter Workflows (Notebooks)

ESCAPE template project

Select ESCAPE JupyterHub Services

MyBinder

Deploy


sap-gui version 2 jul 2021 - 15:00



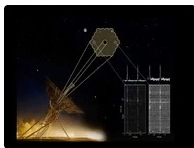
Current Status (IDA)

- REST Request adds a parameter for “jupyter-notebook”
 - `params={'communities': 'escape2020', 'access_token': ZENODO_AUTH_TOKEN, 'keywords': "jupyter-notebook"}`
- Result (currently) is a single workflow:
 - https://gitlab.in2p3.fr/escape2020/wp3/template_project_escape
- Workflow can be deployed on MyBinder




[Archives](#)
[Interactive Analysis](#)
[VO-SAMP](#)
[Log In to Enable Data Selection](#)
[Login](#)

WSRT-Apertif




Apertif Surveys

Data from the Apertif surveys include imaging and time-domain data. The time-domain products consist of high-time resolution filterbank data in the PSRFITS standard. The imaging data products include the raw observations in the measurement set (MS) standard format. In addition, processed data products are available, including calibration tables, calibrated visibilities, multi-frequency synthesis continuum images, polarization images and cubes, and uncleaned neutral hydrogen (HI) line and beam cubes. Full details of these data will be provided in upcoming papers (van Leeuwen et al. 2020, Adams et al. 2020).

[Visit WSRT-Apertif Archives](#)

ASTRON VO




ASTRON Virtual Observatory

The Virtual Observatory defines a set of standards that can be used to download astronomical data. The ASTRON VO contains several image surveys, which are images in the FITS format. Since the VO is currently under development, more data types will be available in the future.

[Visit ASTRON VO Archives](#)


LOFAR-LTA



LOFAR LTA data

The data from all LOFAR cycle, commissioning and DDT projects since 2013 are stored in the archive. The interferometric data products that can be found include raw, pre-processed data in the measurement set (MS) format, and the products from the calibration, imaging and long baseline pipelines. In the case of beamformed observations, raw data are available in HDF5 format as well as higher-level data products including de-dispersed time series, dynamic spectra and folded pulse profiles. More details on the types of data products stored on the archive are provided [here]. ([here] is wherever you put the more detailed description that was sent to you separately).

Zooniverse




Zooniverse Classification Database

The Zooniverse is the world's largest and most popular platform for people-powered research. This research is made possible by volunteers — more than a million people around the world who come together to assist professional researchers. Our goal is to enable research that would not be possible, or practical, otherwise. Zooniverse research results in new discoveries, datasets useful to the wider research community, and many publications.

[Visit Zooniverse Archives](#)

Virtual Observatory (VO)




Virtual Observatory (VO)

The Virtual Observatory defines a set of standards that can be used to download astronomical data.

[Visit Virtual Observatory \(VO\) Archives](#)

RUCIO






Rucio

Built on more than a decade of experience, Rucio serves the data needs of modern scientific experiments. Large amounts of data, countless numbers of files, heterogeneous storage systems, globally distributed data centres, monitoring and analytics. All coming together in modular solution to fit your needs.

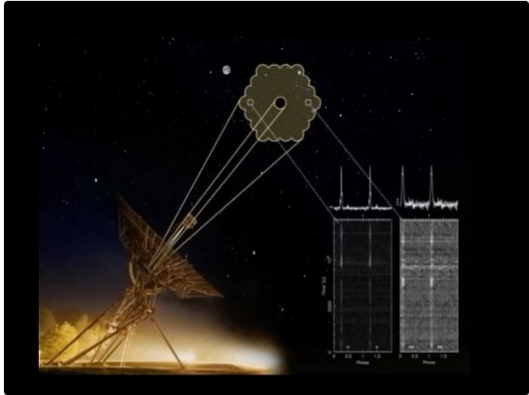
[Visit RUCIO Archives](#)




Archives Interactive Analysis VO-SAMP


 Logout Stelios Voutsinas

WSRT-Apertif




Apertif Surveys

Data from the Apertif surveys include imaging and time-domain data. The time-domain products consist of high-time resolution filterbank data in the PSRFITS standard. The imaging data products include the raw observations in the measurement set (MS) standard format. In addition, processed data products are available, including calibration tables, calibrated visibilities, multi-frequency synthesis continuum images, polarization images and cubes, and uncleaned neutral hydrogen (HI) line and beam cubes. Full details of these data will be provided in upcoming papers (van Leeuwen et al. 2020, Adams et al. 2020).

[Visit WSRT-Apertif Archives](#)

ASTRON VO

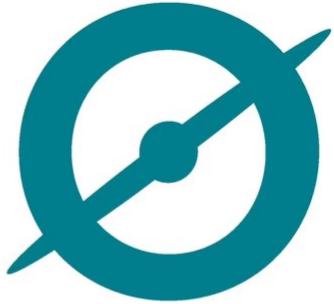


ASTRON Virtual Observatory

The Virtual Observatory defines a set of standards that can be used to download astronomical data. The ASTRON VO contains several image surveys, which are images in the FITS format. Since the VO is currently under development, more data types will be available in the future.

[Visit ASTRON VO Archives](#)

Zooniverse



Zooniverse Classification Database

The Zooniverse is the world's largest and most popular platform for people-powered research. This research is made possible by volunteers — more than a million people around the world who come together to assist professional researchers. Our goal is to enable research that would not be possible, or practical, otherwise. Zooniverse research results in new discoveries, datasets useful to the wider research community, and many publications.

[Visit Zooniverse Archives](#)

esap-gui version 23 jul 2021 - 15:00



Feedback from Implementation by WP5

- Overall a well documented easy process
- Process for adding to the OSSR seems formal (probably rightly so)
 - Is the barrier of entry low enough for the WP5 workflows to be added?
(Is the processes for onboarding a WF the same as for the software is uses?)
- Does the access_token need to be used?
 - If so what is the purpose?
 - Currently seems to work without it
 - If needed, who's access token to be used?
 - User's or ESAP account
- Extra call needed per entry (hit) in the Zenodo records to get the codemeta.json file
 - Needed for getting the Gitlab URL, which we use to deploy on MyBinder
 - Is there a way around this yes or no?
 - Not a problem for now, as we don't anticipate a large influx of "workflows"
 - If we do, does this extra call increase the latency enough to hinder user experience?



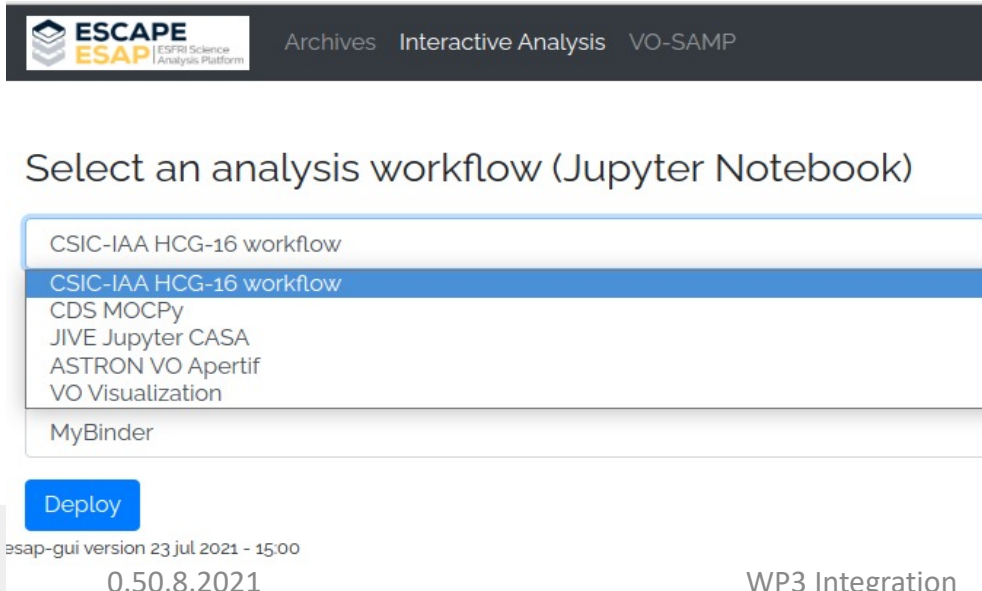
Open Questions

- Do we want a “Find Software” page within ESAP?
 - Harvest Software from the OSSR
 - Display in a separate page
 - Download / Redirect to page rather than “Deploy on facility” that IDA has
- Do we want a dedicated “Find Workflow” page, where the user can add criteria to filter the list of workflows?
- If so what would they want to filter by, and are these parameters covered in the zenodo and codemeta metadata?
- Is it worth setting up a Hackathon session where we invite the publishers of the existing workflows and members from WP3 to add them to the OSSR?



Next Steps

- Add list of existing WP5 workflows to the OSSR
 - Gather Documentation in wiki / issue
 - Invite Authors of each project to contribute their workflow via the documented procedure
- Complex / composite workflows -> Test Science Projects
- Accessing containers provided by WP3
- DAC21: Which use cases will use this?



The screenshot shows the ESCAPE ESAP Analysis Platform interface. At the top, there is a navigation bar with the ESCAPE logo and links for Archives, Interactive Analysis, and VO-SAMP. Below this, a section titled "Select an analysis workflow (Jupyter Notebook)" contains a dropdown menu with the following options: CSIC-IAA HCG-16 workflow (selected), CDS MOCPy, JIVE Jupyter CASA, ASTRON VO Apertif, VO Visualization, and MyBinder. A blue "Deploy" button is located below the dropdown. At the bottom left, the text "esap-gui version 23 jul 2021 - 15:00" and "0.50.8.2021" are displayed. At the bottom right, the text "WP3 Integration" is visible.





Backup Slides



Metadata keywords

Required

- Title
- Authors: list of authors
- Contact (maintainer):
 - name: could be a person or an entity (e.g. E. Garcia, ESCAPE or CTA Observatory)
 - email or other contact channel
- Short description of the software/service
- License

Recommended

- Provide a persistent identifier (such as a Digital Object Identifier (doi), if already available)
- Publication date (if already published elsewhere)
- URL to the (live) development repository
- Type of publication: source code, compiled binaries, container, image
- Grant/funding
- Contributors
- References
- Programming language
- External dependencies (including matching versions)
- Operating system
- Compilation environment
- Hardware requirements
 - General use case (HPC, server, local desktop)
 - CPU, RAM, HDD/SSD requirements
 - GPU requirements
 - FPGA
- Data Type (inputs/outputs)
- Access
 - Open
 - Closed
 - Restricted
 - Embargoed
- Documentation language
- Keywords: to be selected in a defined list of keywords, see below

