ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

ESCAPE Extended Discussion Day

WP3 + WP5 Discussion

28 September 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.

Goals for this session

- Review current status of ESAP/OSSR integration
- Solicit feedback from the wider (technical) ESCAPE community
- Develop roadmap for continued integration

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

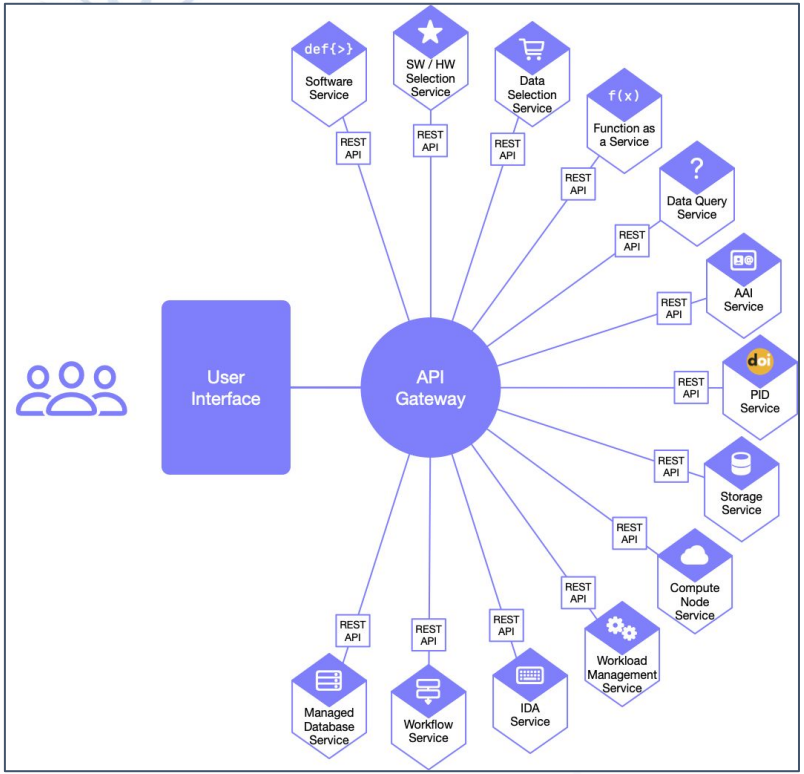
Logistics

- This is a **discussion** session; we want to hear from you!
- As far as possible, we encourage you to engage *on Zoom* — let's discuss your needs and ideas, rather than using it as just a Q&A.
- However, we will keep an eye on `#1-general-discussion` on Slack, so feel free to add thoughts there.

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

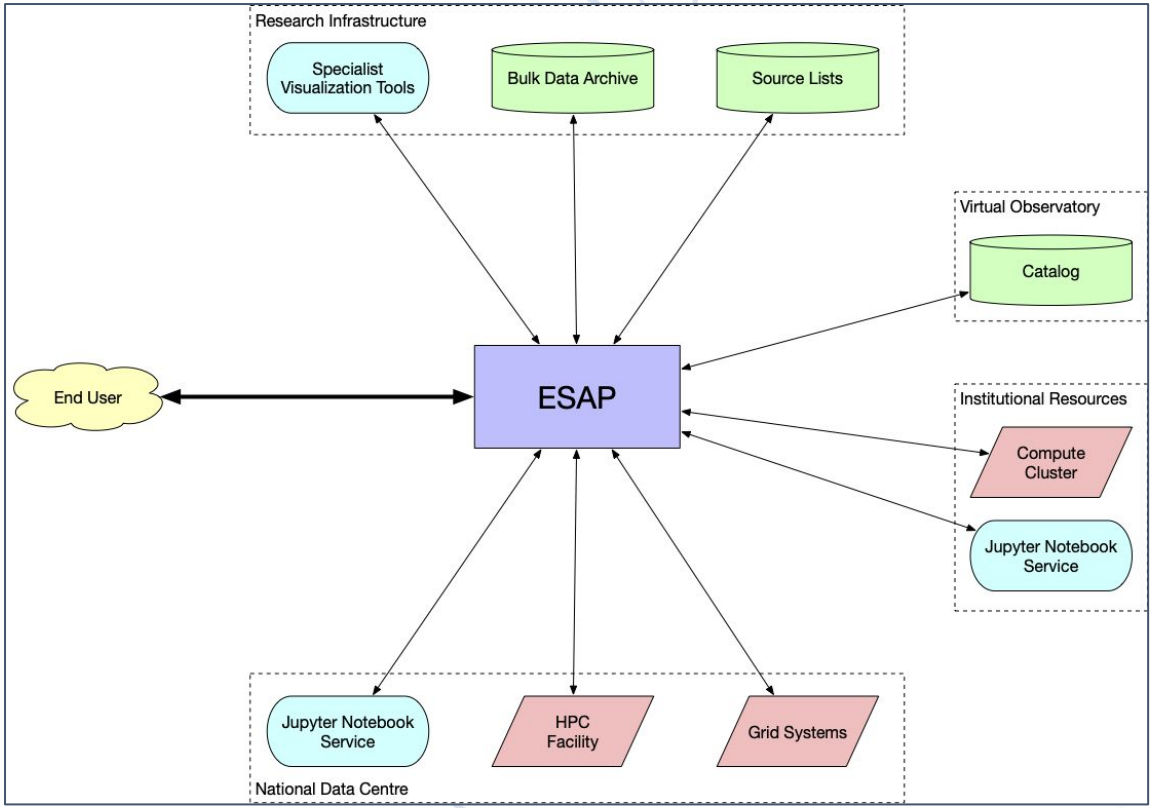
What is this ESAP thing anyway?

- A modular “science platform toolkit”: you bring the services, ESAP provides the “glue” to put them together.
- WP5 provides the (adaptable) user interface, the API gateway, and a number of service integrations… you can add more.
- WP5 does not provide a single, centralized ESAP deployment for “production” use (although various development servers are available).
- Expectation is that multiple ESAP instances can be set up to address specific RI, project, etc needs.



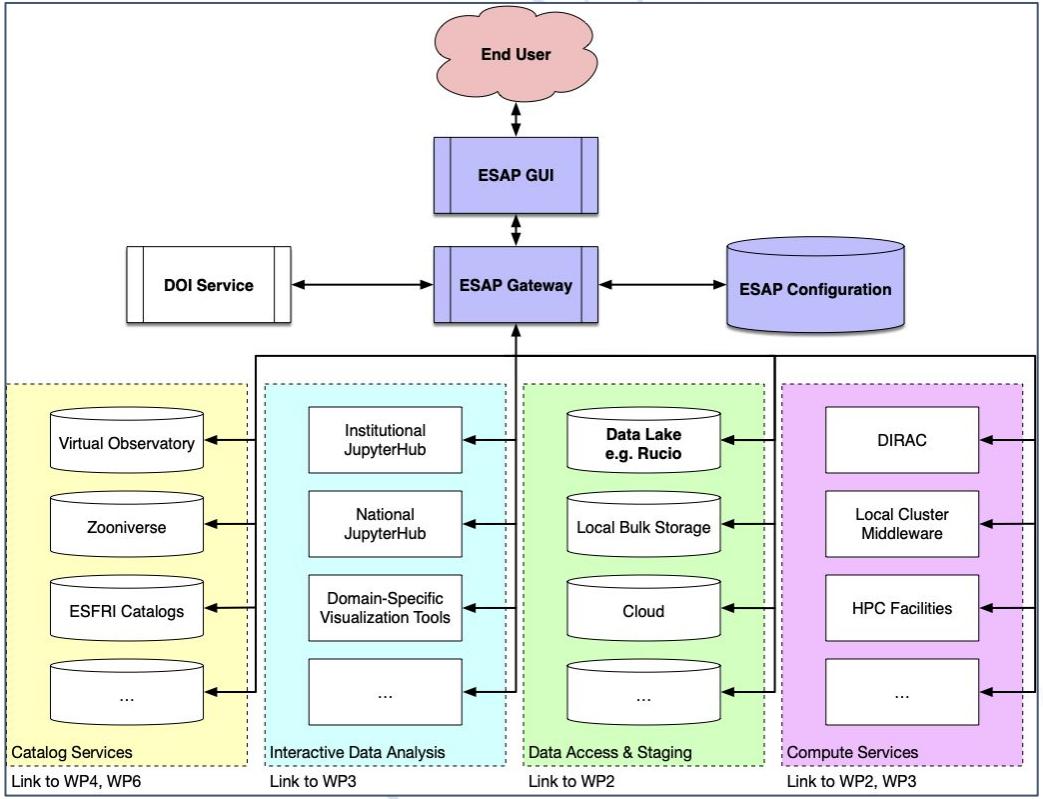
Funded by the European Union’s Horizon 2020 - Grant N° 824064

ESAP in its environment



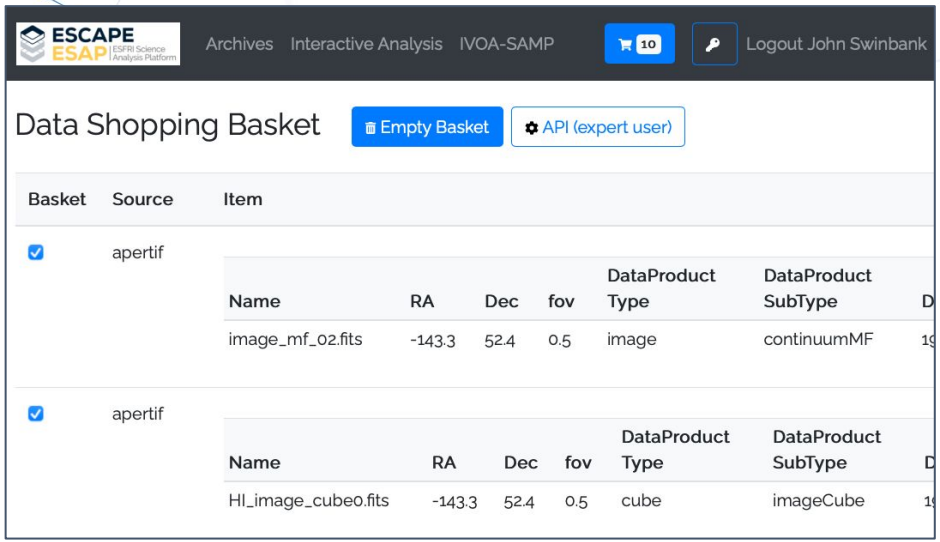
Funded by the European Union's Horizon 2020 - Grant N° 824064

ESAP in its environment



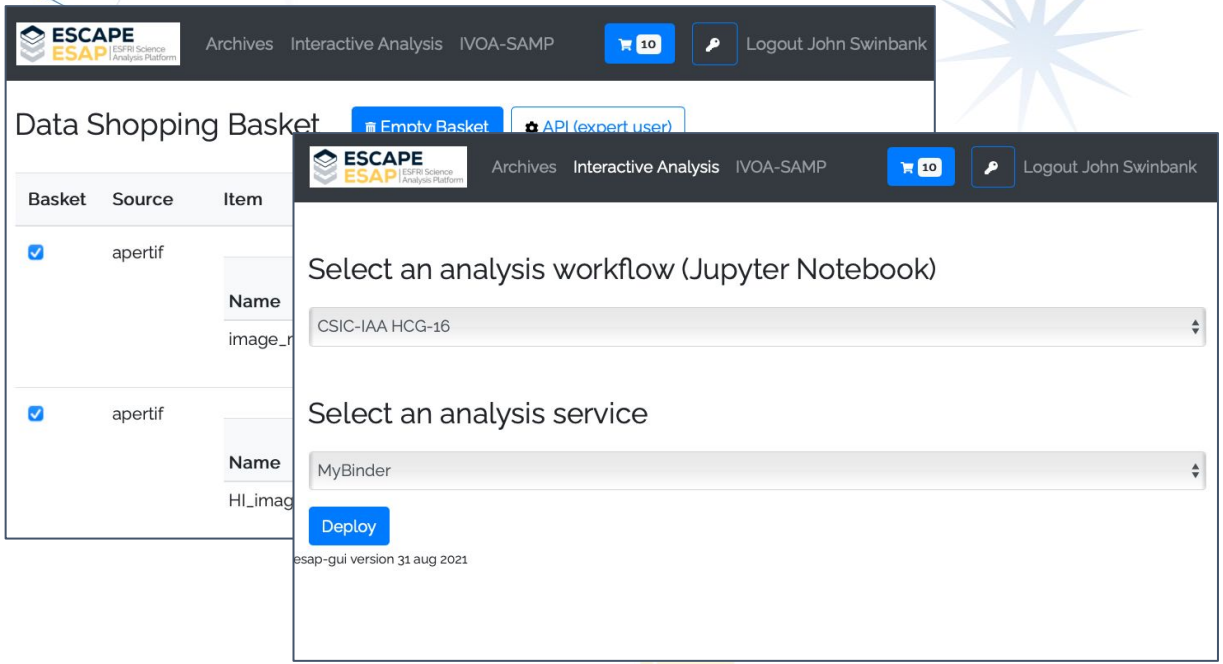
Funded by the European Union's Horizon 2020 - Grant N° 824064

ESAP + OSSR



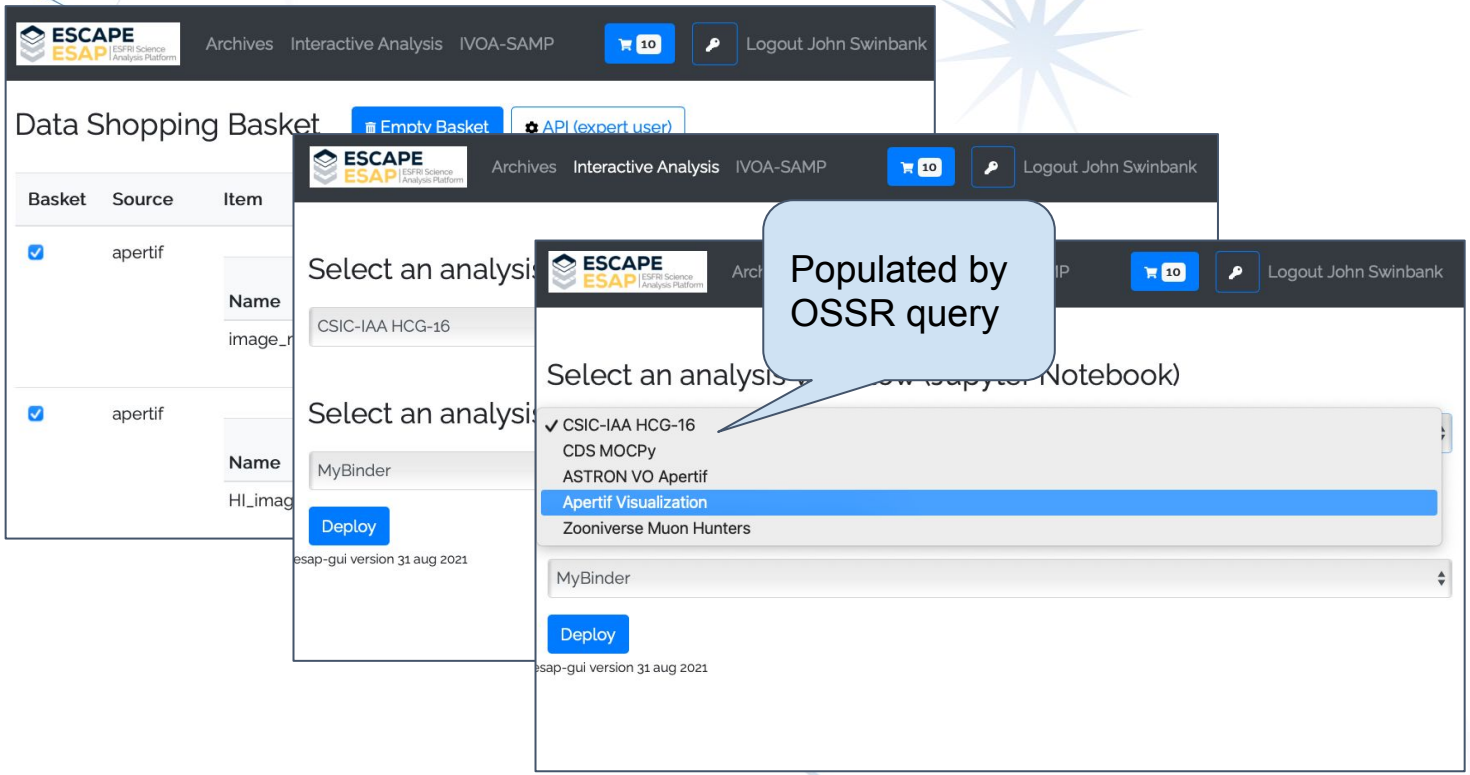
Funded by the European Union's Horizon 2020 - Grant N° 824064

ESAP + OSSR



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

ESAP + OSSR



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

Breaking news!

- ESAP is now using the eossr library to fetch data from Zenodo/OSSR.
- Outstanding merge request: https://git.astron.nl/astron-sdc/esap-api-gateway/-/merge_requests/128.
  - Huge 👋 for Stelios & Gareth!
- Not yet deployed on sdc-dev.astron.nl/esap-gui test system.

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

Discussion Points (1)

- *How do users want to find software through ESAP?*
  - This directly impacts on the metadata which needs to be stored in OSSR and harvested through ESAP.
- For example…
  - Search by keyword.
  - Search by type of software (Jupyter notebook, command line script, …).
  - Identify software which can explicitly process the types of data already in the shopping basket.
    - Or just by keyword?
  - Identify software which can be run given the available services.
  - …?
- Should ESAP warn users if they try to execute a workflow which isn't relevant to their data?

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

Discussion Points (2)

- *What types of software will users run?*
  - The basic model is a Jupyter notebook; select it, and that notebook appears in the browser, with access to the data in the shopping basket.
- But…
  - What if there are two notebooks in the repository?
  - What if the software is a command line tool?
  - What if the software is a library?
  - What if the software is some opaque container?
- What should the user see after they click the “deploy” button in the most general case?
- Can all software be meaningfully containerized?
- Should all software define an “entrypoint” which specifies what happens when ESAP runs that software?
- Can that entrypoint also specify what sorts of inputs (and outputs?) the software handles / produces? MIME types? 🤔
  - ...and what about software that doesn’t take input? Should that still be launchable through ESAP?

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

Discussion Points (3)

- *What are "complex workflows", and how do we enable them?*
- Proto-definition, based on previous discussion:
  - **A "complex workflow" involves combining data from multiple different facilities in one analysis session.**
  - (Do we agree on that definition? Perhaps combining tools, rather than facilities?)
  - For example: *agnpy and gLike are both registered with the OSSR. The user wants to analyze some data using both of them in the same session.*
- Do we need to support composing arbitrary environments from the OSSR contents?
- Do we rely on users to register potentially many different combinations as separate OSSR entries (e.g. somebody registers gammapy+agnpy, somebody else some other combination of tools)

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

Deliverables and Next Steps

Proposed deliverables:

1. A taxonomy of different types of software available through the OSSR.
2. Derived from the above: a list of types of software that will be supported by ESAP.
   - It is not necessary that ESAP can run everything in the OSSR; we should make clear what is supported, and what constraints that puts on the software (e.g. in terms of packaging, metadata, etc).
3. A list of metadata required by ESAP.
   - We've talked about this for a long while, but let's make it a deliverable to stop talking and write it down! (where?)
4. Others?

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

Backup Material

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

- What will be supported by the ESAP?
  - Source code? docker / singularity / containers / images? Packages?
  - Install?
  - From Docker-hub? Registry?

- Metadata
  - What kind of information does the ESAP need?
    - Contains notebooks
    - Containers/Images:
    - Has one associated? --> if yes, URL
      - Is one? / Includes one in the record?
      - Docker/Singularity version?
    - Is workflow?
  - How to provide them in the most exhaustive but precise way?

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