

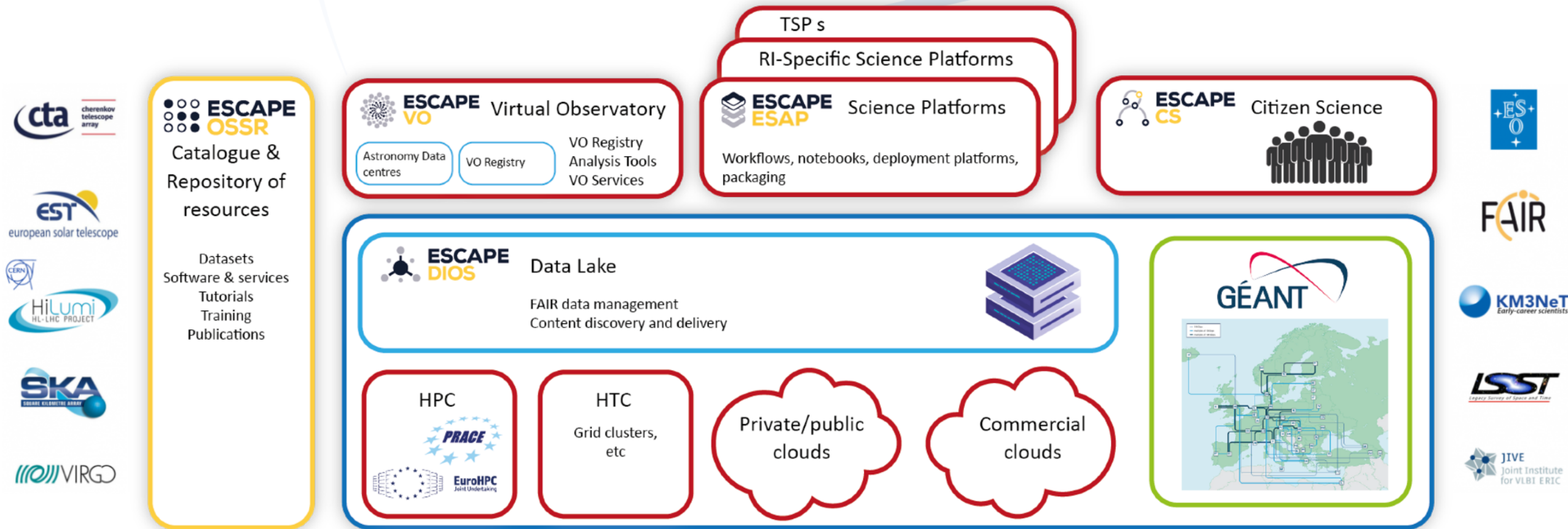


ESAP: ESFRI Science Analysis Platform (WP5)

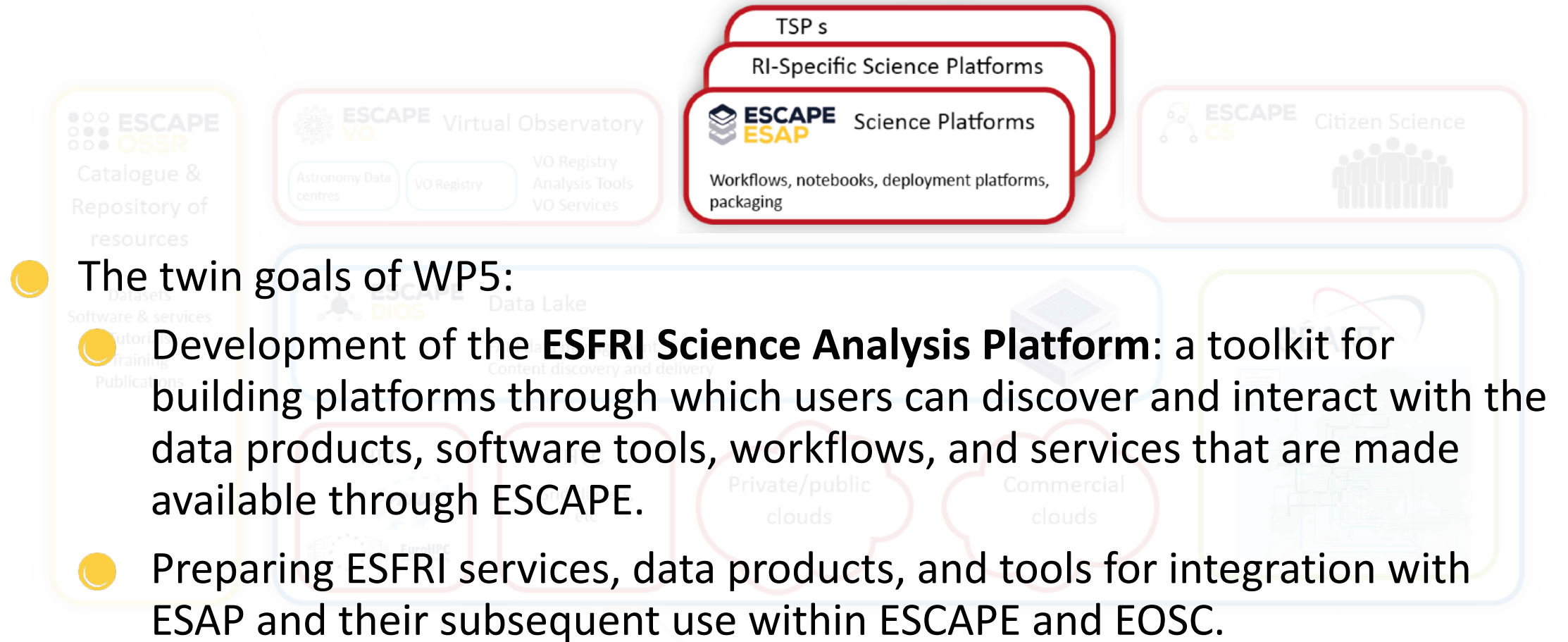
John Swinbank — swinbank@astron.nl



The ESCAPE EOSC Cell

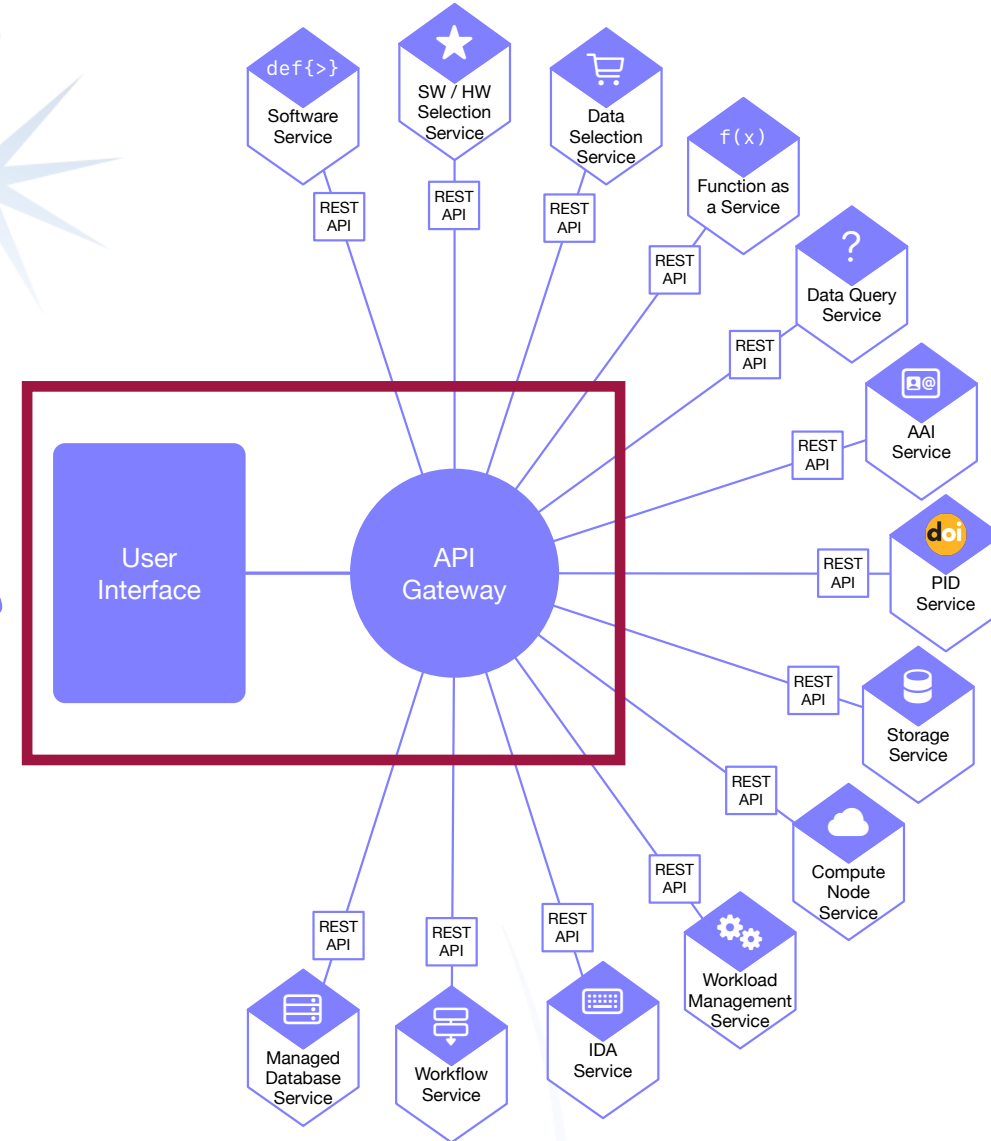


WP5 / ESAP in the ESCAPE “big picture”

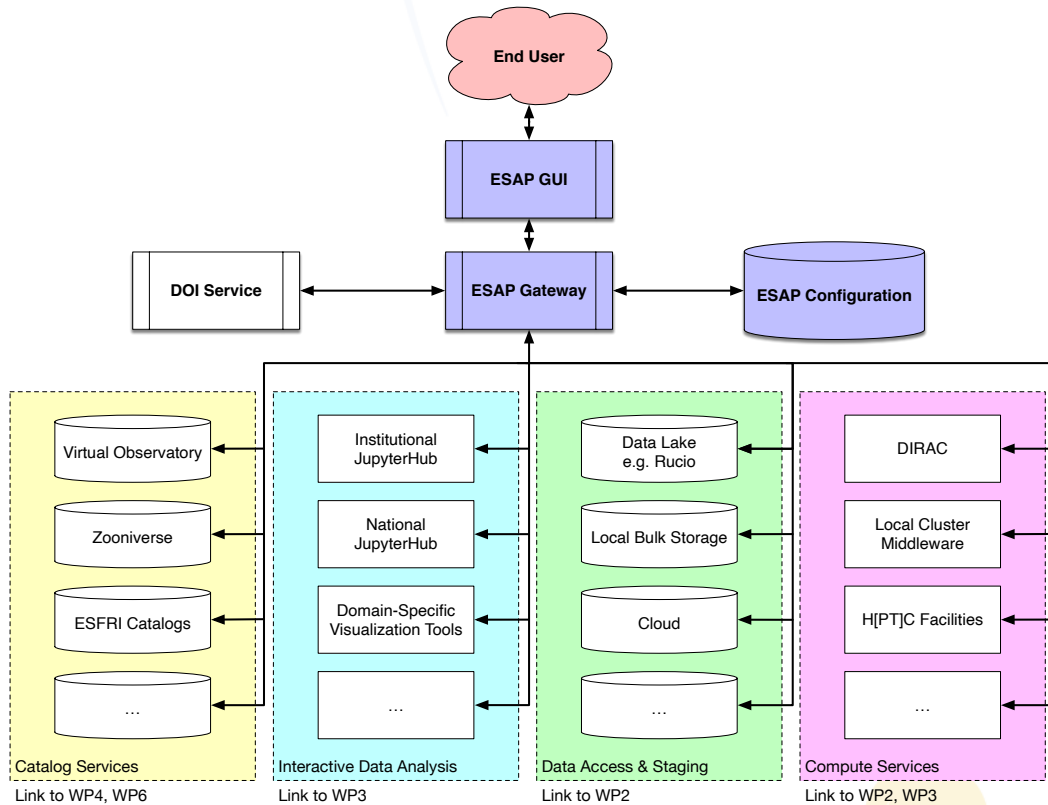


ESAP: Core & Services

- ESAP provides a focal point for integrating diverse services from across the ESCAPE project and from individual ESFRIs.
- The heart of the system is the web-based *User Interface* and the *API Gateway*. These software products are *core WP5 deliverables*.
- The API Gateway modulates communication with a range of external services using REST APIs. A “service connector” system provides a standardized, plugin-based mechanism for integrating new services.
- WP5 members and ESFRI partners collaborate on developing & integrating a variety of external services.
- It is possible to extend ESAP to operate in a variety of different environments by implementing & integrating new service connectors.



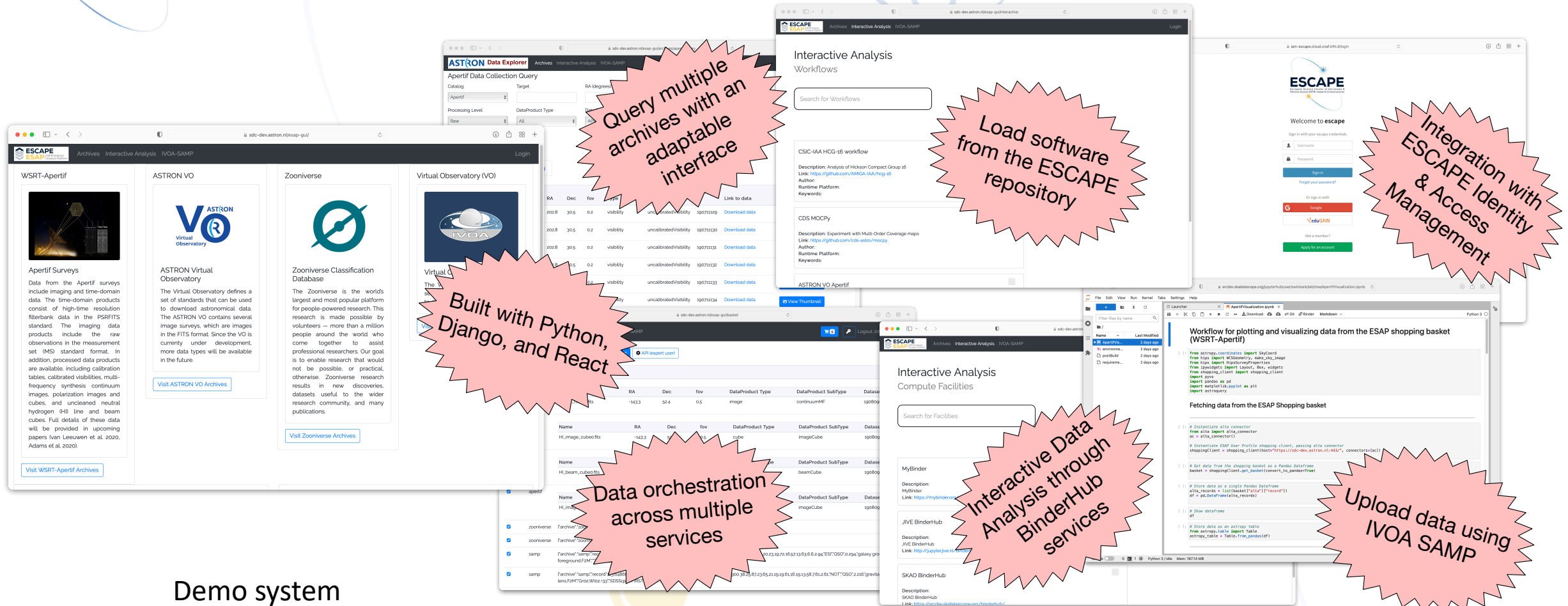
ESAP as a *toolkit*; integration with ESFRIs, EOSC



- WP5 is not resourced to operate or provide science platform services, except for development and test systems.
- The ESAP deliverables are oriented around enabling ESFRIs, other projects, and institutions to *deploy their own science platforms*.
- Individual ESAP instances can be registered with the EOSC Portal.



Current Status & Capabilities



Query multiple archives with an adaptable interface

Load software from the ESCAPE repository

Integration with ESCAPE Identity & Access Management

Built with Python, Django, and React

Data orchestration across multiple services

Interactive Data Analysis through BinderHub services

Upload data using IVOA SAMP

The demo system interface includes sections for:

- WSRT-Apertif**: 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).
- ASTRON VO**: 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.
- Zooniverse**: 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.
- Virtual Observatory (VO)**: The Virtual Observatory (VO) is 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.

Demo system

<https://sdc-dev.astron.nl/esap-gui>



WP5 Structure

● Task 5.1: Data Aggregation & Staging

- *Provide ESAP users with the capability to access and combine data from multiple collections and to stage that data for subsequent analysis.*
- *Links to **WP2/DIOS**, **WP4/VO**.*

● Task 5.2: Software Deployment & Virtualization

- *Make the codes, tools, scripts, and other packages developed in support of the various ESFRIs and other RIs readily available to ESAP users.*
- *Links to **WP3/OSSR**.*

● Task 5.3: Analysis interface, work flows, and reproducibility

- *Integrate the data access services provided in Task 5.1 with the software from Task 5.2 to provide users with a coordinated approach to data analysis in the ESAP context.*
- *Links to **WP2/DIOS**, **WP3/OSSR**, **WP4/VO**.*

● Task 5.4: Integration with HPC and HTC infrastructures

- *Make it possible for ESAP users to deploy their workflows and analysis jobs at scale on a range of HPC and HTC infrastructures.*
- *Links to **WP2/DIOS**, **WP3/OSSR**.*

● Task 5.5: Work Package Management



RP2 Technical Highlights: Data Orchestration

● The ESAP “Shopping Basket”, deployed during RP2, is **the core data orchestration metaphor** within ESAP.

● Results of queries and analysis tasks are place in the Basket and are carried through the system.

● Future analysis jobs or other tasks manipulate and augment the basket contents.

sdsc-dev.astron.nl/esap-gui/basket

ESCAPE

ESAP

ESAP Science Analysis Platform

Archives

Interactive Analysis

IVOA-SAMP

Logout John Swinbank

Data Checkout

Empty Basket

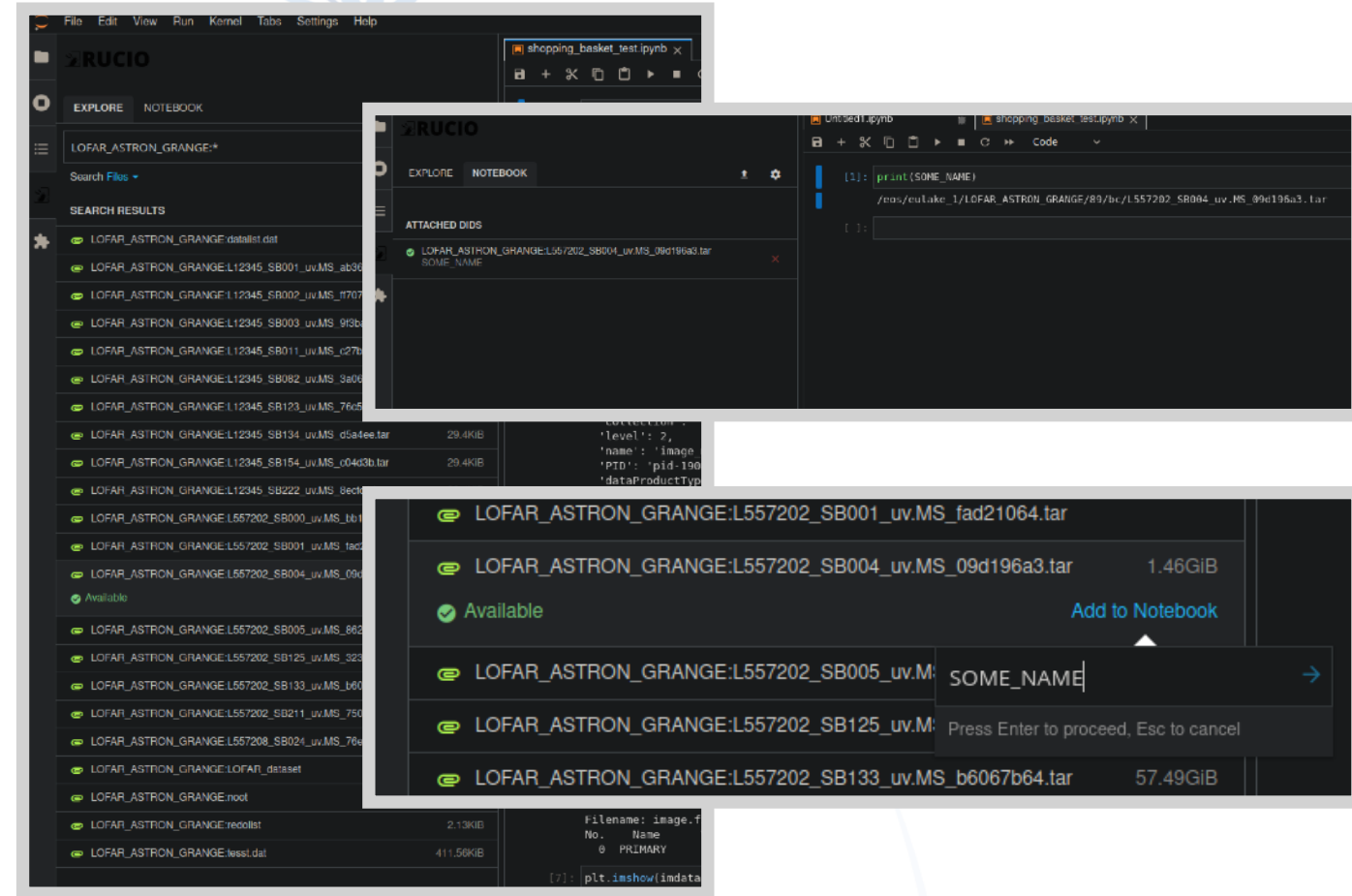
API (expert user)

| Basket | Source | Item | | | | | | | | | | | | | | |
|-------------------------------------|------------|--|------|------------------|---------------------|------------|------------------|---------------------|------------|---------------------|--------|------|-----|-------|-------------|-----------|
| <input checked="" type="checkbox"/> | apertif | <table><tr><th>Name</th><th>RA</th><th>Dec</th><th>fov</th><th>DataProduct Type</th><th>DataProduct SubType</th><th>Dataset ID</th></tr><tr><td>image_mf_02.fits</td><td>-143.3</td><td>52.4</td><td>0.5</td><td>image</td><td>continuumMF</td><td>190809041</td></tr></table> | Name | RA | Dec | fov | DataProduct Type | DataProduct SubType | Dataset ID | image_mf_02.fits | -143.3 | 52.4 | 0.5 | image | continuumMF | 190809041 |
| Name | RA | Dec | fov | DataProduct Type | DataProduct SubType | Dataset ID | | | | | | | | | | |
| image_mf_02.fits | -143.3 | 52.4 | 0.5 | image | continuumMF | 190809041 | | | | | | | | | | |
| <input checked="" type="checkbox"/> | apertif | <table><tr><th>Name</th><th>RA</th><th>Dec</th><th>fov</th><th>DataProduct Type</th><th>DataProduct SubType</th><th>Dataset ID</th></tr><tr><td>HI_image_cube0.fits</td><td>-143.3</td><td>52.4</td><td>0.5</td><td>cube</td><td>imageCube</td><td>190809041</td></tr></table> | Name | RA | Dec | fov | DataProduct Type | DataProduct SubType | Dataset ID | HI_image_cube0.fits | -143.3 | 52.4 | 0.5 | cube | imageCube | 190809041 |
| Name | RA | Dec | fov | DataProduct Type | DataProduct SubType | Dataset ID | | | | | | | | | | |
| HI_image_cube0.fits | -143.3 | 52.4 | 0.5 | cube | imageCube | 190809041 | | | | | | | | | | |
| <input checked="" type="checkbox"/> | apertif | <table><tr><th>Name</th><th>RA</th><th>Dec</th><th>fov</th><th>DataProduct Type</th><th>DataProduct SubType</th><th>Dataset ID</th></tr><tr><td>HI_beam_cube0.fits</td><td>-143.3</td><td>52.4</td><td>0.5</td><td>cube</td><td>beamCube</td><td>190809041</td></tr></table> | Name | RA | Dec | fov | DataProduct Type | DataProduct SubType | Dataset ID | HI_beam_cube0.fits | -143.3 | 52.4 | 0.5 | cube | beamCube | 190809041 |
| Name | RA | Dec | fov | DataProduct Type | DataProduct SubType | Dataset ID | | | | | | | | | | |
| HI_beam_cube0.fits | -143.3 | 52.4 | 0.5 | cube | beamCube | 190809041 | | | | | | | | | | |
| <input checked="" type="checkbox"/> | apertif | <table><tr><th>Name</th><th>RA</th><th>Dec</th><th>fov</th><th>DataProduct Type</th><th>DataProduct SubType</th><th>Dataset ID</th></tr><tr><td>HI_image_cube1.fits</td><td>-143.3</td><td>52.4</td><td>0.5</td><td>cube</td><td>imageCube</td><td>190809041</td></tr></table> | Name | RA | Dec | fov | DataProduct Type | DataProduct SubType | Dataset ID | HI_image_cube1.fits | -143.3 | 52.4 | 0.5 | cube | imageCube | 190809041 |
| Name | RA | Dec | fov | DataProduct Type | DataProduct SubType | Dataset ID | | | | | | | | | | |
| HI_image_cube1.fits | -143.3 | 52.4 | 0.5 | cube | imageCube | 190809041 | | | | | | | | | | |
| <input checked="" type="checkbox"/> | apertif | <table><tr><th>Name</th><th>RA</th><th>Dec</th><th>fov</th><th>DataProduct Type</th><th>DataProduct SubType</th><th>Dataset ID</th></tr><tr><td>HI_beam_cube1.fits</td><td>-143.3</td><td>52.4</td><td>0.5</td><td>cube</td><td>beamCube</td><td>190809041</td></tr></table> | Name | RA | Dec | fov | DataProduct Type | DataProduct SubType | Dataset ID | HI_beam_cube1.fits | -143.3 | 52.4 | 0.5 | cube | beamCube | 190809041 |
| Name | RA | Dec | fov | DataProduct Type | DataProduct SubType | Dataset ID | | | | | | | | | | |
| HI_beam_cube1.fits | -143.3 | 52.4 | 0.5 | cube | beamCube | 190809041 | | | | | | | | | | |
| <input checked="" type="checkbox"/> | zooniverse | [{"archive": "zooniverse", "catalog": "project", "project_id": "15072", "category": "classifications"}] | | | | | | | | | | | | | | |
| <input checked="" type="checkbox"/> | zooniverse | [{"archive": "zooniverse", "catalog": "project", "project_id": "15072", "category": "subjects"}] | | | | | | | | | | | | | | |



RP2 Technical Highlights: Data Lake as a Service

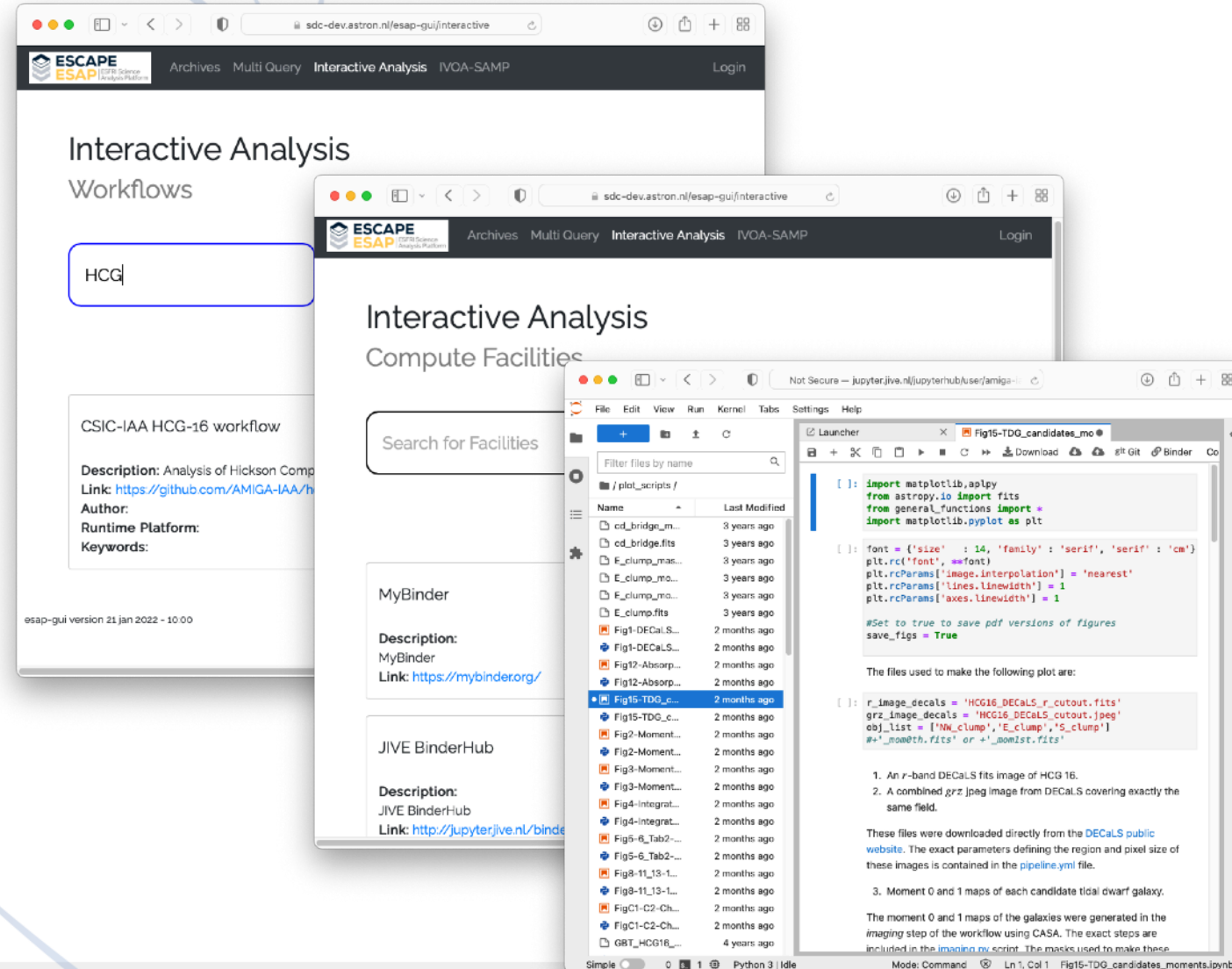
- The key integration with the WP2/DIOS Data Lake.
- ESAP users can access and manipulate Data Lake contents from within ESAP supported analysis environments.
- Includes a data browser, file upload/download, local storage, latency-hiding content-delivery.



RP2 Technical Highlights: OSSR Integration

- Dispatch software discovered in the WP3/ OSSR repository to compute services accessible through ESAP.
- ESAP makes it possible to quickly leverage the variety of general scientific and ESFRI-specific software available through the OSSR.
- Increasing range of searching, sorting, filtering capabilities.
- Work in progress: smart routing of jobs to appropriate compute resources.

ASTRON

The image displays three overlapping screenshots of the ESAP Interactive Analysis interface and a Jupyter notebook.

The top-left screenshot shows the "Interactive Analysis Workflows" page. It features a search bar with "HCG" entered. Below the search bar, there is a section for "CSIC-IAA HCG-16 workflow" with a description, link, author, and runtime platform. The bottom of this screenshot shows the "esap-gui version 21 jan 2022 - 10:00".

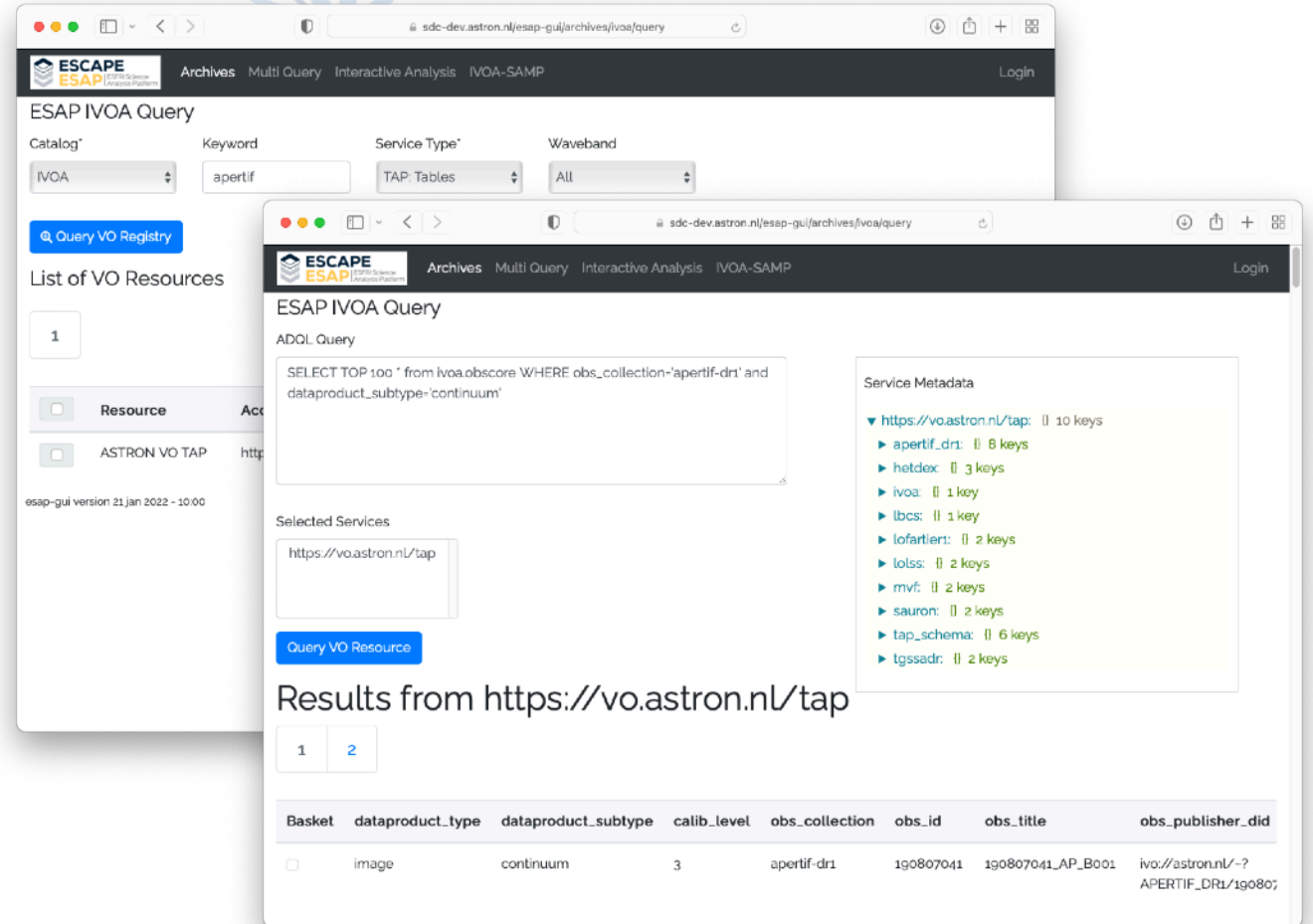
The middle-right screenshot shows the "Interactive Analysis Compute Facilities" page. It has a "Search for Facilities" bar. Below this, there are sections for "MyBinder" and "JIVE BinderHub", each with a description and a link.

The bottom-right screenshot shows a Jupyter notebook titled "Fig15-TDQ_candidates_mo". The notebook contains Python code for plotting astronomical data using Matplotlib and Astropy. The code includes imports, parameter settings, and a plot command. Below the code, there is a section titled "The files used to make the following plot are:" followed by a list of files and a description of the plot.



RP2 Technical Highlights: Virtual Observatory

- Close integration between ESAP and WP4/VO make a wide range of science cases possible.
- Extensive support for querying a variety of VO services, within the ESAP system itself and from analysis environments
- SAMP support provides seamless integration with other VO services and desktop applications.
- Collaboration with **WP4** to develop a technical note exploring concepts for a web-service interface providing a simple way to discover and access computing services. This is an important step towards *common, interoperable standards* emerging from the ESCAPE project.



The screenshot displays the ESAP IVOA Query interface. The top navigation bar includes links for Archives, Multi Query, Interactive Analysis, and IVOA-SAMP, along with a Login button. The main section is titled 'ESAP IVOA Query' and features a search bar with the keyword 'apertif' and a 'Query VO Registry' button. Below this, a 'List of VO Resources' shows a single resource: 'ASTRON VO TAP'. The 'ADQL Query' section contains the following query: `SELECT TOP 100 * from ivoa.obscore WHERE obs_collection='apertif-drl' and dataproduct_subtype='continuum'`. The 'Selected Services' section lists the URL 'https://vo.astron.nl/tap'. The 'Results from https://vo.astron.nl/tap' section shows a table with the following data:

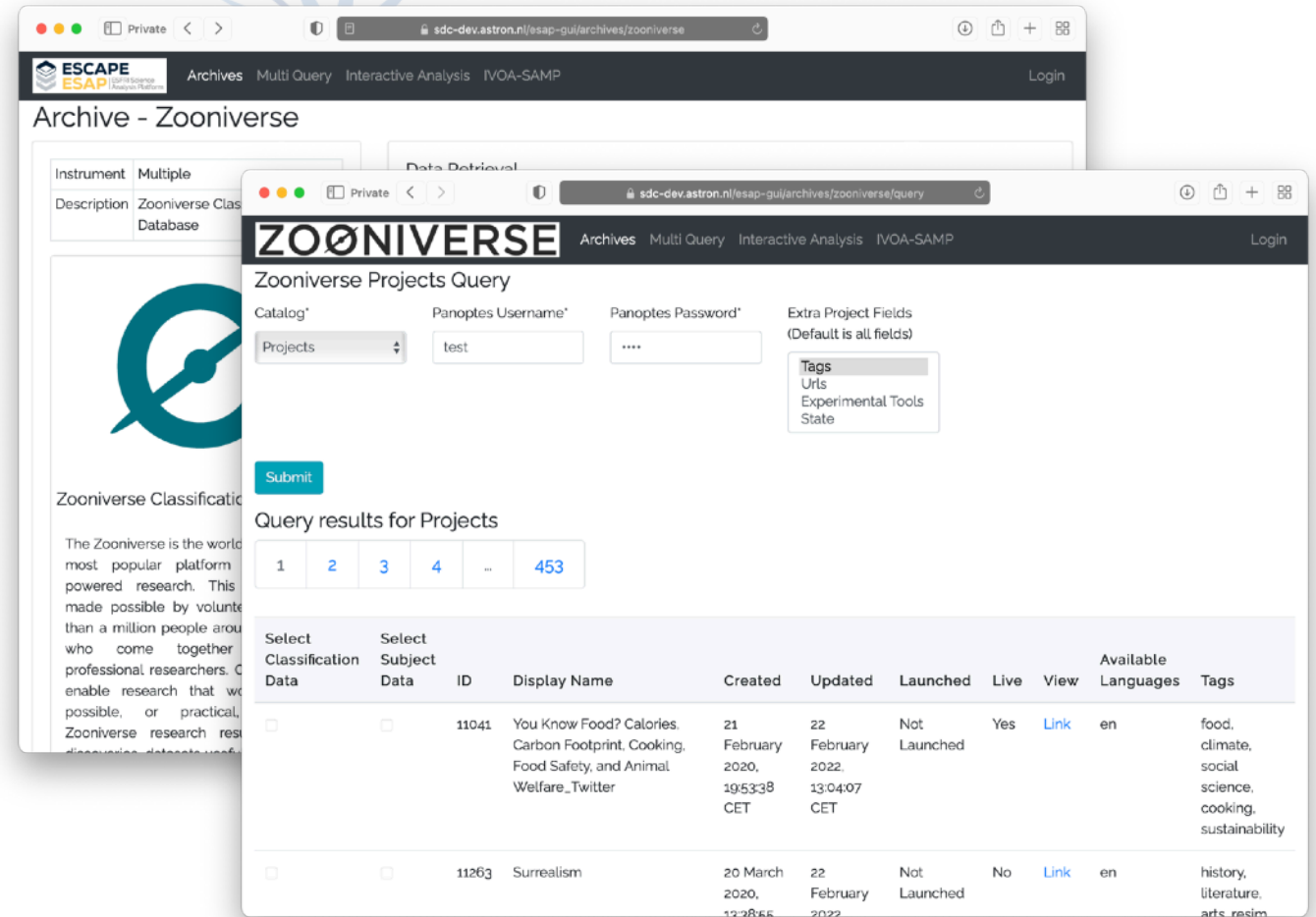
| Basket | dataproduct_type | dataproduct_subtype | calib_level | obs_collection | obs_id | obs_title | obs_publisher_did |
|--------------------------|------------------|---------------------|-------------|----------------|-----------|-------------------|--------------------------------------|
| <input type="checkbox"/> | image | continuum | 3 | apertif-drl | 190807041 | 190807041_AP_B001 | ivo://astron.nl/~?APERTIF_DRL/190807 |

On the right side, the 'Service Metadata' section lists various services and their key counts: `https://vo.astron.nl/tap` (10 keys), `apertif_drl` (8 keys), `hetdex` (3 keys), `ivoa` (1 key), `lbc` (1 key), `lofarlert` (2 keys), `tolss` (2 keys), `mvf` (2 keys), `saaron` (2 keys), `tap_schema` (6 keys), and `tgssadr` (2 keys).



RP2 Technical Highlights: Zooniverse

- **Full integration of ESAP with the WP6/Zooniverse “Panoptes” system.**
- ESAP users can directly query into the Panoptes database and add data to their basket.
- Notebook users can use Python APIs to access the shopping basket, combine with astronomical data, etc.




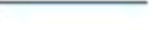






The screenshot shows the Zooniverse Projects Query interface. The top navigation bar includes 'Archives', 'Multi Query', 'Interactive Analysis', and 'IVOA-SAMP'. The main heading is 'Archive - Zooniverse'. Below this, there's a 'Data Retrieval' section with a 'Zooniverse Classification' description. The 'Zooniverse Projects Query' form includes fields for 'Catalog*' (set to 'Projects'), 'Panoptes Username*' (set to 'test'), and 'Panoptes Password*' (masked with '....'). There's also an 'Extra Project Fields' dropdown menu with options: 'Tags', 'Urls', 'Experimental Tools', and 'State'. A 'Submit' button is present. Below the form, the 'Query results for Projects' are displayed as a table with 453 results. The table has columns for 'Select Classification Data', 'Select Subject Data', 'ID', 'Display Name', 'Created', 'Updated', 'Launched', 'Live', 'View', 'Available Languages', and 'Tags'.

| Select Classification Data | Select Subject Data | ID | Display Name | Created | Updated | Launched | Live | View | Available Languages | Tags |
|----------------------------|--------------------------|-------|---|--------------------------------|--------------------------------|--------------|------|----------------------|---------------------|--|
| <input type="checkbox"/> | <input type="checkbox"/> | 11041 | You Know Food? Calories, Carbon Footprint, Cooking, Food Safety, and Animal Welfare_Twitter | 21 February 2020, 19:53:38 CET | 22 February 2022, 13:04:07 CET | Not Launched | Yes | Link | en | food, climate, social science, cooking, sustainability |
| <input type="checkbox"/> | <input type="checkbox"/> | 11263 | Surrealism | 20 March 2020, 12:28:55 CET | 22 February 2022 | Not Launched | No | Link | en | history, literature, arts, resin |



Research Infrastructure Engagement

- Wide representation of RIs in WP5.
- ESAP is *customizable for different RIs*.
- The partners contribute to the development of a common core — the UI & API gateway — and collaborate to develop plugins to address their particular needs.

| WP: | WP5 | |
|-----------------|---|-----------------|
| RI: | | |
| CTA |  | ESFRI PROJECTS |
| EST |  | |
| KM3NET |  | |
| ELT and ESO |  | ESFRI LANDMARKS |
| FAIR |  | |
| HL-LHC and CERN |  | |
| SKA |  | |
| JIVE |  | ERIC |
| EGO |  | Others |
| LSST-Europe |  | |



Research Infrastructure Engagement Results and Ongoing Activities



- Data and analysis notebooks published and made available through ESAP.
- Initial batch processing capabilities based on DIRAC integrated with ESAP in test form; full ESAP batch interface will be forthcoming.
- Development of CONCORDIA, a containerized system for execution the CORSIKA air-shower simulation system.
- Integrating ESAP for use in the CTAO Second Science Data Challenge.
- EGO/VIRGO pipelines packaged and made available; integration with the WP2/DIOS Data Lake.
- Managed database service in support of EGO/VIRGO processing currently in early prototype.
- Detailed use cases describing potential evolution of ESAP into ESO role.
- Batch processing on miniCBM data, including integration with the WP2/DIOS Data Lake.
- Batch processing of PANDA experiment data.
- Dockerized FAIR applications (R3BRoot, CMBRoot).
- Data Lake as a Service provides integration between analysis environments and Data Lake storage.



Research Infrastructure Engagement Results and Ongoing Activities



- Released a Jupyter plugin providing seamless access to the European VLBI Network archive.
- BinderHub/Jupyter system at JIV-ERIC integrated with ESAP, provides specialist radio astronomy Jupyter kernel (with CASA 6.4).
- Developing the capability to archive and provide persistent identifiers to notebook analysis runs.
- LOFAR DDFacet pipeline packaged and containerized for deployment on ESCAPE infrastructures.
- Rubin Science Platform deployment at CNRS-LAPP; investigating integrations with ESAP.
- Interfacing persistent, network-attached storage (NFS, Ceph) to interactive data analysis environments.
- Making use of an ESAP instances in the STFC cloud and at the Spanish SKA Regional Centre.
- ESAP identified as a technology of interest for SKA Regional Centre prototyping during 2022/23.
- Solution package for SKA Data Challenge 1 published to WP3 Software Repository and made available through ESAP.



Research Infrastructure Engagement

Summer 2021 workshop, MS31/D5.3

- The *Second WP5 Workshop* (2021-08-05) presented the current status of ESAP development & solicited input from the community.
- <http://indico.in2p3.fr/e/SecondWP5Workshop> for agenda, materials, etc.
- D5.3 summarizes the results of that workshop, and compares the current status of ESAP to:
 - *Requirements*, captured in D5.2 (Detailed Project Plan);
 - *Use cases*, captured in the [Project Platform](#).
- D5.3 made a number of suggestions for minor “course corrections”, but concludes that the overall direction of travel and functionality is appropriate.



Deliverables

| Number | Title | Due | Completed |
|--------|---|-----------------|----------------|
| D5.1 | Preliminary report on D5.1 requirements for ESFRI WP5 science analysis use cases | M6 (July 2019) | M6 (July 2019) |
| D5.2 | Detailed project plan for WP5 | M9 (Oct. 2019) | M9 (Oct. 2019) |
| D5.3 | Performance assessment of initial Science Platform prototype | M31 (Aug. 2021) | Submitted M31 |
| D5.4 | Final assessment of the performance of the Science Platform prototype and plan for deployment of production version within the EOSC | M42 (July 2022) | - |



Milestones

| Number | Title | Due | Completed |
|--------|---|-----------------|-----------------|
| MS27 | First WP5 workshop on Science Platform design and requirements | M4 (May 2019) | M4 (May 2019) |
| MS28 | Review of preliminary report on requirements for ESFRI science analysis use cases | M5 (June 2019) | M5 (June 2019) |
| MS29 | Initial science platform prototype with discovery and data staging | M22 (Nov 2020) | M22 (Nov 2020) |
| MS30 | Deployment of initial set of ESFRI software on prototype platform | M23 (Dec. 2020) | M24 (Jan. 2021) |
| MS31 | Second WP5 workshop to analyse prototype performance | M31 (Aug. 2021) | Submitted M31 |
| MS32 | Integration of Science Platform with OSSR repository | M32 (Sep. 2021) | Submitted M32 |
| MS33 | Integration of Science Platform with Data Lake expanded prototype | M34 (Nov. 2021) | Submitted M34 |
| MS34 | Delivery and integration of new ESFRI visualization and analysis tools | M39 (Apr. 2022) | - |
| MS35 | Final WP5 ESFRI user training workshop on the Science Platform | M40 (May 2022) | - |



Risks Analysis: Grant Agreement

| Risk | Description | Current Status |
|------|--|--|
| 17 | Poorly defined requirements on Science Platform functionality to support ESFRI science | <ul style="list-style-type: none"> Primarily addressed through D5.1 & Project Platform, which extensively document the use cases being addressed. Ongoing engagement of ESFRIs with WP5 to refine and address use cases. Successful summer 2021 workshop validated project goals with ESFRIs. |
| 18 | Lack of commonality between ESFRI requirements | <ul style="list-style-type: none"> Modular design of ESAP makes it possible to service multiple ESFRI needs from a common core. |
| 19 | Lack of availability or immaturity of underlying core EOSC services | <ul style="list-style-type: none"> Development and test is isolated from EOSC services. Active engagement with the EOSC Future TSPs will drive EOSC integration. |
| 20 | Insufficient ICT infrastructure available to support Science Platform deployment and testing | <ul style="list-style-type: none"> Adequate infrastructure is available thanks to resource made available by a variety of project partners (thank you!). |



Risks Analysis: Detailed Project Plan

- Lower-level risks are identified in D5.2 (Detailed Project Plan).
- “Retention of key staff, with domain knowledge” is (partially) realised through extended leave of WP5 Technical Lead.
- Mitigations put in place included proactive engagement of the WP Coordinator, development of a common ESAP vision, and formation of a regular technical leadership team.
- Milestone MS31 and Deliverable D5.3 were delayed relative to the original scheduled, but delivered in August.
- No other major concerns.
- We continue to monitor and proactively respond to risks.



Future plans

- *Focus on integration and stabilization of the various capabilities than have been developed across the work package and the ESCAPE ecosystem:*
 - Complete integration of the various data and analysis services which have been developed by project partners.
 - (Continue to) engage with ESFRIs on ESAP deployment and service provisioning.
 - Provide support to EOSC Future Test Science Projects (we have already had significant engagement with the Extreme Universe TSP).
- *ESAP technical improvements:*
 - Better support for asynchronous services.
 - Integrate more execution engines, providing a wider variety of compute and analysis task types.
 - More advanced software selection and dispatching: deeper integration with OSSR and with DIOS data locality.
 - Deploy a batch processing management system based on DIRAC.



Conclusions

- ESAP provides a central coordinating element across the diverse ESCAPE ecosystem, bringing together work from WPs 2, 3, and 4 into a single user-facing system.
- During RP2, the WP5 team has made important technical progress...
 - ...on development of the ESAP system itself.
 - ...on packaging and providing a range of scientific analysis capabilities.
 - ...on engaging with the other work packages.
- Over the next several months, we will focus on consolidating and integrating this work so that we deliver a compelling package to all ESCAPE ESFRIs.

