



ESCAPE

European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures

ESAP Training Workshop

John Swinbank — swinbank@astron.nl



Welcome & Logistics



Welcome

- *Final public meeting of ESCAPE Work Package 5.*
- Introduce and explain the work we've been doing since 2019.
- Our aim is to inspire you...
 - ...to build on this work in your own projects.
 - ...to join with us in future development.
- Day 1: Deep dive on ESAP, the “ESCAPE Science Platform”.
- Day 2: Related work and future prospects.
- None of this is “finished”; there's enough here to be interesting, but plenty more work to be done.
- Plenty of time for discussion: brief questions after each presentation; extended discussion blocks at the end of each session.
- <https://indico.in2p3.fr/e/EsapTraining> for the schedule.



ESAP & ESCAPE: An Introduction



ESCAPE Goals & Fast Facts

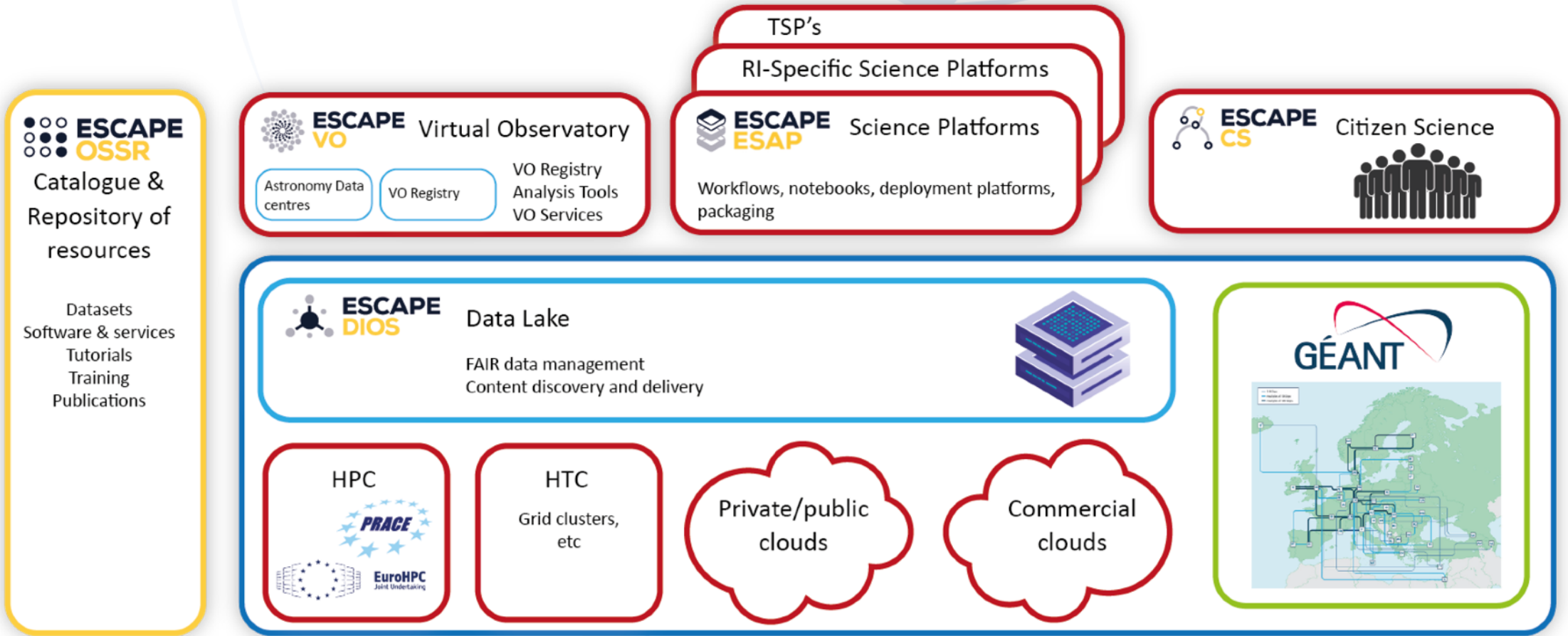
- Develop common “e-infrastructure” solutions that benefit a wide range of particle physics & astronomy research facilities.
- 31 different partners across 8 countries.
- 10 ESFRI projects and landmarks.
- €16M budget.
- Project runs from 2019 until 2023.



ESCAPE Research Infrastructures



ESCAPE Work Packages



Acronyms

- **DIOS: Data Infrastructure for Open Science**
- **CEVO: Connecting ESFRI projects to EOSC through a Virtual Observatory framework**
- **ESCAPE: European Science Cluster of Astronomy and Particle physics ESFRI research infrastructures**
- **ESFRI: European Strategy Forum on Research Infrastructures**
- **EOSC: European Open Science Cloud**
- **ESAP: ESFRI Science Analysis Platform**
- **OSSR: Open source scientific Software and Service Repository**

- So that means that “ESCAPE ESAP” is the European Science Cluster of Astronomy and Particle physics European Strategy Forum on Research Infrastructures research infrastructures European Strategy Forum on Research Infrastructures Science Analysis Platform.

...watch out for a quiz later!



The Scientific Data Lake



Manage large datasets
without worrying of your
data storage space
limitations



Support for joint
processing of data drawn
from multiple sources

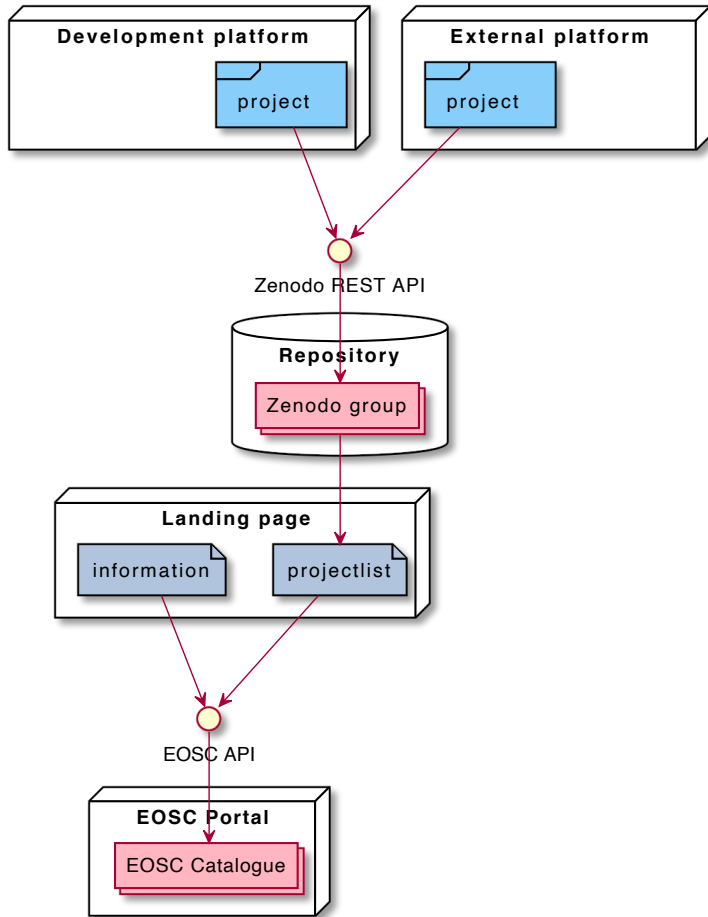


Apply FAIR data
management principles

A **federated data infrastructure** following **FAIR** data principles and providing a flexible and robust system to **efficiently and securely store and transfer large data volumes**. Support orchestration to make data accessible across different communities.



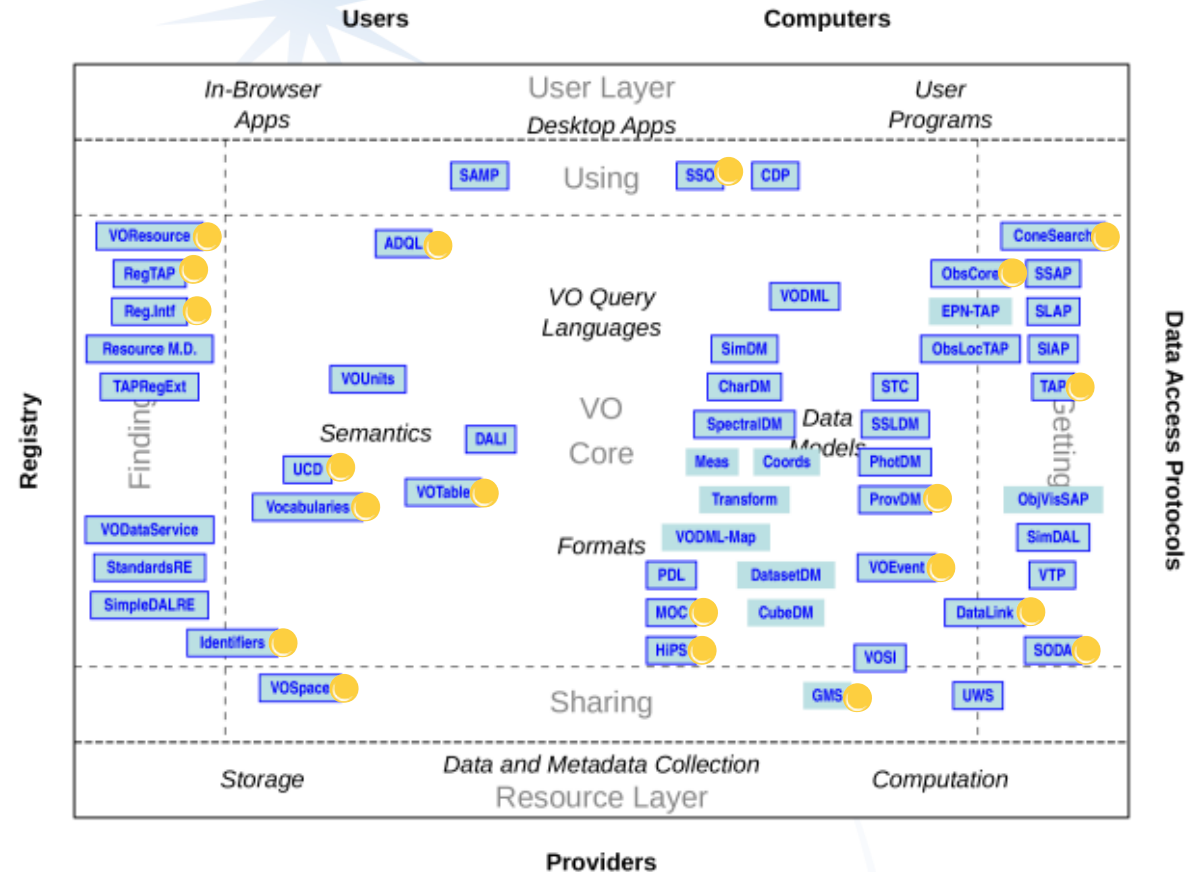
Software & Services Repository



- A common repository for software across the various ESCAPE infrastructures.
- Designed both to *make software available* to practising scientists and to *preserve* software for future reproducibility.
- Built-in to common CI/CD systems (GitLab, Hub, etc).
- Built around the Zenodo system for long-term stability and preservation.
- Programmatic access through the eOSSR library.
- Integrated with the ESAP platform... see later.



- ESCAPE aims to build around open standards wherever possible.
- We have used and contributed to a bunch of IVOA standards. In particular, the project has pushed on:
 - Using IVOA semantic UCD metadata for solar physics.
 - Developing IVOA standards for radio astronomy metadata.
 - Deploying IVOA-compliant services in the ESO science archive.
- Education, education, education, ...



Citizen Science



Join ambitious
experiments



Contribute to
extraordinary scientific
discoveries



Take part in the Open
Science revolution

An ambitious programme that allows the **science-inclined public** to take part in **research projects** to perform tasks that could not otherwise be solved algorithmically and that will contribute to making a wide variety of **fundamental scientific discoveries**.



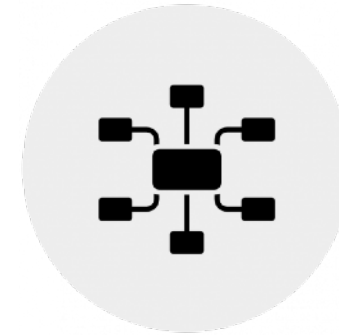
What I talk about when I talk about science platforms



Access and share your data and results



Work with advanced analysis tools



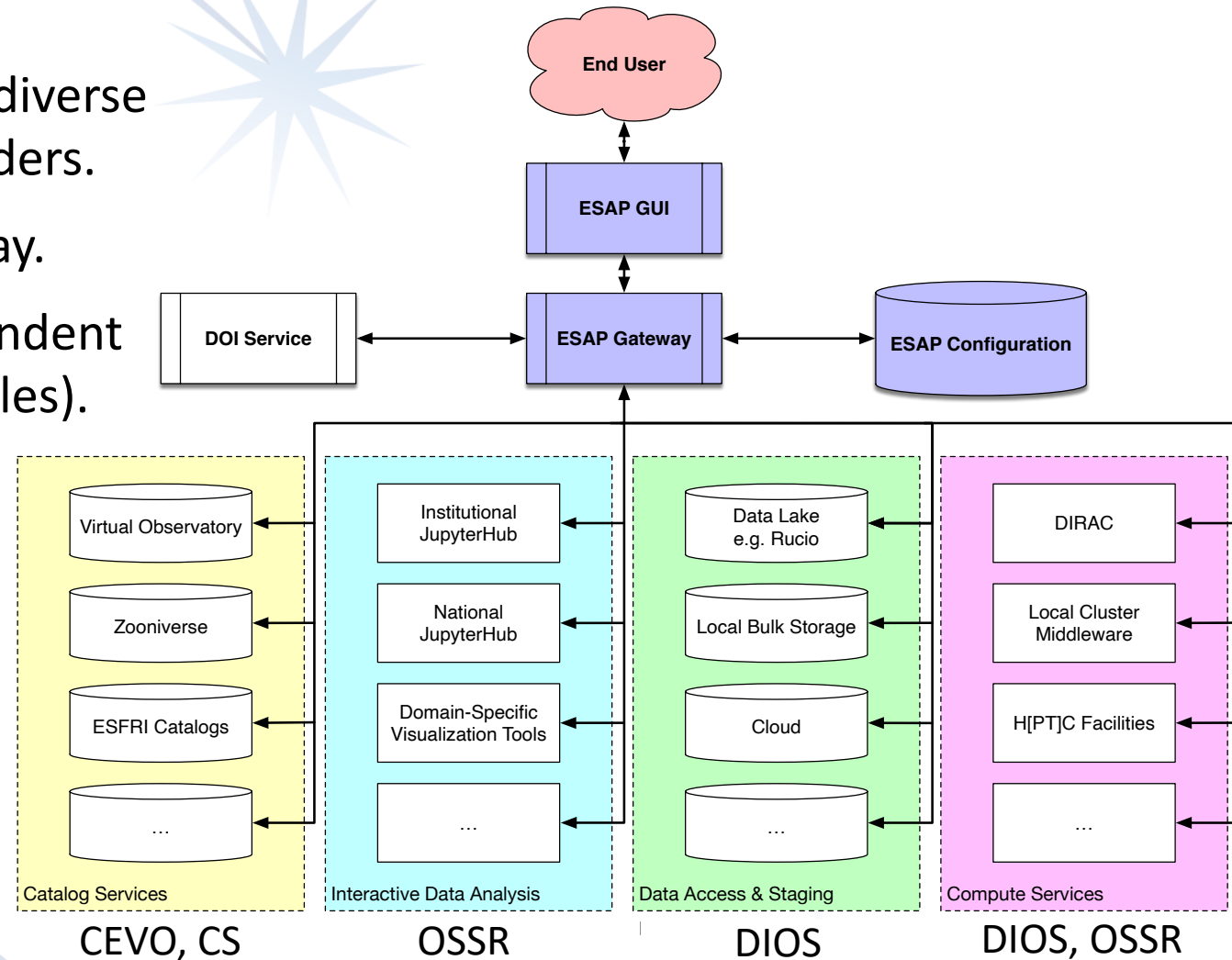
Integrate with a wide variety of other services



Logo credits to upstream projects



- ESAP provides a focal point for integrating diverse services which are drawn from other providers.
- Two part structure: web UI, and API gateway.
- Focal point of a range of pluggable, independent services (some of which are WP5 deliverables).
- Designed to be flexible, robust and extensible.
- WP5 does not provide an “operational” service: ESAP is a software development & integration project, which is made available to ESFRIs and successor projects (EOSC Future, etc) for them to customize and deploy.



A platform... or a toolkit?

- ESCAPE's mission is not to build and maintain Jupyter, batch computing services, or similar for common/EOSC access.
 - Many ESCAPE partner institutions do make available systems for testing, developing, and experimenting on.
- **ESAP is a *toolkit* for building “science platforms” which are customized to particular applications.**
- Deployable at a variety of scales:
 - “Centralized ESAP”, providing flexible and convenient access to a wide spectrum of ESCAPE services.
 - “Project ESAP”, providing a way for individual infrastructures, projects, etc to quickly integrate diverse capabilities into a unified service offering.



Example 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

Demo system

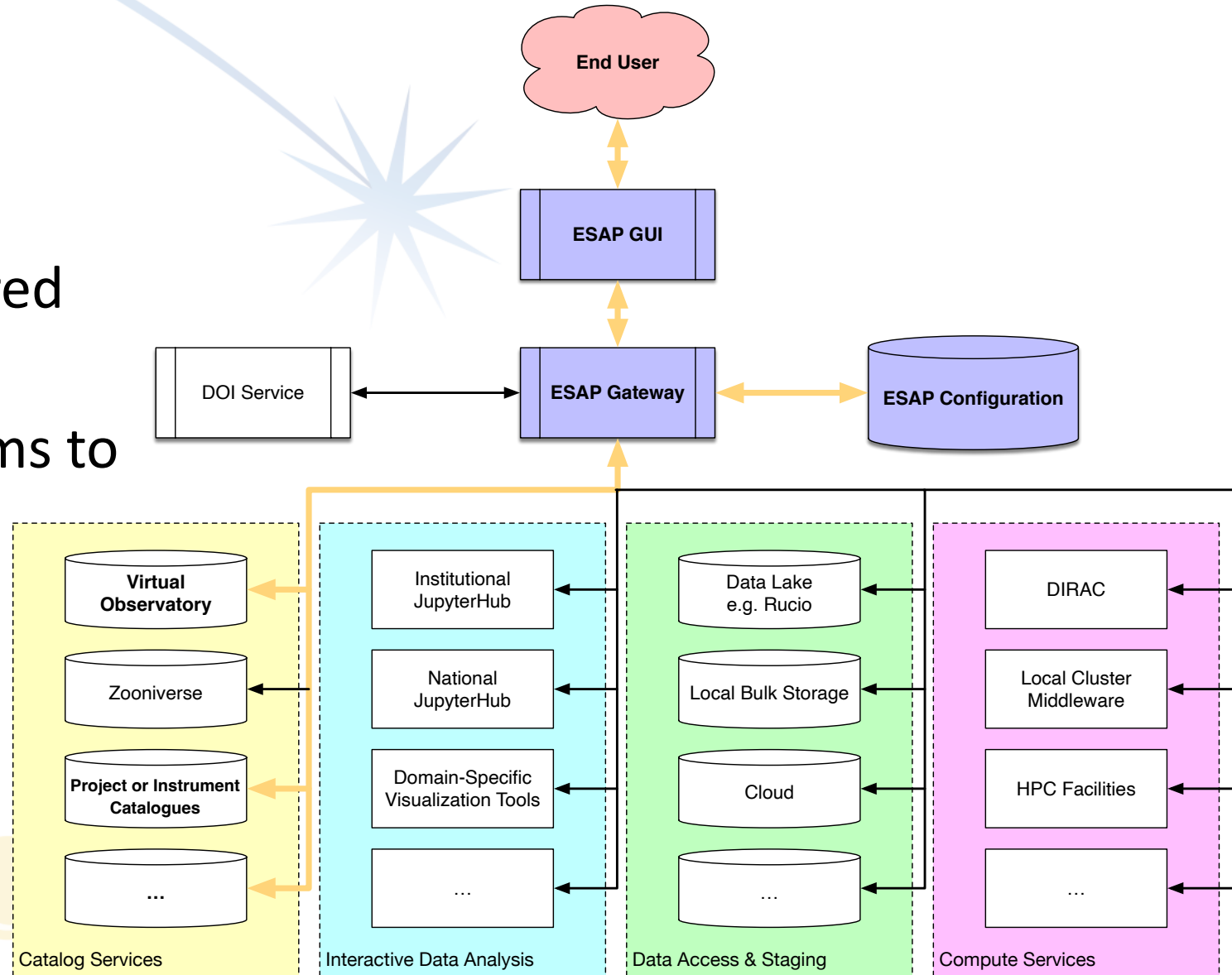
<https://sdc-dev.astron.nl/esap-gui>



Example Workflow

1. Query

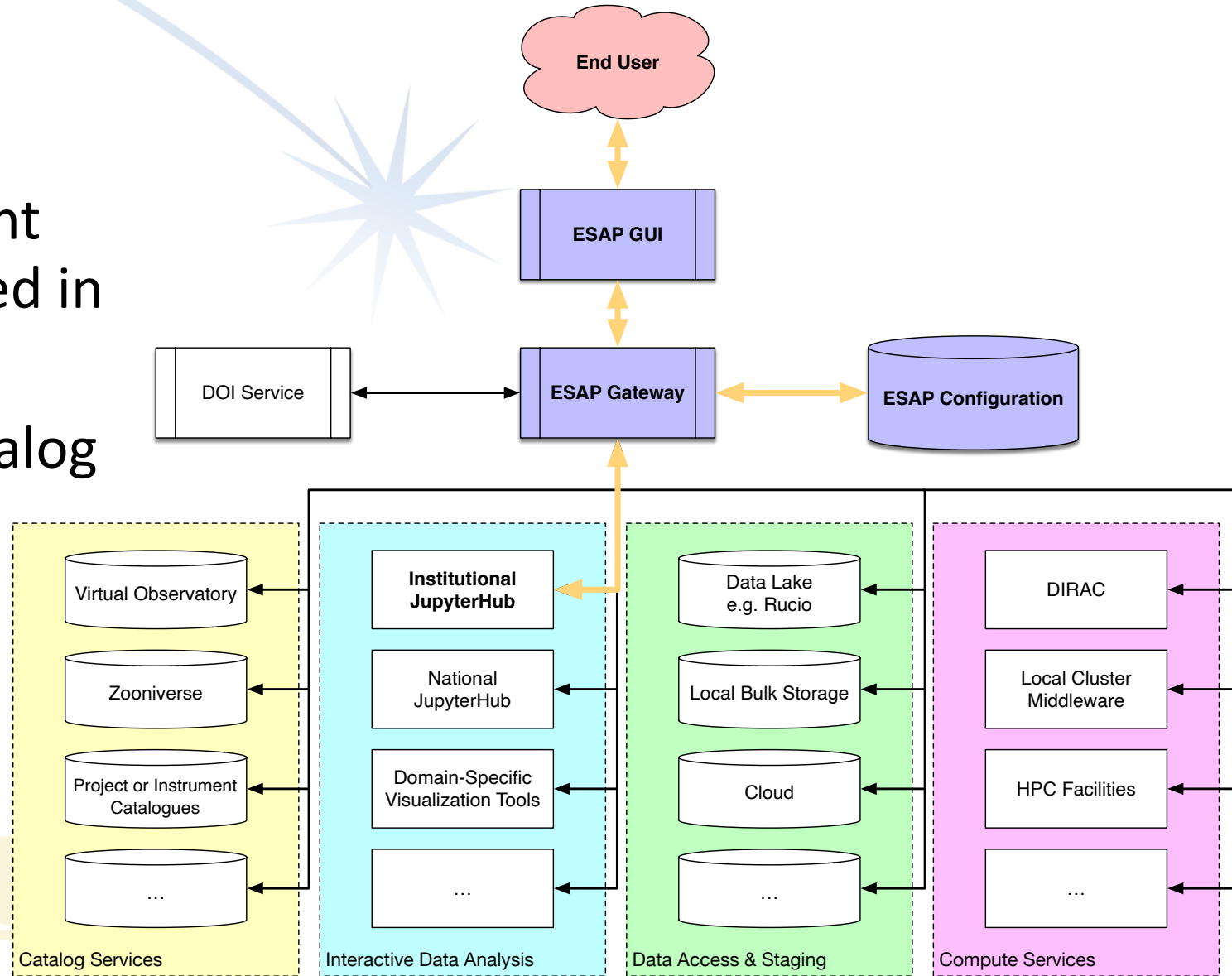
- User identifies relevant catalog services configured in this instance of ESAP.
- User submits search terms to multiple catalogs using consistent ESAP UI.
- Search results returned to user and displayed in unified form.



Example Workflow

2. Winnow

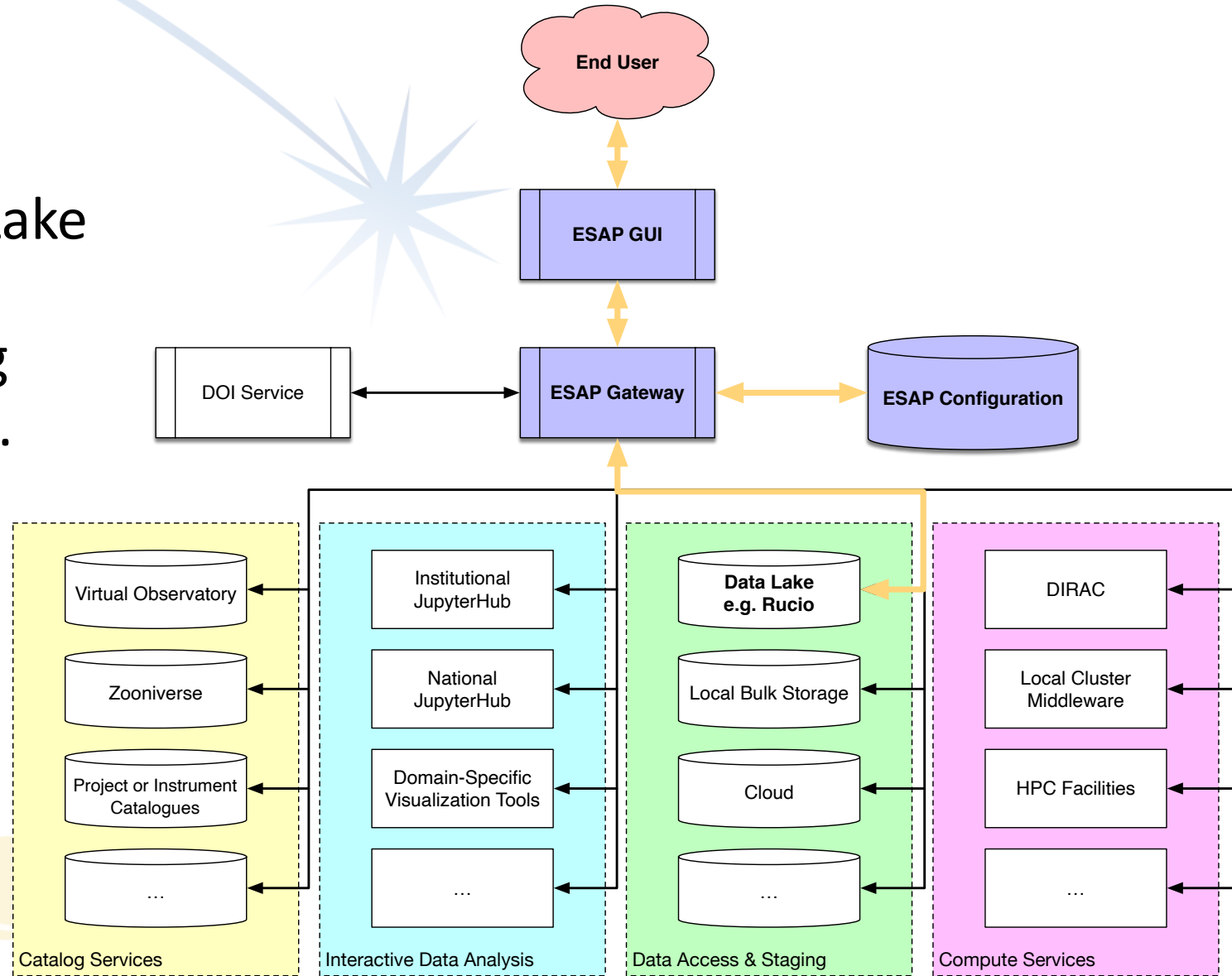
- User identifies convenient Jupyter system configured in this instance of ESAP.
- User sends retrieved catalog data to notebook.
- Interactive analysis session in notebook identifies most promising candidates for bulk processing.



Example Workflow

3. Stage Data

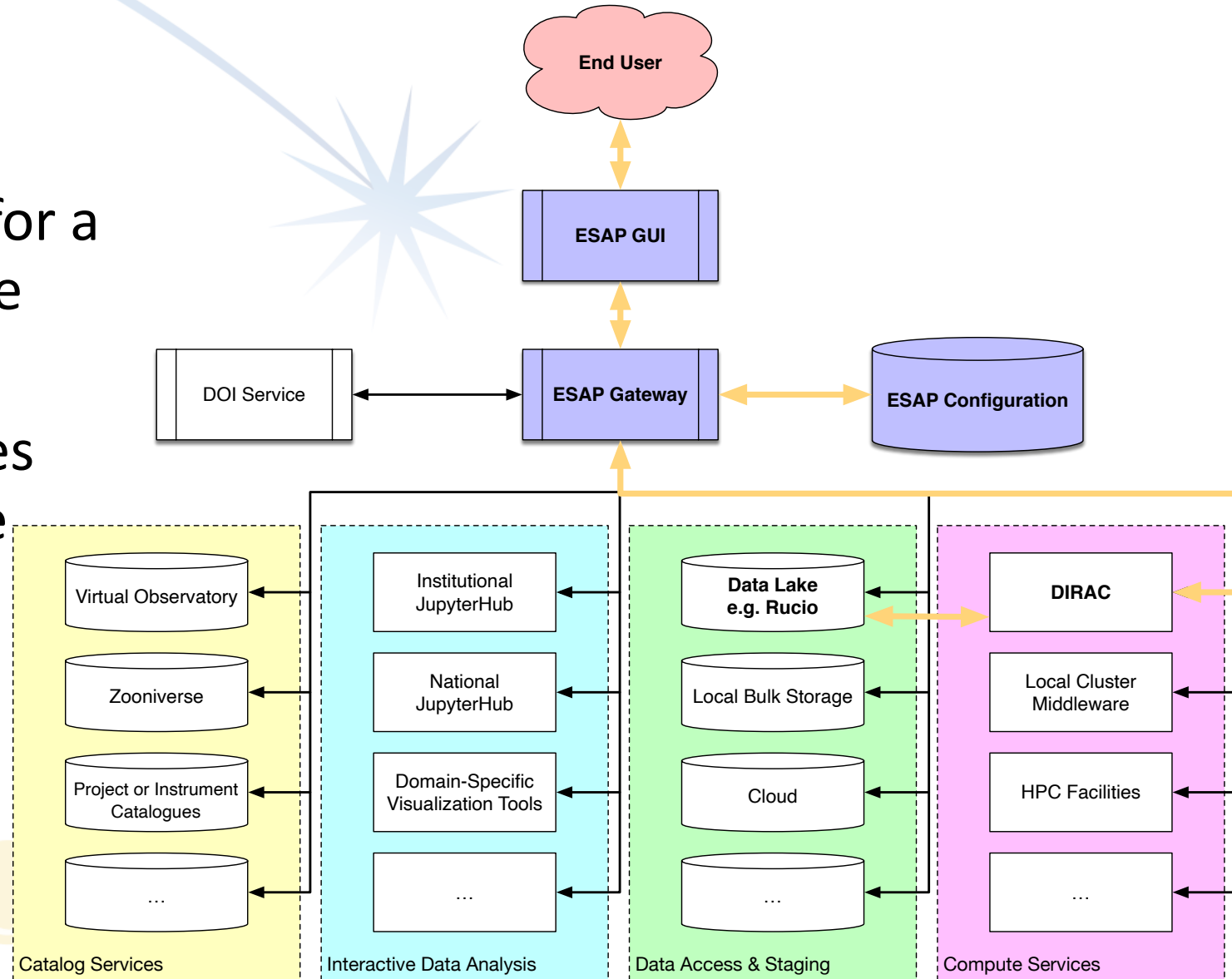
- User instructs the Data Lake to “stage” archived data corresponding to catalog entries to online storage.



Example Workflow

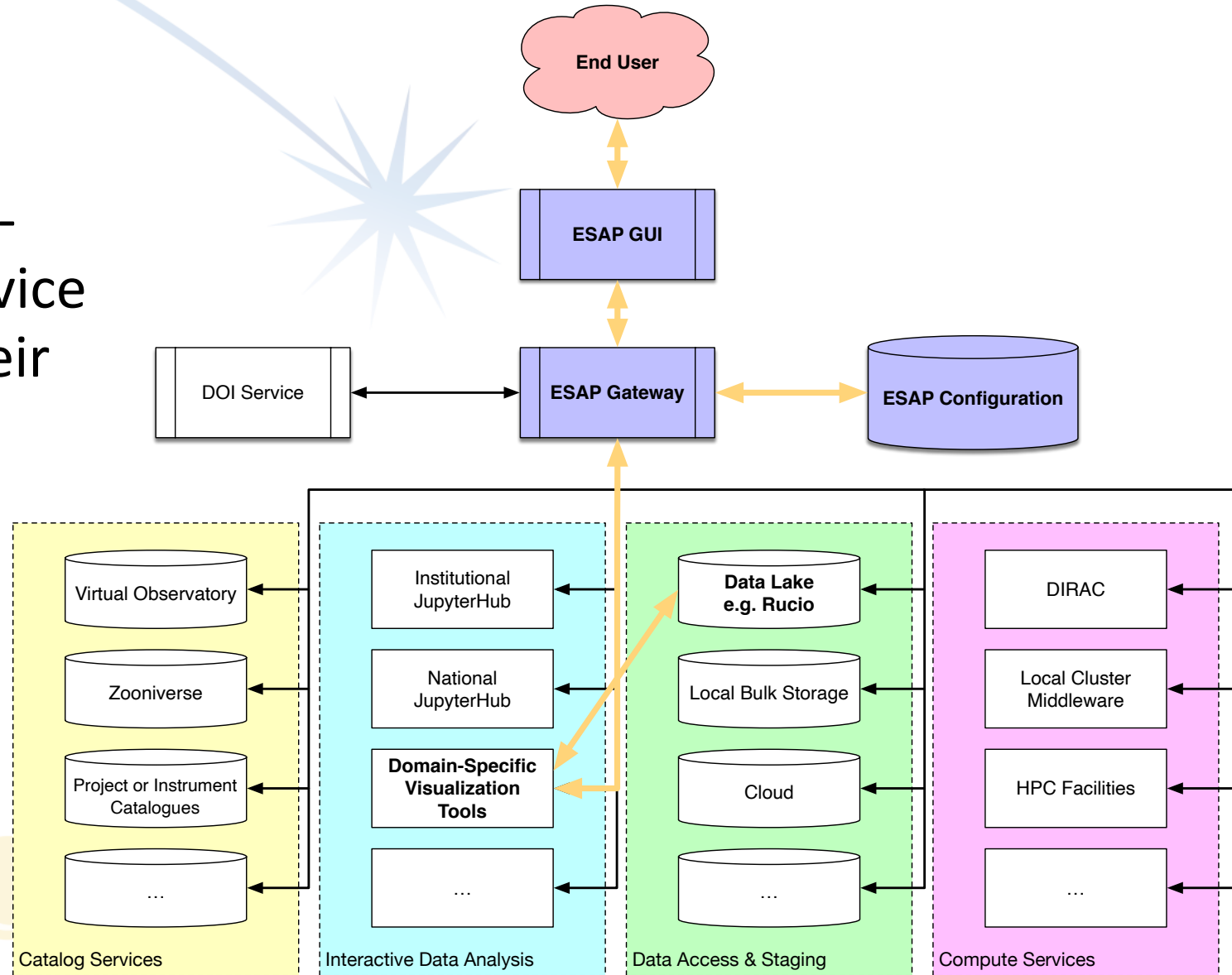
4. Compute

- User sends instructions for a batch compute job to the DIRAC cluster.
- Compute cluster retrieves the staged data from the Data Lake.
- Batch processing happens.
- Results are stored to Data Lake, and the user notified of completion.



5. Visualize Results

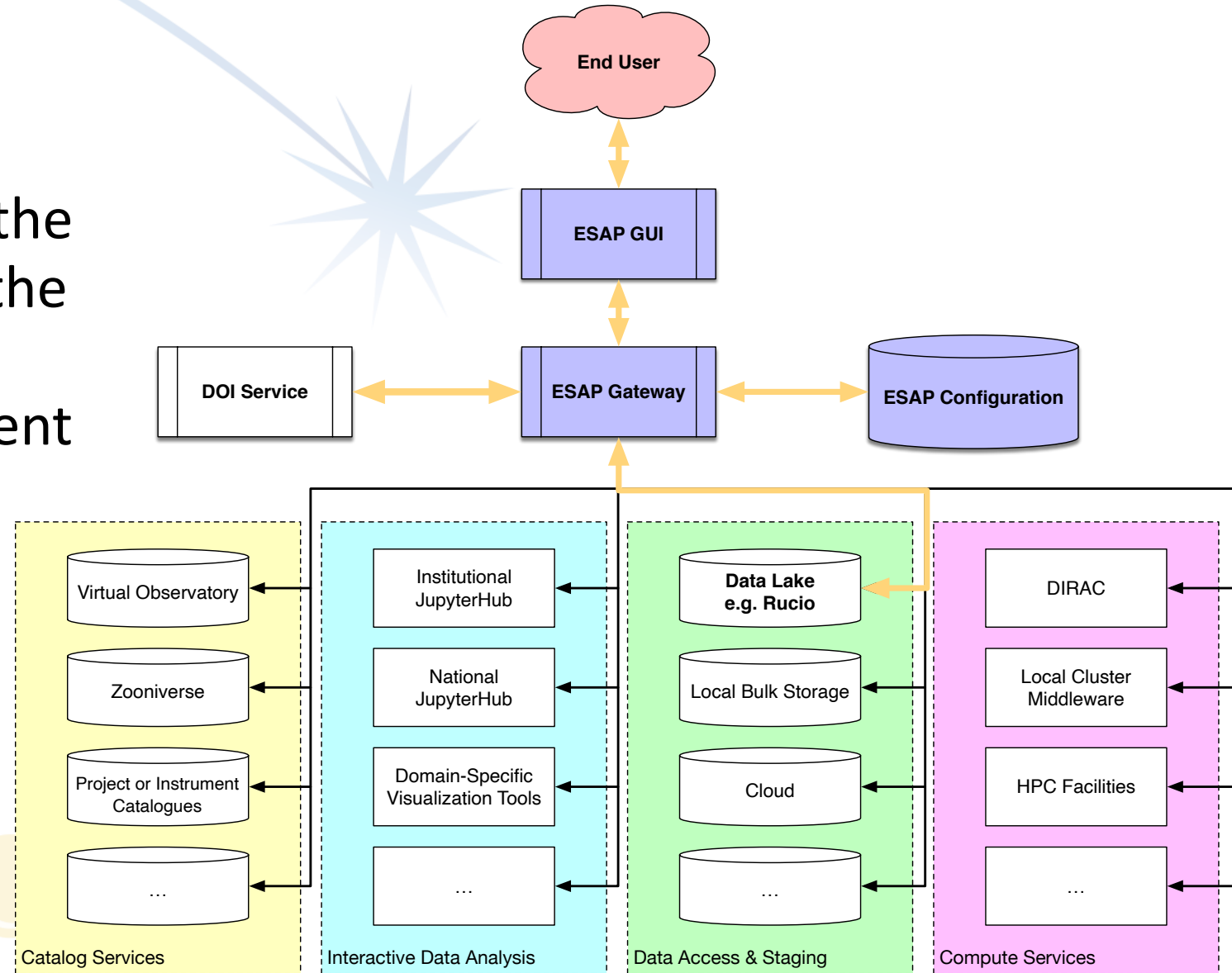
- User identifies a domain-specific visualization service that can help analyze their data.
- User initiates a visualization session, passing location of compute results.



Example Workflow

6. Publish

- Having established that the results are noteworthy, the user instructs the DOI Service to mint a persistent identifier for them.
- The results are made available to the wider community.



References

- ESCAPE:
 - <https://www.projectescape.eu>
- ESAP documentation
 - <https://git.astron.nl/astron-sdc/esap-api-gateway/-/wikis/home>
- WP5 code repositories
 - <https://git.astron.nl/astron-sdc/escape-wp5>

And now on with the show...

