



ESCAPE

European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures

IVOA standards in ESCAPE-ESAP functionalities

**S. Bertocco, K.Kliffen, D.Morris, S.Voutsinas, S.Sanchez,
M.Parra, J.Swinbank**

IVOA Interoperability Meeting

25-29 April 2022

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.



ESCAPE EU Project

European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures

▶ **Data Lake (DIOS: Data Infrastructure for Open Science):**

Build a scalable, federated, data infrastructure for open science for the ESFRI
Enable connection to compute and storage resources.



▶ **Software Repository (OSSR: Open-source Scientific Software and Service Repository):**

Open source scientific software and service repository.



▶ **Virtual Observatory (CEVO: Connecting ESFRI projects to EOSC through VO framework):**

Extend FAIR standards, methods, tools of the Virtual Observatory;
demonstrate EOSC ability to include existing platforms



▶ **Science Platforms (ESAP: ESFRI Science Analysis Platform):**
Flexible science platforms to enable the analysis of open access data



▶ **Citizen Science (ECO:Engagement and COmmunication):**

Open gateway for citizen science on ESCAPE data archives and ESFRI CS projects



Summary

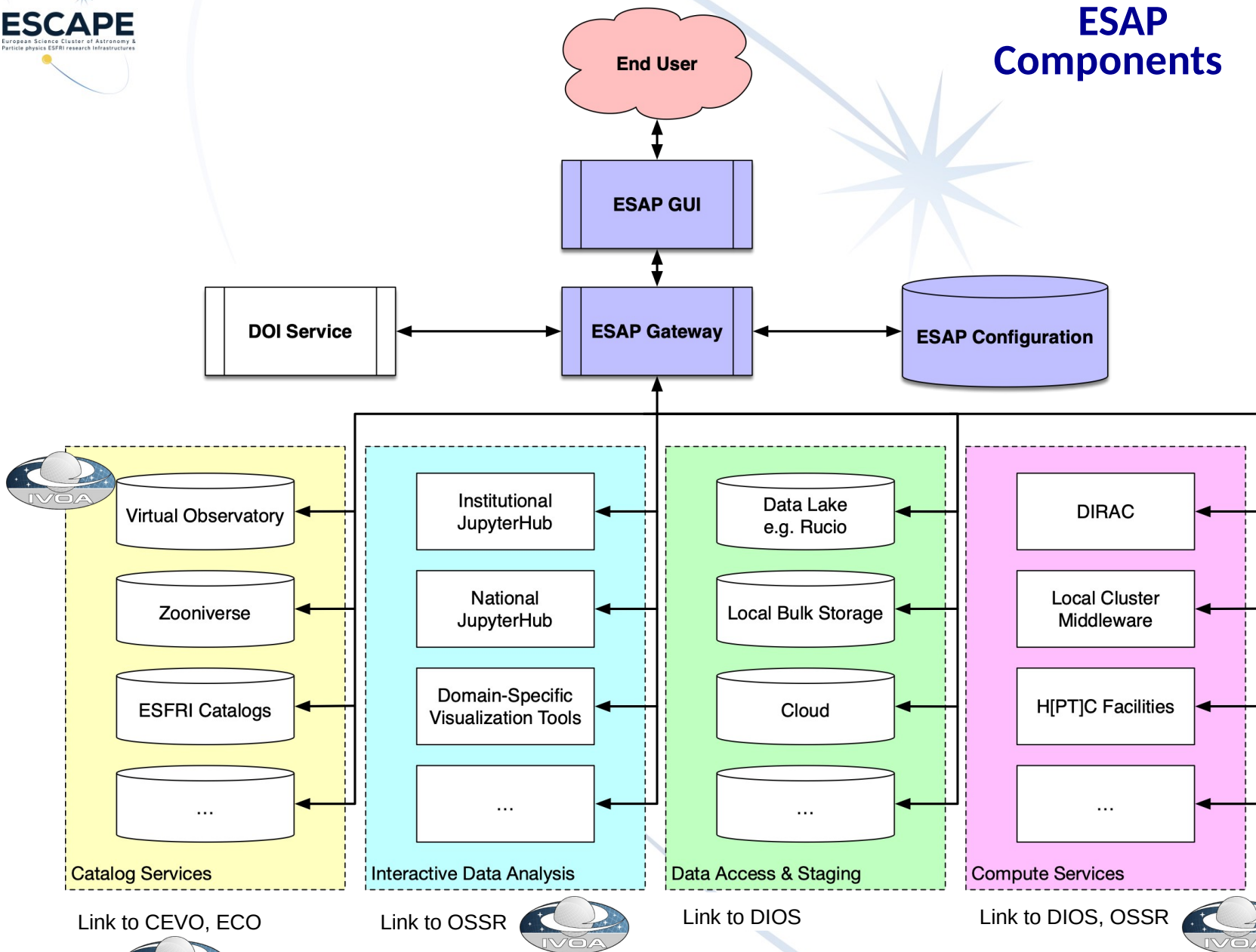
- ▶ Focus on ESAP (ESFRI Science Analysis Platform)
 - components
 - features

- ▶ Where IVOA is in ESAP
and
- ▶ where IVOA could be in ESAP

- ▶ Suggestions for discussion



ESAP Components



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 screenshots show various components of the system:

- ASTRON Data Explorer:** A search interface for astronomical data with filters for Catalog, Target, RA (degrees), Dec (degrees), and Visibility.
- Interactive Analysis Workflows:** A page listing workflows such as 'CSIC-IAA HCD-16 workflow' and 'COS MOCPY' with search and 'View Tutorial' buttons.
- Interactive Analysis Compute Facilities:** A page listing facilities like 'MyBinder', 'JVE BinderHub', and 'SKAO BinderHub' with search and 'View Tutorial' buttons.
- Workflow Editor:** A Python-based interface for creating workflows, showing code for fetching data from the ESAP Shopping basket.
- Virtual Observatory (VO) Overview:** A page listing various VOs like WSRT-Apertif, ASTRON VO, Zooniverse, and Virtual Observatory (VO).
- ASTRON VO Archives:** A table listing data products with columns for Name, RA, Dec, lmv, DataProduct Type, DataProduct SubType, and Dataset.

Demo system

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



Existing Query Capabilities

ESAP can query the registry for services and data

ADQL Query

```
SELECT TOP 100 * from ivoa.obscore WHERE obs_collection='apertif-dr1' and dataproduct_subtype='continuum'
```

Service Metadata

<https://vo.astron.nl/tap> | 11 keys

Selected Services

<https://vo.astron.nl/tap>

Query VO Resource

ESAP IVOA Query

Archives Multi Query Interact

Catalog* Keyword

IVOA radio

Query VO Registry

List of VO Resources

1

Results from https://vo.astron.nl/tap

Basket	dataproduct_type	dataproduct_subtype	calib_level	obs_collection	obs_id	obs_title	obs_publisher_did	obs_creator
<input type="checkbox"/>	image	continuum	3	apertif-dr1	190807041	190807041_AP_B001	ivo://astron.nl/-7/APERTIF_DR1/190807041_AP_B001/image_mf_02.fits	

Resource	Access URL	Waveband	Title	Service Type	Content Types
<input type="checkbox"/> CSIRO ASKAP TAP	https://casda.csiro.au/casda_vo_tools/tap	radio	CSIRO ASKAP Science Data Archive TAP Service	vs.catalogservice	[other]
<input type="checkbox"/> CSIRO Pulsar TAP	https://data.csiro.au/psrdavo/tap	radio	CSIRO Parkes Pulsar Data Archive	vs.catalogservice	[other]
<input type="checkbox"/> HALCA	http://jvo.nao.ac.jp/skynode/do/tap/halca	radio	HALCA VSOP (the VLBI Space Observatory Programme) Correlated Data	vs.catalogservice	[archive]

ESCAPE ESAP Archives Multi Query Interactive Analysis IVOA-SAMP

Archive - Virtual Observatory (VO)

Instrument Virtual Observatory

Description Virtual Observatory (VO)

Virtual Observatory (VO)

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

Data Retrieval

Data from the VO can be downloaded using the query functionality of the web page. Also, the VO standard protocols can be used to find the data that is shared through the service using external tools like TOPCAT, Aladin or DS9.

Data in Virtual Observatory (VO)

Dataset or Category	Catalog	Query Access
VO Registry	Virtual Observatory (VO)	Query this Dataset

VO Registry

The table access protocol (TAP) defines a service protocol for accessing general table data, including astronomical catalogs as well as general database tables. Access is provided for both database and table metadata as well as for actual table data. This version of the protocol includes support for multiple query languages, including queries specified using the Astronomical Data Query Language ADQL within an integrated interface. It also includes support for both synchronous and asynchronous queries. Special support is provided for spatially indexed queries using the spatial extensions in ADQL. A multi-position query capability permits queries against an arbitrarily large list of astronomical targets, providing a



ESAP Software Execution (1)

Interactive Analysis Workflows

Search for Workflows

Next

Advanced Search

CSIC-IAA HCG-16 workflow

Description: Analysis of Hickson Compact Group 16
Link: <https://github.com/AMIGA-IAA/hcg-16>
Author:
Runtime Platform:
Keywords: jupyter-notebook

CDS MOCpy

Description: Experiment with Multi-Order Coverage maps
Link: <https://github.com/cds-astro/mocpy>
Author:
Runtime Platform:
Keywords: jupyter-notebook

Getting Started
 September 28, 2021 Software Open Access

AMIGA-IAA/hcg-16: Repo synced with Zenodo

Mike Jones; Sebastian Luna-Valero; Julián Garrido; Susana Sánchez Expósito

This zenodo registry is related to the gitHub repository hcg-16 (<https://github.com/AMIGA-IAA/hcg-16>), which hosts a pipeline to reproduce the data reduction and analysis of Jones et al. 2019

This and future releases will be archived with Zenodo.

Preview

- hcg-16-v1.2.1.zip
- AMIGA-IAA-hcg-16-b9f7e10
 - LICENSE 1.1 kB
 - README.md 4.2 kB
 - casa
 - calibration_flag.py 28.7 kB
 - imaging.py 10.0 kB
 - cgatcore
 - pipeline.py 3.9 kB
 - pipeline.yml 460 Bytes
 - codecheck.yml 4.2 kB
 - docker
 - Dockerfile.casa.5.1 588 Bytes
 - Dockerfile.casa.5.4 570 Bytes
 - Dockerfile.sofia 1.1 kB
 - environment.yml 261 Bytes
 - plot_scripts
 - Fig1-DECaLS_grz_image.ipynb 3.4 kB

Files (64.6 kB)

Name	Size
AMIGA-IAA/hcg-16-v1.2.1.zip	64.6 kB

84 views 23 downloads

Available in: GitHub, OpenAIRE

Publication date: September 28, 2021
 DOI: 10.5281/zenodo.5534682
 Keyword(s): Astronomy and Astrophysics, VLA, Radioastronomy, Compact groups, Galaxies
 Grants: European Commission
 • ESCAPE - European Science Cluster of Astronomy & Particle physics ESFRRI research Infrastructures (914641)

Workflow metadata description needed:

- OSSR & CEVO working on it
- codemeta.json
- <https://escape2020.pages.in2p3.fr/wp3/eossr/v0.4/metadata.html>



Interactive Analysis

Compute Facilities



Deploy

JIVE BinderHub

Description:

JIVE BinderHub

Link: <http://jupyter.jive.nl/binderhub/>

MyBinder

Description:

MyBinder

Link: <https://mybinder.org/>

Rosetta @ INAF OATS



Compute facilities metadata description needed:

- ExecutionPlanner

<https://github.com/ivoa/ExecutionPlannerNote>

& Dave Morris presentation



Asynchronous Tasks in ESAP

- ▶ Since now ESAP only performed synchronous queries
- ▶ VO defines a number of asynchronous protocols (e.g. async TAP) for longer running queries, wish to implement these in the ESAP Query App
- ▶ Store references to large results (1000+ rows) in Shopping basket
- ▶ Starting (and keeping track of) Batch computation jobs (e.g. simulations)
- ▶ Orchestrating Interactive Analysis platforms and data staging (e.g. jupyter hubs)



- Provides a specification for a REST API
- Data model and control flow for Jobs
- Existing clients
- Don't reinvent the wheel for ESAP

Universal Worker Service Pattern Version 1.1

IVOA Recommendation 24 October 2016

Interest/Working Group:

<http://www.ivoa.net/twiki/bin/view/IVOA/IvoaGridAndWebServices>

Author(s):

P. A. Harrison, G. Rixon

Editor(s):

P. A. Harrison

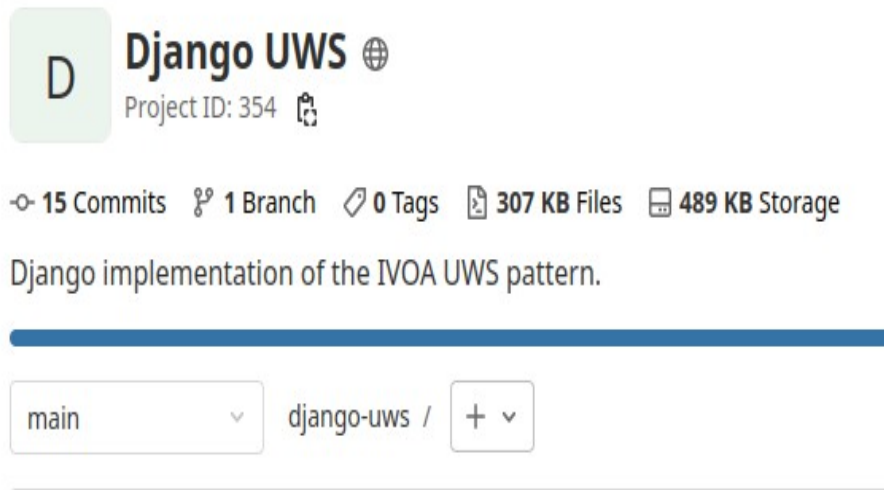
DOI:



10.5479/ADS/bib/2016ivoa.spec.1024H








IVOA-UWS in ESAP




- Portable Django App (integrates easily in existing Django Projects, such as ESAP)
- Uses Django REST Framework for serialization and REST interface
- Alpha version: not fully compliant with the specification; not on PiPy yet
- Uses JSON instead of XML



D Django UWS 
Project ID: 354 

 15 Commits  1 Branch  0 Tags  307 KB Files  489 KB Storage

Django implementation of the IVOA UWS pattern.

main  django-uws /  

<https://git.astron.nl/astron-sdc/django-uws>



UWS Worker & Job Description Language

- Initial Job Description Language single parameter: type:

Query.tap

Orc.jupyter

- Job implementations should specify more parameters

```
def init_workers() -> Dict[str, Worker]:  
    tmp = {}  
    for worker_name in getattr(settings, "ESAP_WORKERS", []):  
        worker: Worker = import_string(worker_name)()  
        tmp[worker.type] = worker  
    return tmp
```

```
REGISTERED_WORKERS = init_workers()
```

```
37         try:  
38             worker_type = parameters["type"]  
39             worker = REGISTERED_WORKERS[worker_type]  
40             worker.run(job)  
41  
42             job.phase = EXECUTION_PHASES.COMPLETED  
43         except Exception as err:  
44             # TODO: set error object  
45             job.phase = EXECUTION_PHASES.ERROR
```



Future of UWS in ESAP

- ▶ Implement various kinds of Jobs in the workers and formalise the Job Description Language (JDL)
- ▶ Implement all parts of the REST API in the Django package
- ▶ Implement XML serialization for compatibility with existing clients
- ▶ Worker facing API replacing the Database connection and Django ORM

Suggestions:

- ▶ Extend the specification with a JSON data schema for easier integration with modern Javascript Frameworks (React, Angular etc.)
- ▶ Endpoint for registered jobs should be able to specify the parameters they allow



Conclusion

- ▶ How IVOA standards “helped” ESAP features development
- ▶ How IVOA could help science platform development with new standards
- ▶ Enforce ExecutionPlanner development and standardization
- ▶ Suggest to introduce JSON in messages exchange in IVOA standards





ESCAPE

European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures



Thank you !



Backup Slides



ESCAPE EU Project

European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures

▶ **Data Lake (DIOS: Data Infrastructure for Open Science):**

Build a scalable, federated, data infrastructure for open science for the ESFRI
Enable connection to compute and storage resources.



▶ **Software Repository (OSSR: Open-source Scientific Software and Service Repository):**

Open source scientific software and service repository.



▶ **Virtual Observatory (CEVO: Connecting ESFRI projects to EOSC through VO framework):**

Extend FAIR standards, methods, tools of the Virtual Observatory;
demonstrate EOSC ability to include existing platforms



▶ **Science Platforms (ESAP: ESFRI Science Analysis Platform):**
Flexible science platforms to enable the analysis of open access data




▶ **Citizen Science (ECO:Engagement and COmmunication):**

Open gateway for citizen science on ESCAPE data archives and ESFRI CS projects

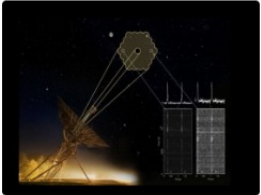


ESAP features



[Logout Sara Bertocco](#)


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).

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](#)

Virtual Observatory (VO)



Virtual Observatory (VO)

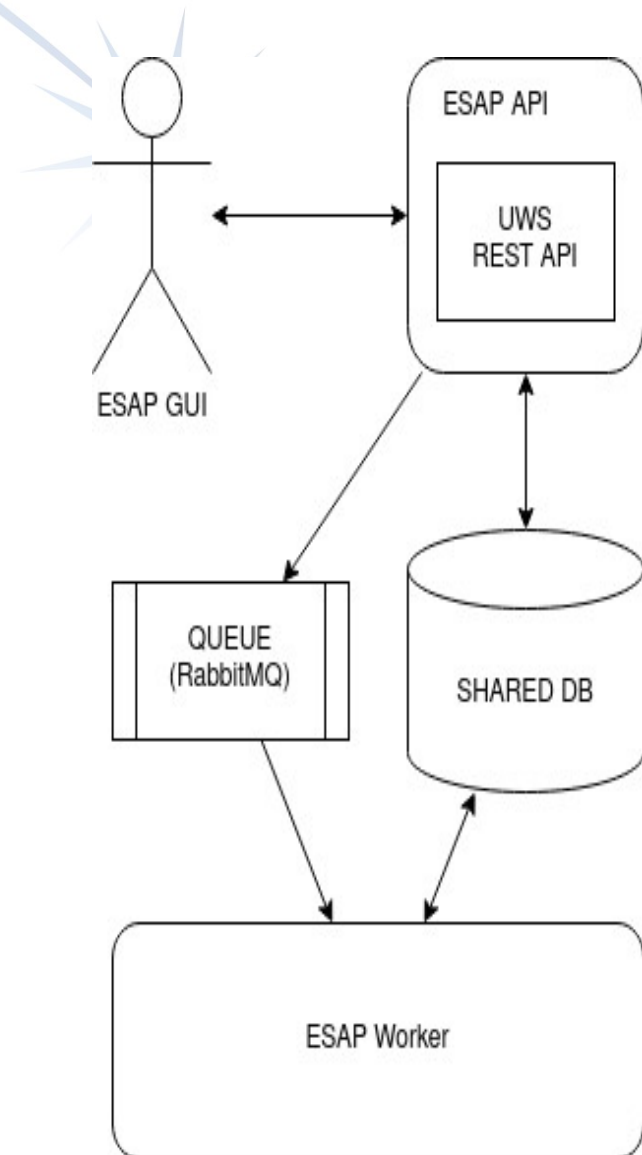
The Virtual Observatory defines a set of standards that can be used to download astronomical data.

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



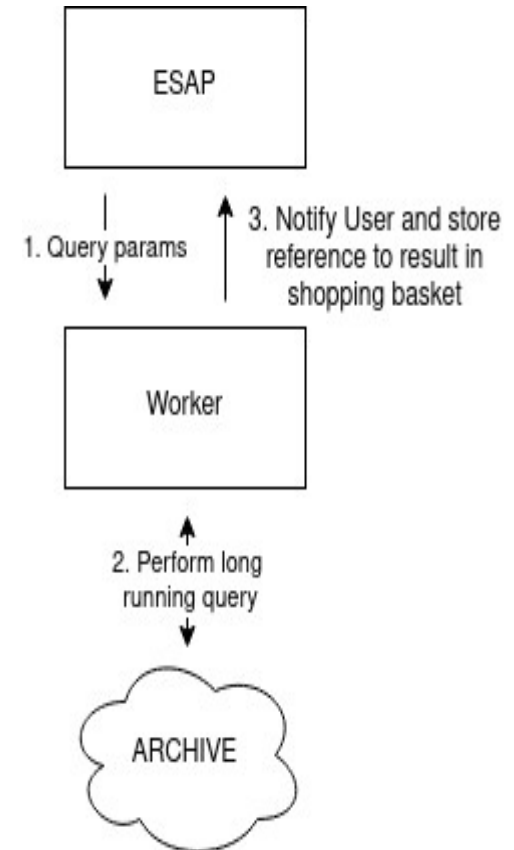
Asynchronous ESAP - Architecture

- REST API
 - **UWS specification**
 - Build as Portable Django Application
- Celery Queueing system for message brokering to Worker
 - Support for various queue software: RabbitMQ, Redis, AmazonSQS, etc.
- Celery Worker for performing tasks



Async Query Workflow

- User specifies query parameters in the ESAP GUI
- Parameters are sent to the async Worker
- Worker performs the Query
- Worker stores reference to results in the shopping basket
- Worker in addition can notify the user that the query is done



Orchestration Workflow: BinderHub example

- Current focus on BinderHub: git repository -> JupyterLab interface
- User selects data in shopping basket, software workflow, target system
- Job started:
 - Clone the repository
 - Create a “postBuild” script to download data from Shopping Basket
 - Makes the repository available to BinderHub service (needs to be public)
 - Performs request to BinderHub to start a server and wait for it to be ready
 - Reports back with the URL so ESAP GUI can redirect the user
 - Cleans up the modified repository.

