# Aim & Outline

Aim: *To assess the feasability of using DIRAC as part of the ESAP.*

- What is DIRAC?
- CTA & DIRAC
- Relation to other Workpackages and Focus Groups
- Identified open issues
- Planning
- Use case example for CTA

# Today

- 15min Overview of WP5 DIRAC task (Gareth, Matthias)
  - mainly our efforts to find a place for DIRAC
- 15min Overview of WP3 DIRAC efforts (Luigi, Andrei)
- 10min KM3NeT Use Case (Jutta)
- 10min CTA Use Case and (Johan)

A session to trigger discussions..

# What is DIRAC?

- DIRAC: **D**istributed **I**nfrastructure with **R**emote **A**gent **C**ontrol
- An open source **software framework** for distributed computing
  - http://diracgrid.org/
- Started at CERN, LHCb
  - Used by a large number of high energy and astronomy experiments
- Systems include:
  - workload management
  - data management (ESCAPE uses RUCIO)
  - Job management API
  - accounting (provenance)
  - +++ …. Much more
- DIRAC Workload Management System (WMS)
  - uses novel approach of pilot jobs
  - allows for detailed job monitoring
  - can submit to a wide range of computing centers (HPC/HTC/Cloud)
  - able to create workflows: full automatization of multi-step workflow execution
- DIRAC is not just a WMS

# Applicable Requirements from the ESAP Project Plan

Many existing workload management systems are proven to be capable of dealing with this, which makes the development of ESAP in terms of connecting to different computer infrastructures easier. **ESAP will investigate the existing workload management systems and select one to interface with**. In this way, batch job submission will be handled by the selected workload management system that in turn connects to the underlying HTC, HPC or Cloud platforms.

- Requirement R5:
  - Users should be able to execute a job on a given dataset, including but not limited to: batch or real-time queries & pipelines, depending on the capabilities of the facility, which need to be made clear to the user.
    - Use Cases: U-4, U-5, U-7
    - Effort: 21

- Requirement R11:
  - Users should be able to select computing facilities on the basis of their capacity. E.g. She needs an HPC resource with a specific acceleration (GPU) because the software to be run requires it.
    - Use Cases: U-15, U-14, U-8
    - Effort: 13

- Requirement R13
  - Users should be able to schedule computational tasks at regular intervals e.g. to periodically retrieve new classification data from a Citizen Science experiment.
    - Use Cases: U-18
    - Effort: 13

# UI Service Components

- DIRAC part of ESAP
  - A) as a **workload management system** part of the Data Processing?
    - A.1 Batch Processing
    - A.2 Interactive Processing
  - B) as a **service** part of the offered Software and workflows?
    - DIRAC as a Service, e.g. for the simulation of data
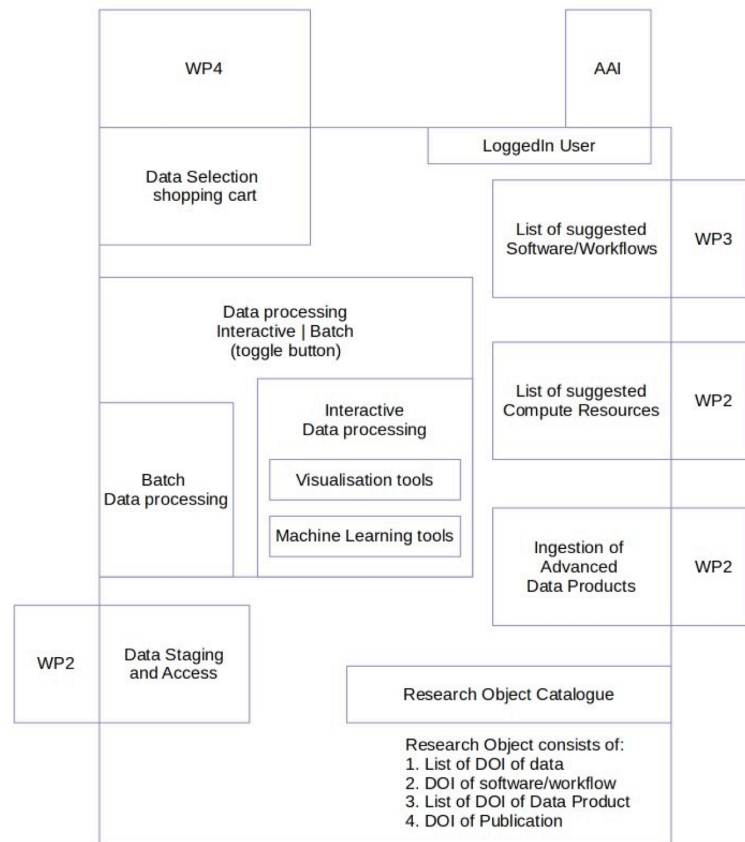
# Architectural Design

- Architectural design for Data Discovery and Data Staging

- Extension for data processing and federated computing, access to computing resources
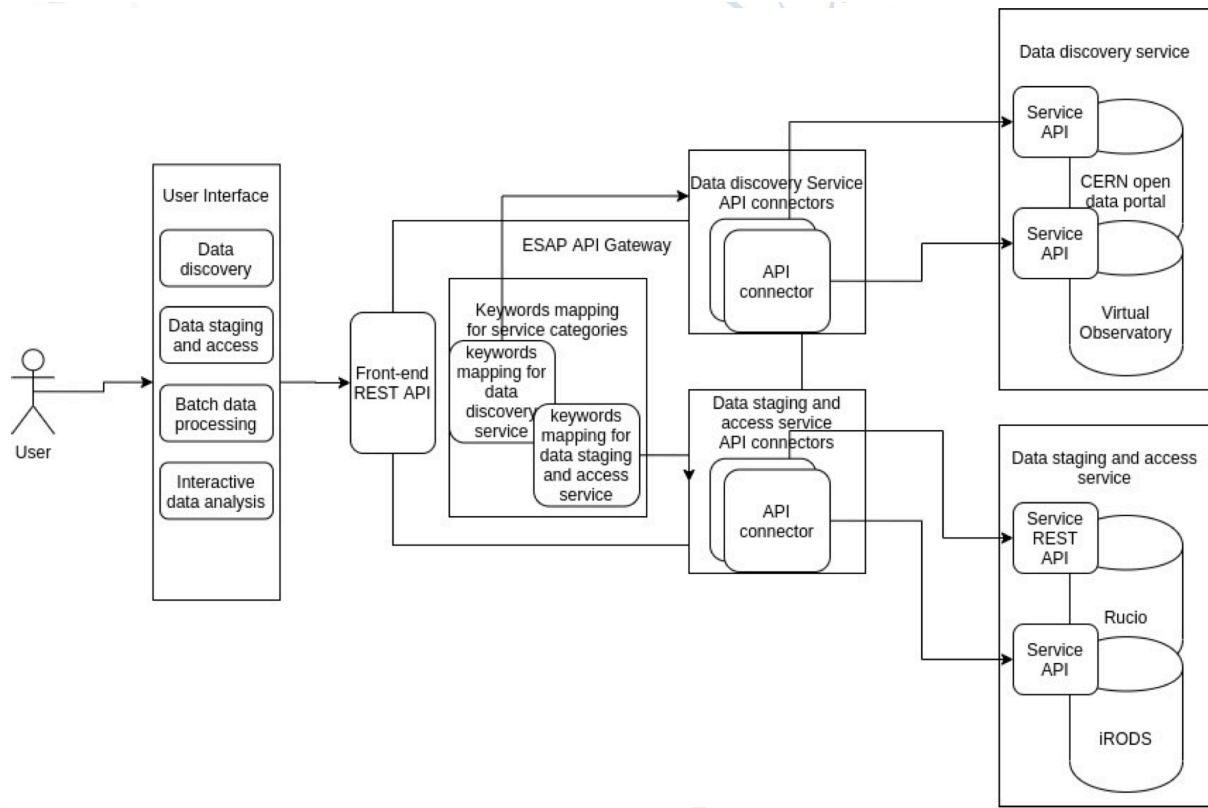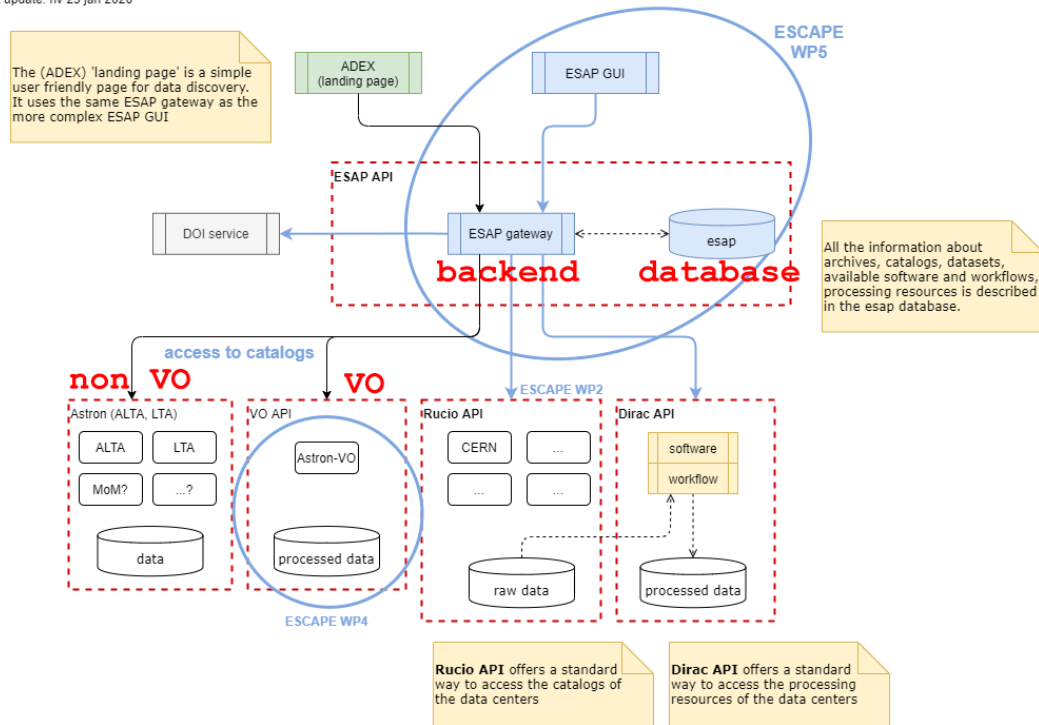
# API Integration

- DIRAC API integration
  - Standard way to access the processing resources of the data centers
  - Software and workflow
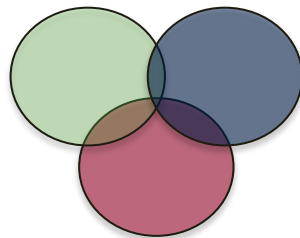- Interaction with Rucio
- Overlap with WP2 and with WP3

# Cross collaboration

- WP5 Interactive Data Analysis Group [Stelios Voutsinas et al.]
- WP5 Focus Group Batch [Sara Bertocco et al.]
- WP3 CORSIKA Focus Group 5 [Cristiano Bozza, Andrei Tsaregorodtsev et al]
  – Resulted in crossover WP meetings … ongoing



- CTA members who are DIRAC experts [Johan Bregeon & Luisa Arrabito]
  – Offer of consultancy help

# DIRAC as a Workload Management System

- Discussions with WP5 Interactive Analysis and Batch Processing
  - Workload management system or scheduler are part of these focus groups
  - These groups are currently looking in direct implementations (Jupyter to Rucio) or other schedulers (Rosetta)
  - Batch processing is currently working on concept (see later)
- Agreed to wait until concept is mature
  - In the meantime started to write a requirements document for a workload management system
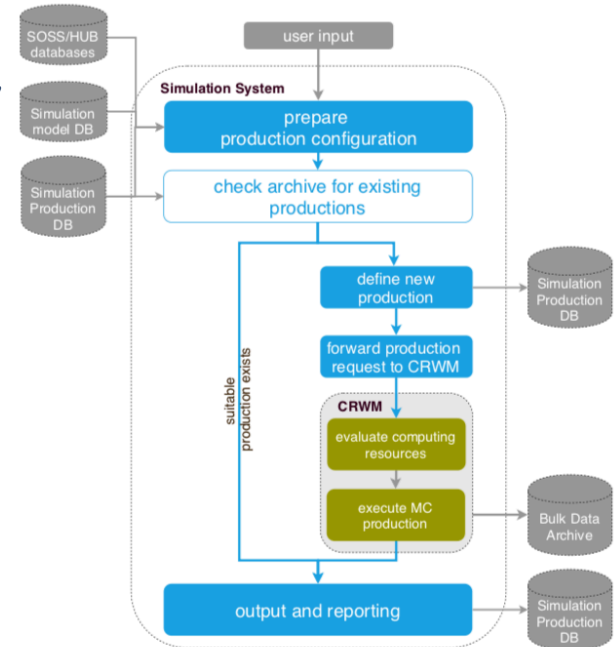
# DIRAC as a Service
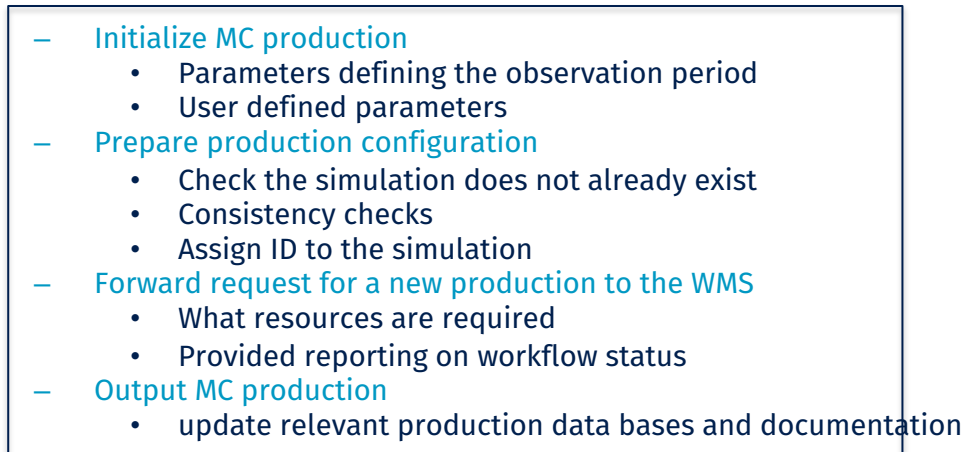
- Alternative is to offer DIRAC as a service from the platform to be able to run certain workflows on distributed computing
  - Self-contained service
  - Examples: simulation and sub-sequent analysis of data
- Overlap with WP3
  - See talk by Luigi and Andrei
- Agreed to first collect use cases from the ESFRIs
  - find commonalities
  - See talks by Jutta (KM3NeT) and Johan (CTA)

# Use Cases | ESAP & WMS

- Currently collecting use cases from experiments CTA, KM3NeT, SKA,

- Bare bones CTA **example**: IRF generation:

  - Initialize MC production
    - Parameters defining the observation period
    - User defined parameters
  - Prepare production configuration
    - Check the simulation does not already exist
    - Consistency checks
    - Assign ID to the simulation
  - Forward request for a new production to the WMS
    - What resources are required
    - Provided reporting on workflow status
  - Output MC production
    - update relevant production data bases and documentation

- Collection location: https://git.astron.nl/astron-sdc/esap-gui/-/wikis/Use-Case-for-CTA-DIRAC

# Discussion / Planning

- Step 1: Next steps
  - Finish collecting use cases
  - With WP3 integrate the DIRAC-CORSIKA container with the ESAP

- Step 2: Further integration and tech
  - Can we link the data from the shopping cart?
  - Do we need some testbed infrastructure (e.g. a server)?

- Step 3: Can we get some glue?
  - RUCIO
  - Interactive (jupyter etc …)

Difficulty

# DIRAC & ESAP | Open Questions & Issues

- CTADIRAC running on the grid: **need a certificates/tokens/IAM**
  - **For testing you can use a fake certificate**
  - **There is a run in local mode (submit jobs to local machine)**

- Glue needed: RUCIO-DIRAC integration – **nontrivial**
  - it is being worked on at Belle2
    - https://indico.cern.ch/event/756635/contributions/3391594/attachments/1844883/3026456/DIRAC-Rucio_Multi-VO.pdf

- Where would the jobs go? What is a job? (script/workflow/container)
  - Same issue for the Batch working group

- DIRAC can work in 'local mode'
  - Jobs would run on the server running the API – could be used a very basic test?

- Interactive submission (e.g. notebook) - also **nontrivial**
  - yet to be solved problem – more glue needed (containers?)

# Conclusion

- Focus group to discuss and define a role of DIRAC
  - Not pushing for DIRAC as a self-service, but trying to fill gap if there is one
- Discussions started within WP5 focus groups and with WP3
- Role of DIRAC yet to be defined
  - As workload management system
  - As service
- Next steps to develop in more detail the use cases
  - Discussions with WP2 (DIRAC-RUCIO integration) may be needed depending on the outcome
- DIRAC not part of minimum prototype, but might come with the extension