LOFAR processing software

Yan Grange, Vishambhar Nath Pandey

ASTRON

E-OSSR Onboarding Presentation

3 September 2021

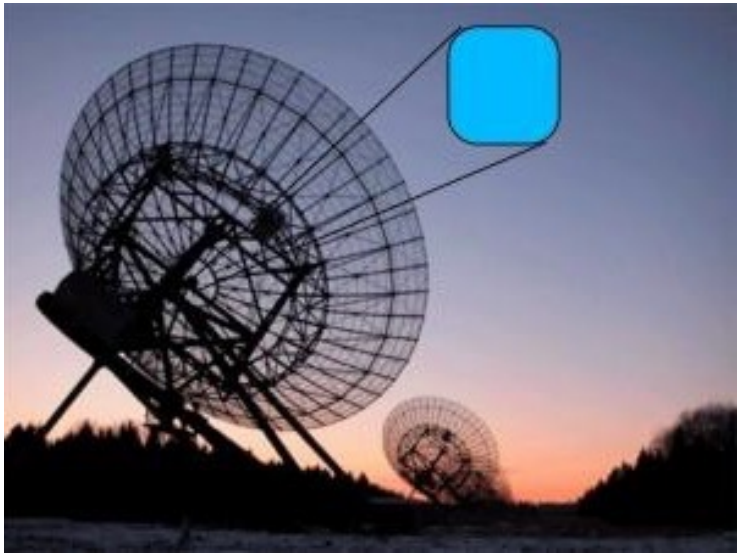
ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

ASTRON

“Making discoveries in radio astronomy happen”

2 main facilities operational right now



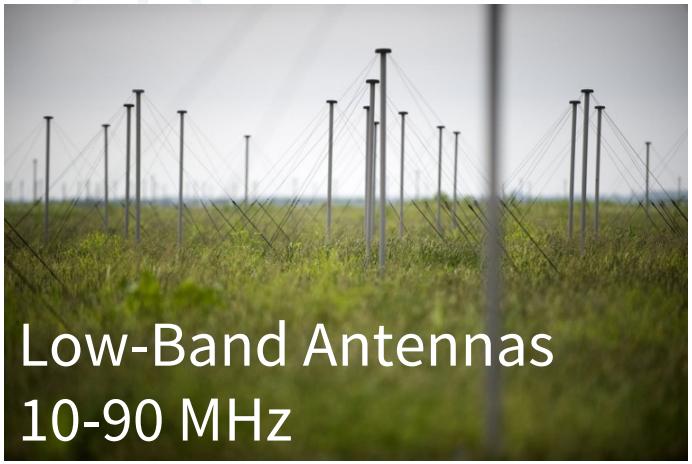
LOFAR

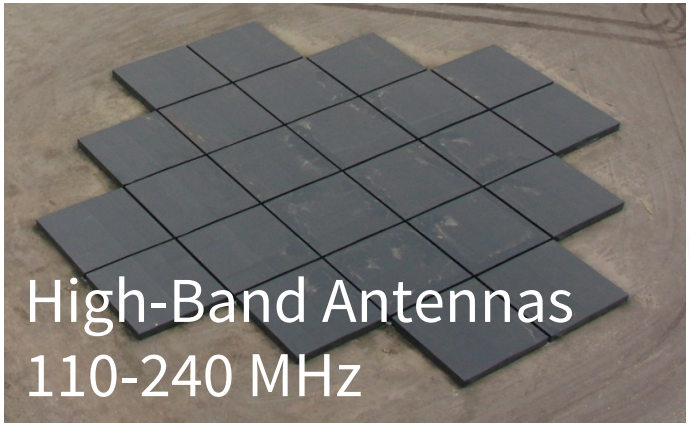
ASTRON
Netherlands Institute for Radio Astronomy

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

LOFAR - instrument

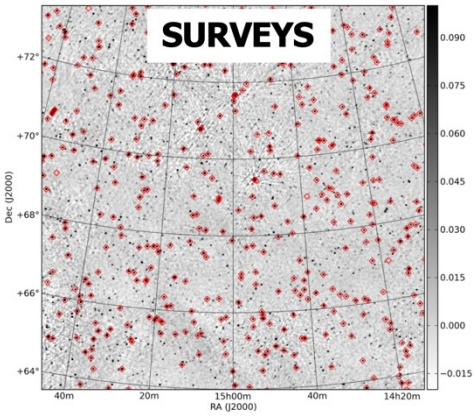
The International Lofar Telescope is the formal entity managing the telescope.
ASTRON is one of the partners and takes care of its day to day operations.

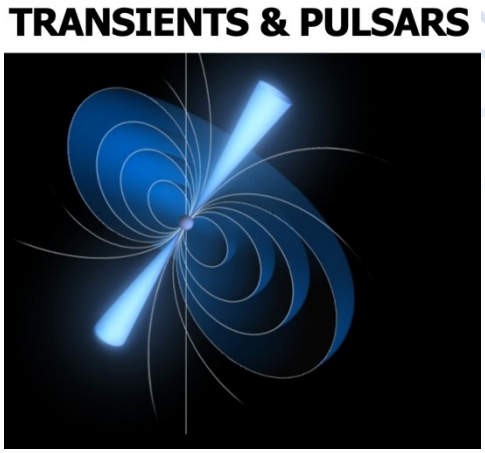


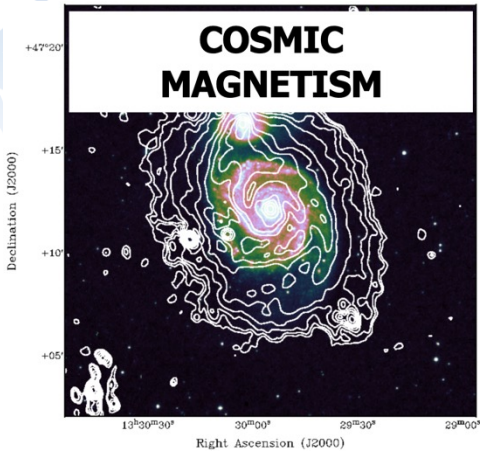


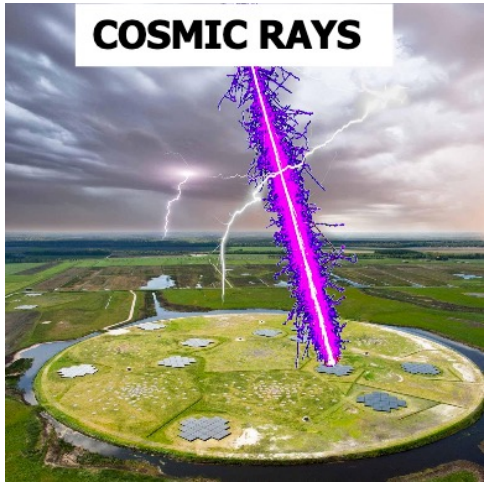
Funded by the European Union's
Horizon 2020 - Grant N° 824064

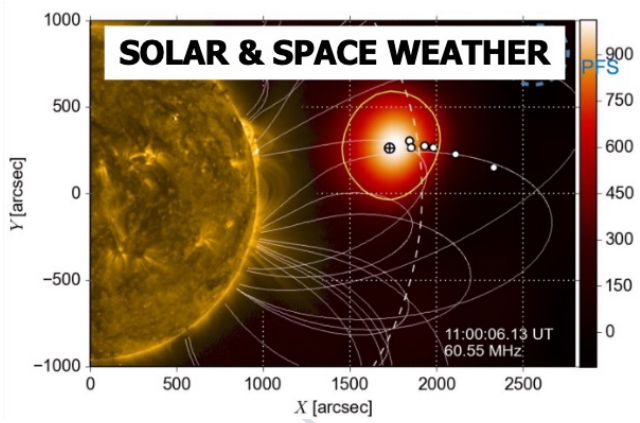
LOFAR - Science

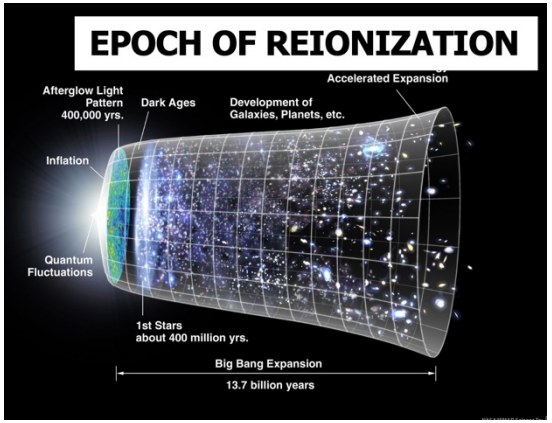












In essence: we can observe anything between the processes in our atmospere and the formation of the first stars…

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

LOFAR - software

- LOFAR is a ”software telescope” so ”the LOFAR software” can be anything of:
  - Station software (signal processing of antenna outputs)
  - Correlator software (combining outputs of all antennas)
  - Processing software (RFI, calibration, imaging, RM cubes, etc.) ← **This is what we talk about here!**
  - Long-term data archiving, data access for users, processing on cloud/grid (e.g. ESCAPE services)
- LOFAR has (currently) ~ 150 users across the globe.

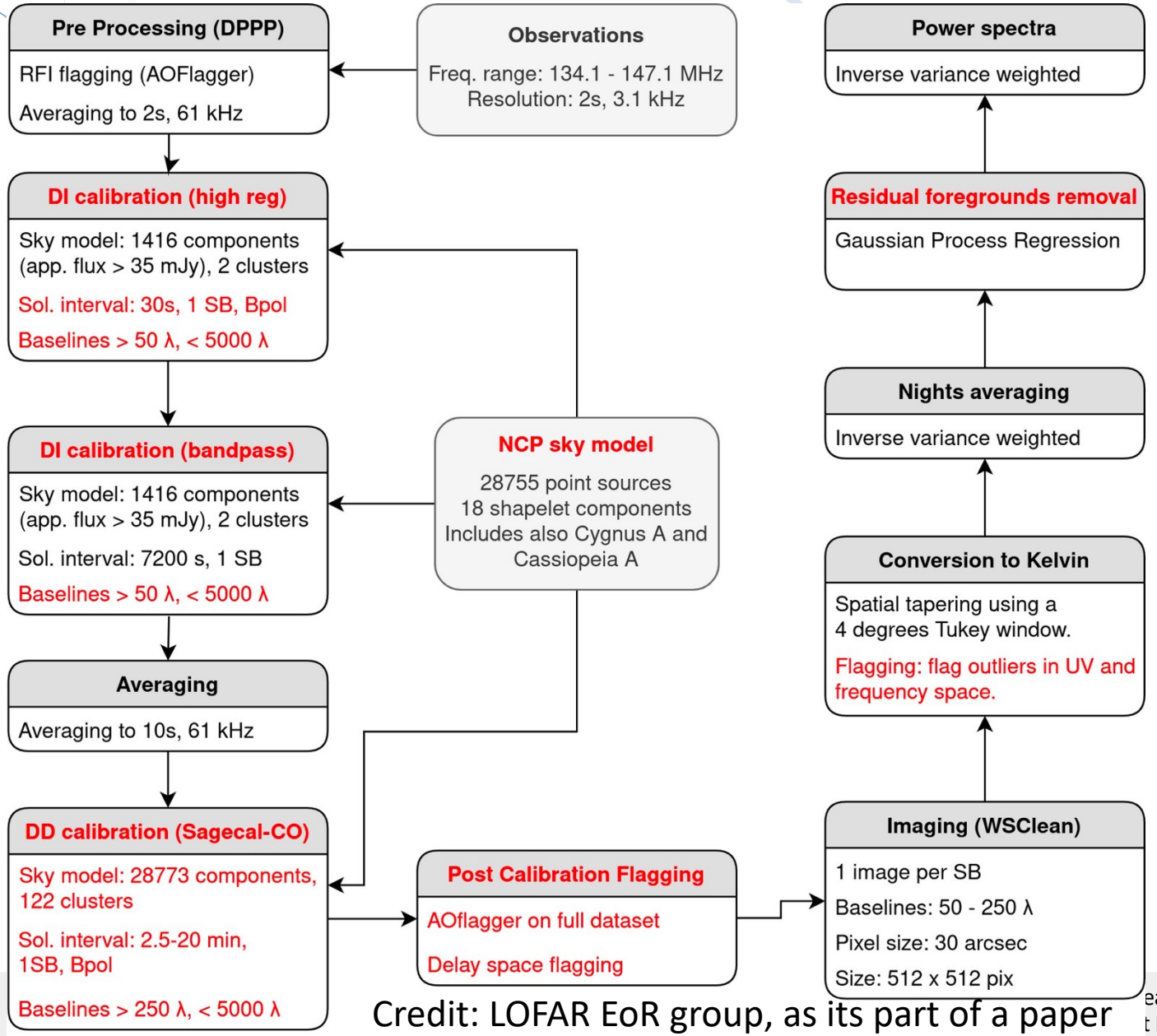
Funded by the European Union’s Horizon 2020 - Grant N° 824064

LOFAR – processing software

- Use case is imaging based on raw data:
  - RFI mitigation (mostly suppressing humans)
  - Calibration
  - Imaging
  - Source catalogs
- Workflow
  - Each step is a different module, which run one after the other.
  - Currently running using custom-built pipeline framework.
  - CWL implementation is on its way
- The package contains a bit more than only the “LOFAR processing software”. Mostly separate tools used in the pipeline for e.g. imaging (WSClean)

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

EOR processing pipeline



Credit: LOFAR EoR group, as its part of a paper

ean Union's
t N° 824064

Software – development process

- ‘We’ have been working on it for a while

```
commit 0a343975ad89ac435a205160a072b93dfdce7070
Author: (no author) <anonymous@astron.nl>
Date:   Thu Dec 21 14:33:22 2000 +0000
    New repository initialized by cvs2svn.
```

- Nowadays we use git.
- LOFAR software currently being developed in a SCRUM team (“with all roles and meetings”; 3wk sprints)
  - Two sprints per year for maintainance
  - Support, fix bugs, etc based on schema and wiki for documentation
  - CI/CD using Jenkins
  - Start using Docker more often in production
- Mostly supported on our local systems. 6 roll-out moments per year (not all necessarily used). Manual testing before roll out

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

Software – methods

- General guidelines: there is not much written down on how our workflow is but I've asked around.
  - Unit tests and integration tests on all new code. If old code is crucial, tests added
  - Code reviews, juniors activelt supervised
  - Coding styles: PEP8 for python. Not much more than that.
- Not very modular, mostly one big package. Work ongoing to make it more modular
- Licenses: ASTRON standardised on Apache 2.0, some external libraries are GPL (e.g. Casacore). LOFAR is also GPLv3 to make life easier.

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

Software - requirements

- Development mostly happens for Centos (some parts Ubuntu).
  - Since the software is packed in a container that should not be an issue
- For the pipelines, we use our custom framework
  - CWL version is being developed
- Data access using gridftp with X509 certificates, working on replacing that with webdav with macaroons.

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

Hardware requirements

- HW requirements are very much use case dependent.
  - Data sets can be rather big (~150GB per file. Often 244 files per observation).
  - Some parts need (Nvidia) GPUs to execute (in finite time).
- Processing is very often embarassingly parallel (each file is a wave band).
  - However after each wave band is calibrated, the solutions are smoothened/averaged to prevend weird jumps. So some communication between nodes is needed.
  - Smaller runs (using a subset of data, trying out calibration parameters, etc) can be done on a single node.
  - Full imaging runs could (and pronbably: should) run on a cluster.
- Some, more complex pipelines do hoever highly rely on MPI

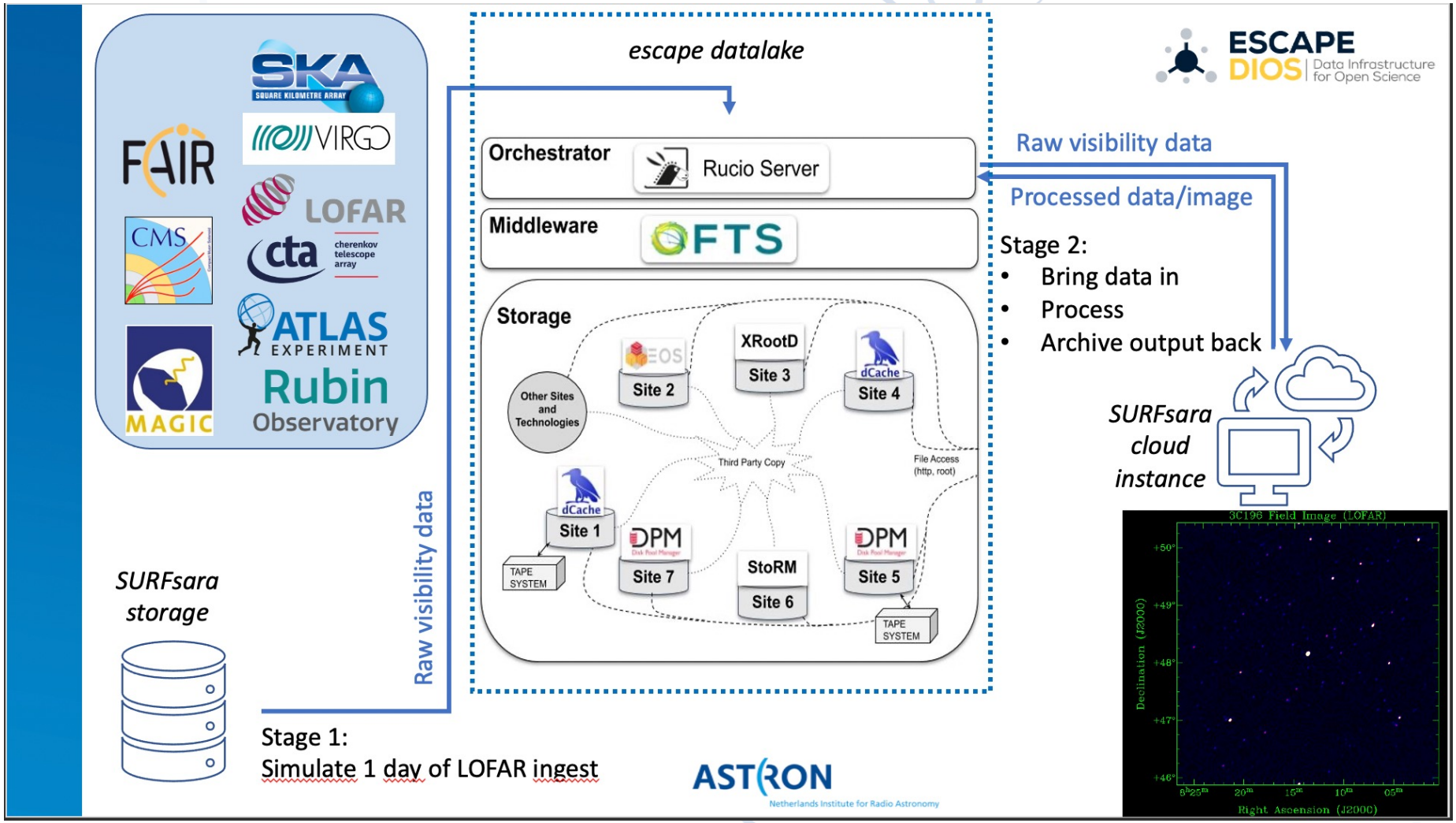
Funded by the European Union's Horizon 2020 - Grant N° 824064

OSSR Integration

- Source code, container-based images to be onboarded.
  - We haven't yet defined a test work flow. If the goal is mainly to show that all components work a limited data set and pipeline configuration can be added to the package.
- User story
  - As a LOFAR user I want to create an image based on raw data in the LOFAR Long-Term Archive so that I can use it to do science.
  - Access to LTA data, processing raw data to image and making a source list from that image

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

Time for a short demo (~10 min)



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

Time for a short demo (~10 min)

- (did a live demo)

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

Open Points and Discussion Time

- Would be nice to have it integrated in a Jupyter notebook environment that plays well with WP2, WP5, WP4.
  - Data input from ESAP shopping basket, through Rucio and VO tooling.
  - What is the WP3 vision on how to handle supporting software in entries? (i.e. hierarchy of notebooks)
- Platform integration and metadata => I guess we want to discuss that. I have little opinion on that…

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