20211005 - DM TSP - Introduction/kick-off meeting

Caterina Doglioni - Introduction to Science Project

See slides: 20211003_DM-TSP_Intro.pdf

Ian Bird - Interaction with ESCAPE tools

Slides: ESCAPE-SPs-v1.pdf

Notes on slides:

Who is involved in EOSC-Future/what it is:
- ESCAPE = 1 of 5 clusters from different sides of science
- E-infrastructures (e.g. EGI, EUDAT, OpenAIRE)

Working together to build the European Open Science Cloud

Our goal (ESCAPE/Test Science Project): integrate what we are doing in the EOSC Cloud

What is also going on in parallel:
- High Performance Computing (we want to access them / use them)
- INFRAEOSC-07 projects also funded to provide resources in different ways. E.g. commercial procurement of cloud resources, extensions of e-infrastructures...we can use those to get resources

Everything needs to work together in EOSC. This is a broad task, but we know what we are going to do and we can split the task in smaller pieces.

(Bigger) pieces of EOSC-Future:
- EOSC Core: key functions, what everyone will be using (e.g. common authentication, common policies, helpdesk…)
- EOSC-Exchange: "horizontal services" like HPC environments, computing resources…
    - Community & Clusters sit on top of this
    - We would like to use some services & contribute to the horizontal services if we have something useful
- EOSC-Interoperability
- [others as well]

How do we (DM TSP) fit in WP6, and how do we benefit from EOSC-Future?
- In WP6 in EOSC-Future we have 6 postdocs for "science funding" (shared between two TSPs)
- Also consolidation activities

Our goal for the next 2 years

Integrate what we do as much as possible - do the best we can, it doesn't have to be perfect or 100% done, but the fundamental layer should be working. All of our tools should be using the common authentication interface (but this is what we're doing anyway in all community). There will be also tools that we have that will be useful for the exchange layers.

From the ESCAPE side, we want to demonstrate that the tools work → Test Science Projects are useful.
From the ESFRI side (e.g. CERN, experiments) we want to have science done with analysis tools that work in the same environment.
- RECAST/REANA are specific so far, but eventually could be used more

TODO: make a list of requirements and milestones, see end of minutes.

Future meetings

ESCAPE will have a TSP meeting soon, probably on the last week of October 18th.

Individual plans for partners and postdocs

Jared Little + supervisors - LHC/ATLAS plans

DM TSP postdoc Jared Little, working with Arturo Sanchez Pineda, Stephane Jezequel, Tanya Berger-Hryn'ova, Caterina Doglioni + her team

Goal: Make ATLAS DM analyses code/results available in ESCAPE SW catalogue to reinterpretation & comparison with other experiments

- ATLAS dilepton resonant analysis - 1st test-case
    - Make dilepton RECAST available in ESCAPE SW catalogue for reinterpretation [might need ATLAS reconstructed signal MC samples as input]
    - Make dedicated truth-based code available in ESCAPE SW catalogue to reinterpret existing ATLAS fiducial xsection limits [input truth signal MC samples]
    - Decide on the benchmark DM MC samples to use
- ATLAS monojet analysis- 2nd test-case
- Extend this approach to other ATLAS DM analyses.

High-level milestones:
- Putting the data in the data lake (embargoed data is fine, only visible to people who have access)
- Putting the code into the software catalogue

Potential first milestone: take advantage of the dilepton analysis which can be used as an exercise for Software Catalogue. This has RECAST implementation that does not need data → could just use an open signal MC sample.

Misha Smirnov + supervisors - KM3NeT plans

Misha is working on both Test Science Projects, supervised by Kay Graf (KG), started 01/10/2021

In KM3NeT we don't have a complete reproducible pipelines yet - we want to negotiate that with the collaboration first, what/how to make public → we can probably do the milestones after that

- Plans:
  - coordination of KM3NeT contributions (CNRS-CPPM, INFN, NWO-Nikhef and FAU) - total 18 PM
  - harden KM3NeT pipelines for DM Instrument Response Functions (IRFs) and sensitivities (fast response, reproducibility, FAIRness)
    - start with standard DM searches and simulated data
    - extend to common analyses (cross-experiments)
      - Q CD: what do you mean with cross-experiment analysis?
      - A KG: work on many different analyses at once. E.g. including systematics in a common way. Targeting specific source/particle/mass range, but if you want to go deeper in the future for common analyses with events and instruments at the same time and including systematics as priors we need a discussion. Common data format and systematics in a coherent way would be good.
      - Comment CD: we can join forces with discussions happening in other Eu/US activities, e.g. iDMEu (https://indico.cern.ch/category/12787/)
  - orchestration of processing and results with the ESCAPE services

Regarding requirements/milestones: start with one analysis, then add more analyses. We need to see if we want to produce the whole chain from the simulation, otherwise resources are smaller.

An intermediate milestone could be: do analysis at the higher level possible (closer to the final results than to the detector) as that is simpler. We can demonstrate progress with that first. Those are things that are achievable, and we can make statements about these in a shorter timescale.

Elena Gazzarrini + Supervisors - CERN plans

We have data lake challenge in the next 2 months. In parallel start talking to REANA/RECAST friends. Data injection in the next month could be an opportunity.

Strategy questions:
- after ESCAPE finishes (end of July, with unfunded extension), who will provide resources?
- Data lake / Rucio instance in CERN, could this be prototyped outside CERN?

Organization: how is the coordination with WP6 happening?
Ian: raised at the cluster meeting last week, director of EOSC-Future is aware that we aren't really talking. At the moment it's not an issue because we all know what we need to do, and at the technical level we can manage this.

General discussion

Q KG: What kind of timescales do we need for our own internal milestones?
A Ian: No pressure from WP6, not very formal. Just tying things over the google spreadsheet will be enough. Each subproject could have a small text file that we can link to the spreadsheet lines. What we're doing on month 6, 18, 30.

Q CD: What is the status of data lake / software catalogues

Xavier Espinal: we have a working prototype of the data lake - had about 1 year to leverage this infrastructure, and we should test it already.

KG: OSSR - we have an onboarding process which is open for everyone, last Friday we had an onboarding talk for ATLAS software. We can already try some things, we have a sandbox to play with this procedure - Arturos is involved. We can give a presentation at the next meeting.

Notes from Kay:
- Here is the landing platform of OSSR: purl.org/escape/ossr
Here is the overview of the onboarding status:
https://project.escape2020.de/projects/ossr-onboarding

TODO list for all ESCAPE partners:

1) Quantify capacity requirements for each sub-projects. We could get the resources, "contributions" from the partners (for example CERN or other institutes could give us N TB of storage) and from cloud / HPC resources → we need to tell them what we need. See preliminary estimates in the backup slides of Caterina's talk - do they make sense?

2) We should work on a milestones table (attached to the agenda and here: Draft TSP timeline & milestones). High-level milestones (6/18/30) "we need to do A, B,

C...". Examples: on month 18 we should have a test prototype using service A and some level of workflows that can be demonstrated with some practical science case.

For next month's meeting:

- Pooja Bhattacherjee + supervisors - LAPTh plans
- Discussion on individual milestones and requirements
- Xavier Espinal - readiness of Data Lake and data challenge
- [December meeting] Kay Graf - Tutorial on OSSR