ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

The Dark Matter Test Science Project

Caterina Doglioni - Lund University

Input from: Tanya Hryn'ova, Stephane Jezequel, Simone Campana, Ian Bird, Xavier Espinal, Kay Graf (+ KM3Net), Valerio Ippolito, Francesca Calore, Pasquale Serpico, Sam Meehan, Lukas Heinrich, Stephen Serjeant, + many others

@CatDogLund, she/her

http://www.hep.lu.se/staff/doglioni/

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.

First of all, welcome to the new postdocs!

CERN: **Elena Gazzarrini**

Working with: Simone Campana, Xavier Espinal, Ian Bird

ATLAS @ LHC / LAPP: **Jared Little**

Working with: Tanya Hryn'ova, Stephane Jezequel, Caterina Doglioni

[2 more coming later…]

Gamma rays / LAPTh: **Pooja Bhattacharjee**

Working with: Francesca Calore, Pasquale Serpico, Christopher Eckner

Neutrino / FAU: Name to be revealed soon…

Working with: Kay Graf & al.

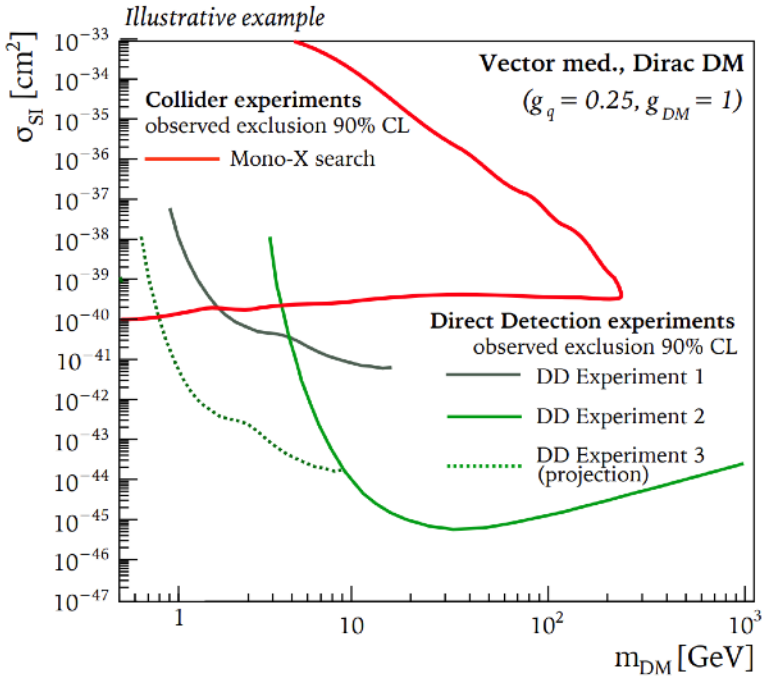
DARKSIDE / INFN: Will be recruited soon

Working with: Valerio Ippolito, Tommaso Boccali

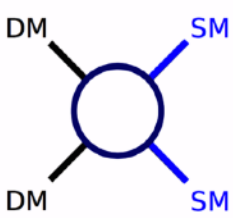
Scientific question: dark matter

Many hypotheses for dark matter
→ many ways to detect it
→ many different experiments
→ many different data / workflow needs
→ many different data / result sharing policies

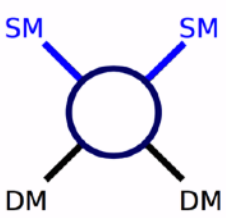
Scientific content of DM-TSP:
New plots of dark matter discoveries / constraints



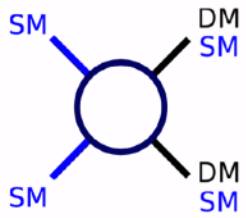
one of many models predicting Weakly Interacting Massive Particles (WIMP) - could also use others...



Indirect Detection

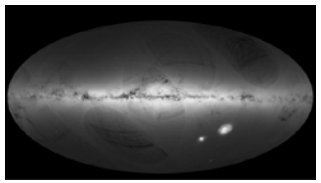


Direct Detection

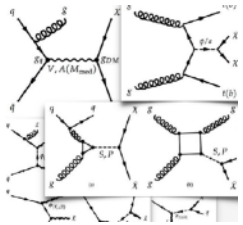


Colliders

Credit: ESA/Gaia/DPAC.



Astrophysics

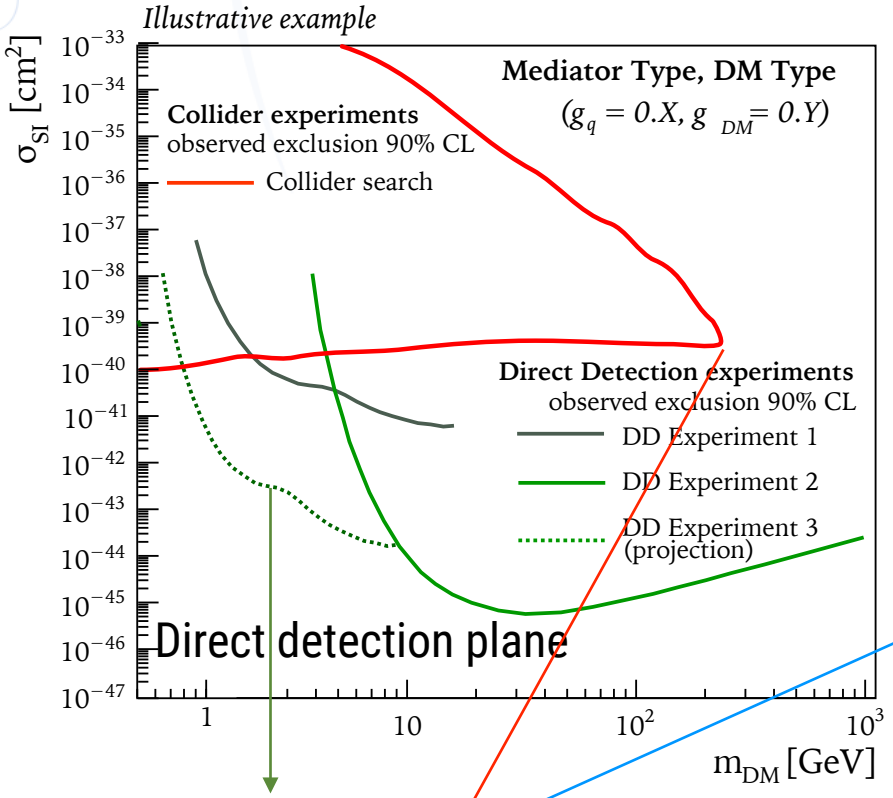


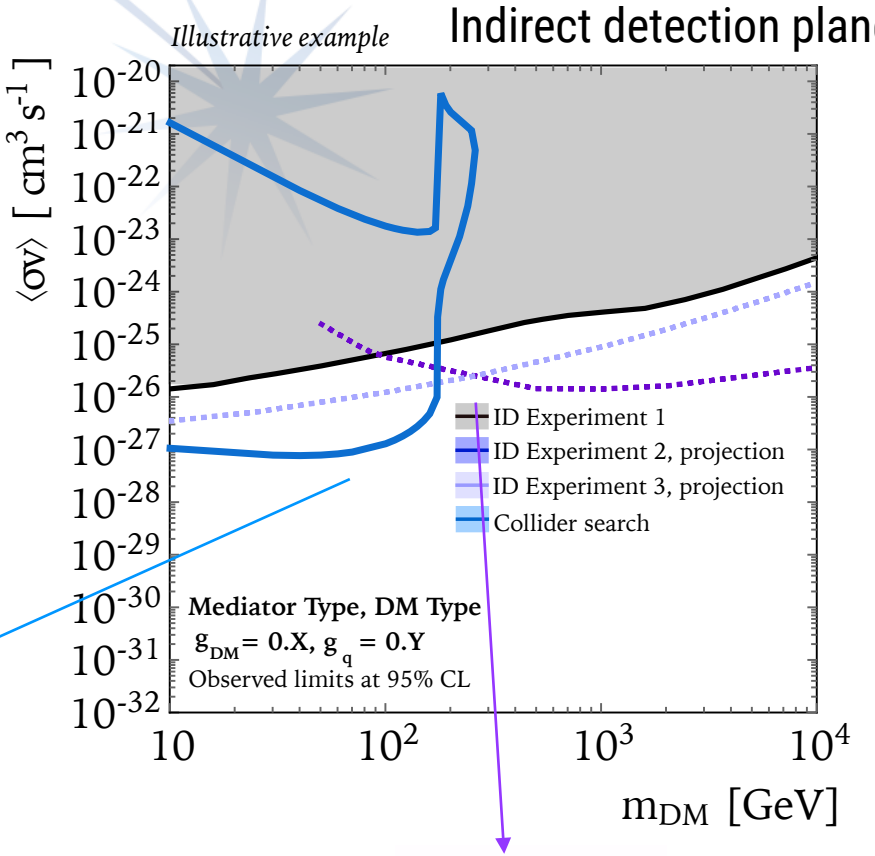
Theory

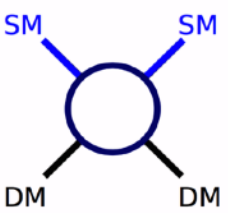
[arXiv:1912.12739 & refs therein](arXiv:1912.12739)

- There are many combinations/comparisons of results on the market...
  but none that sees them all work together with FAIR data & end-to-end workflows!
- This is where our Test Science Project comes in!

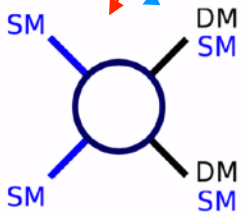
Planned science outputs (to begin with)



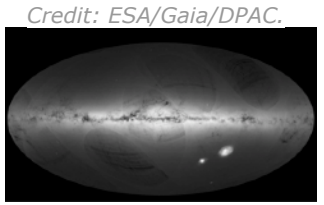




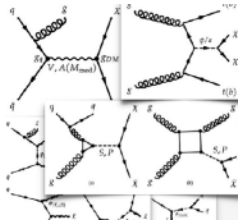
Direct Detection



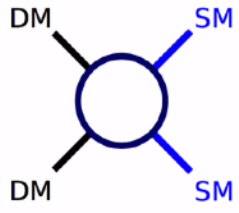
Colliders



Astrophysics



Theory



Indirect Detection

Necessary for both interpretations

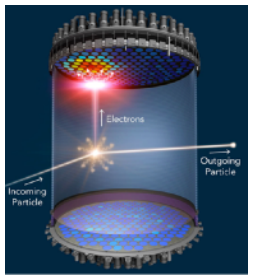
n Union's
l° 824064

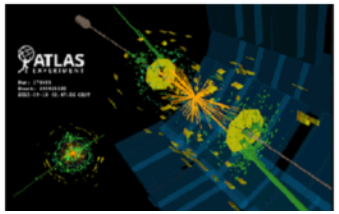
Analysis workflows for DM-TSP (feedback?)

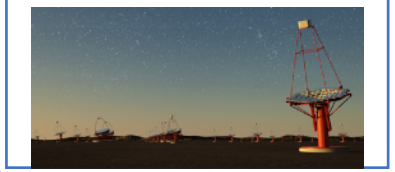
Integration work, foreseen in EOSC-Future

Generation & simulation of events

Experimental data







Credit: SLAC/LZ/ATLAS/CTA

Data sharing and data processing challenges

Data **processing** (including reconstruction & calibration if possible)

Data analysis, preservation and interpretation challenges

Analysis of events/ distributions (including background subtraction, background estimation, statistical analysis)

Interpretation of results

Combination of results with other searches/ experiments

Comparison of results with other searches / experiments

Consolidation work in EOSC-Future and ESCAPE

Data Lake

Software Catalogue
Analysis Platform

Meta-Workflow for DM sub-projects: Direct detection

Discussions on improving interpretations/plots:
iDMEu (JENAA EoI), existing working groups

Theory input & code [LAPP/LAPTH]

Data Lake

ATLAS/CMS [LAPP+Lund/INFN]

CERN

LHC data

recast

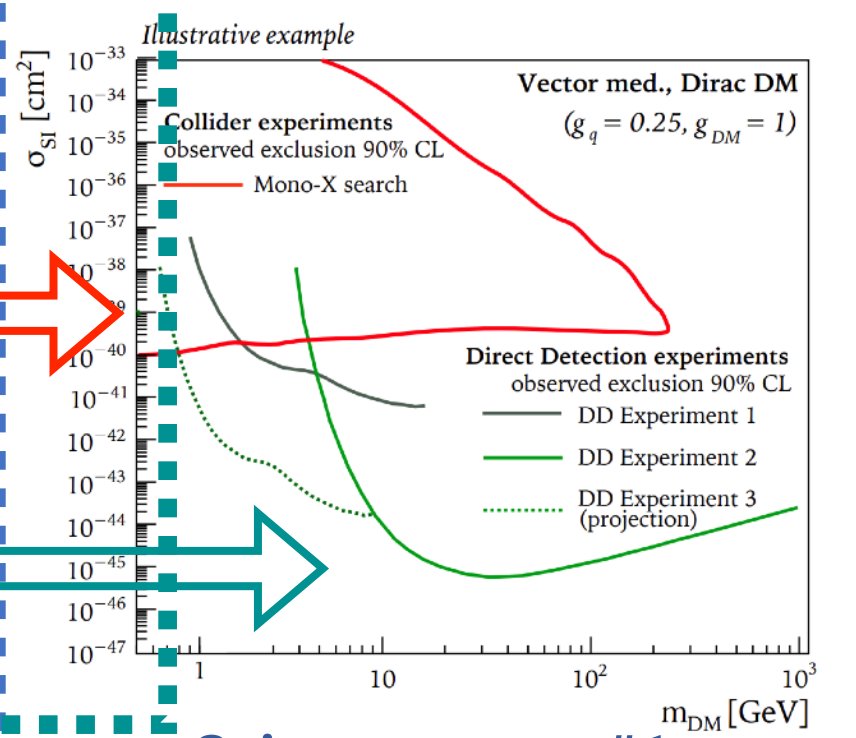
reana

Direct Detection data

Discussions on using common steering (declarative?) analysis code

HSF

Darkside [INFN]



Discussions on standardized data formats for other experiments (iDMEu)

Software Catalogue

Science output #1

Cross-Science Project innovative algorithms (e.g. ML)

Meta-Workflow for DM sub-projects: Indirect detection

Discussions on improving interpretations/plots:
iDMEu (JENAA EoI), PhyStatDM, existing working groups

Data Lake

LHC data

Indirect Detection data

ATLAS/CMS [LAPP+Lund/INFN]

CERN

recast

reana

HSF

Discussions on using common steering (declarative?) analysis code

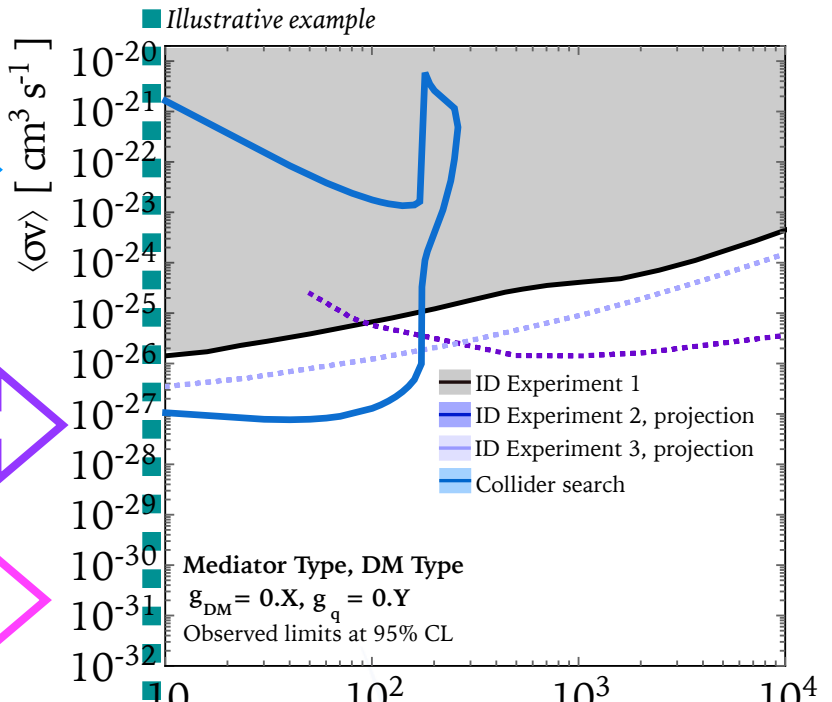
CTA [LAPP/LAPTH]

KM3NeT [FAU]

Theory input & automated codes for interpretation [LAPTH]

Discussions on standardized data formats for other experiments (iDMEu)

Software Catalogue



Science output #2

Cross-Science Project innovative algorithms (e.g. ML)

Rough, preliminary timescale for DM Science Project

Months 1-6 (April-October 2021) - first phase completed

- Organise recruitment, define datasets, resources and algorithms
  - For draft of datasets, resources and algorithms: see backup slides

Months 6-12 (October 2021-April 2022) - where we are now

- CERN postdoc tests REANA elements with an existing RECAST implementation (probably a collider search) using existing resources → will need some discussion with experiments
  - Subsequently, REANA interfaced with Data Lake tested as ESCAPE challenge
  - Can serve as example for others who wish to use the same structure
- Other subproject postdocs get trained on and make progress in *data analysis*
  - Regular (monthly) meetings with a round table on progress
  - We'll hear more about the plans in the round table today
  - Note: there is room also for more ideas/projects if you know of anyone who would be interested!
- Discussion + presentation results integrated within the Snowmass project
  - How many people would be interested in joining a "dark matter complementarity report"?
  - Concrete idea: use a Snowmass whitepaper from this group as progress report to be delivered in March 2022, can also be submitted to a journal

More on how this works together with deployment of ESCAPE tools in Ian's contribution

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

Details on Test Science Sub-projects

Coming after Ian's talk: round-table - more details are in this Google Document

Subproject 1 [Infrastructure and support + colliders]
Partner: CERN
Title: *Enabling dark matter science on the data lake*
PIs: Xavier Espinal (CERN), Simone Campana (CERN), Ian Bird (LAPP)

Subproject 2 [Indirect Detection]
Partner: KM3NeT - CNRS-CPPM, FAU, INFN, NWO-Nikhef
Title: *Determination of KM3NeT Sensitivity to Dark Matter via Open-Science Tools*
PIs: Kay Graf (FAU) **Partner PIs:** Cristiano Bozza (INFN), Pascal Coyle (CNRS-CPPM), Aart Heijboer (NWO-Nikhef)

Subproject 3 [Theory tools, Indirect Detection]
Partner: LAPP
Title: *Indirect dark matter search with gamma rays via open-science tools*
PIs: Francesca Calore (CNRS, LAPTh), Christopher Eckner (CNRS, LAPTh), Pasquale Serpico (CNRS, LAPTh)

Subproject 4 [Colliders]
Partner: LAPP + Lund
Title: *Reproducible ATLAS dark matter searches for visible and invisible particles*
PIs: Tanya Hryn'ova, Stephane Jezequel, Giovanni Lamanna **Partner PIs:** Caterina Doglioni (w/LU resources)

Subproject 5 [Direct Detection]
Partner: INFN
Title: *Opening DARKSIDE and collider experiment data and software towards dark matter discoveries*
PIs: Valerio Ippolito, Tommaso Boccali **Partner PIs:** Shahram Rahatlou (CMS -> colliders)

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

ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

Backup slides

@CatDogLund, she/her
http://www.hep.lu.se/staff/doglioni/

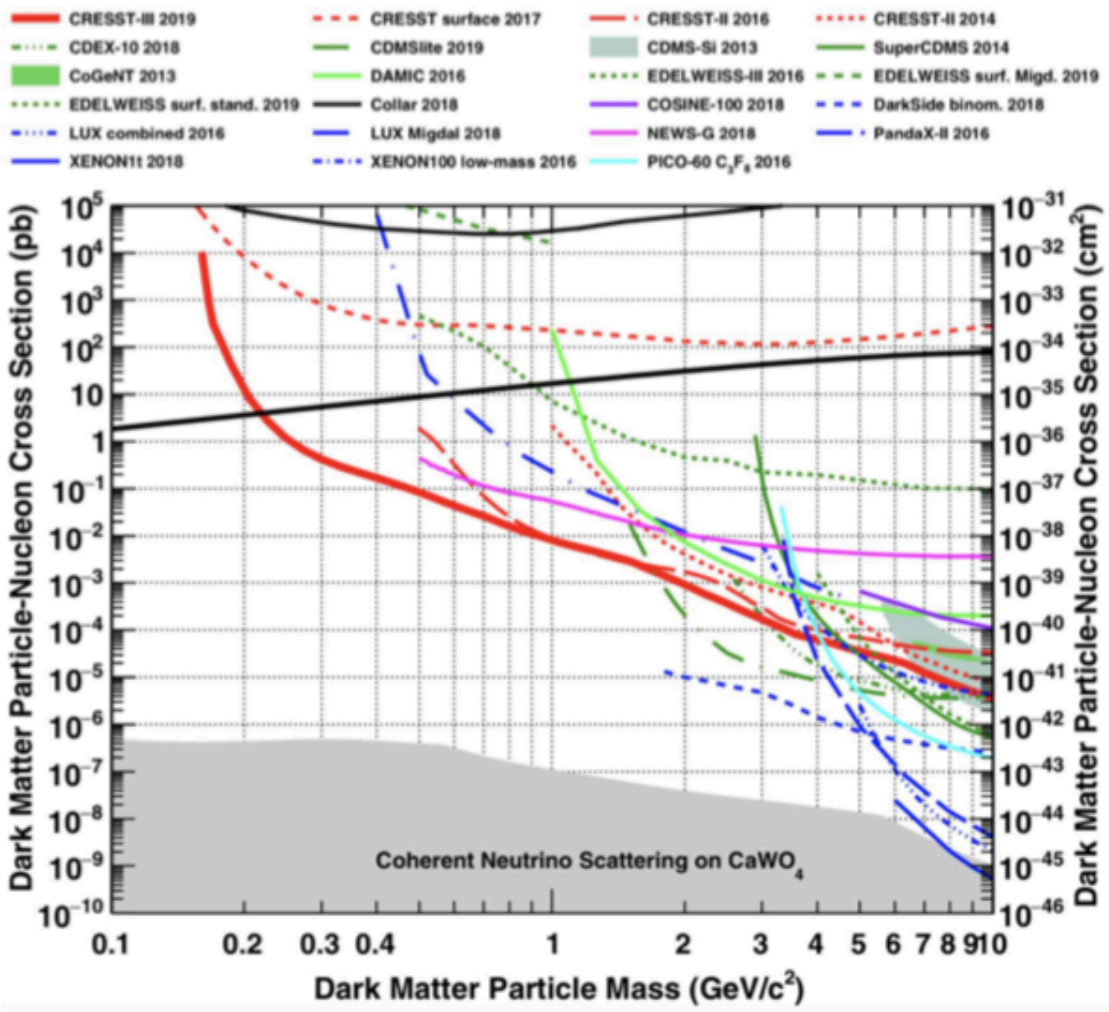
European Research Council
Established by the European Commission

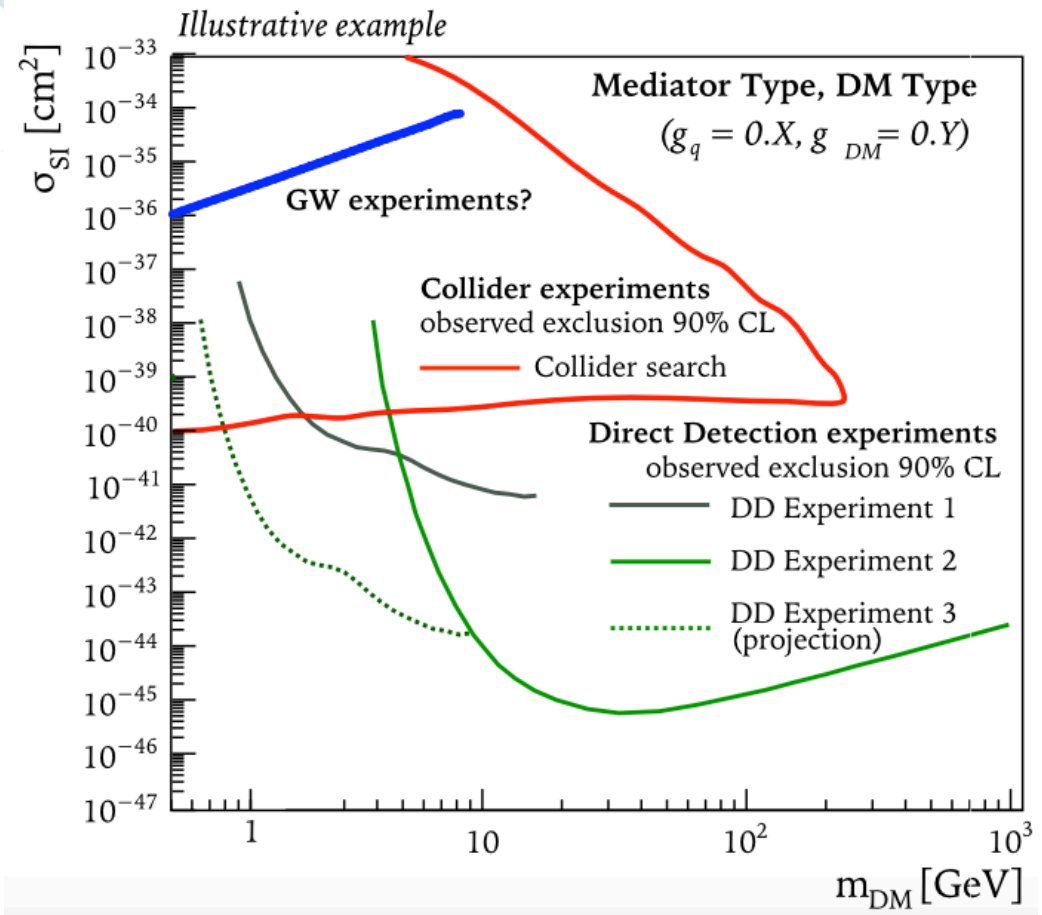
ATLAS EXPERIMENT

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.

Connection to gravitational waves / extreme universe

Example from: arXiv:1909.00654
More implications in arXiv:1907.10610





Knowledge exchange if we want to make this kind of plots is **essential**!

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

What to expect from ESCAPE / EOSC-Future

Global ESCAPE Science Project Document by Ian Bird

1. AAI (WP2): A fully developed AAI (*identification/authentication*) solution following the AARC blueprint is fundamental. In EOSC-Future we must ensure that the ESCAPE solution is fully interoperable with EOSC. Scientists in the TSP's should be using a single user identity for all aspects of work.

2. (ESFRIs, WP4): Publication of data sets into the Data Lake - required from the TSP partners of the ESFRIs and WP4.

3. Data Lake (WP2): federated storage services should be made available to the TSPs, allowing all of the data sets needed to be openly accessible to all participants, except for specific cases where datasets are embargoed but still can be used by TSPs

4. ESCAPE (WP3) software catalogue should publish all of the needed analysis components, and make them available for the various groups involved in the TSP work.

5. (WP5) An analysis environment, with a Jupyter notebook deployment, and access to scalable compute resources behind. *DM Science Project "workhorse" (to be tested and discussed): REANA*

6. Virtual Research Environment for each of the TSP as the outcome of the integration of the above together with publication services (WP3) for the scientific results and outputs of the work.

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

Resources: data lake needs

- Indirect detection (KM3NeT) [**Subproject 2**]
  - Embargoed simulated data:
    - Low-level data products: dedicated MC simulations to generate Instrument Response Functions: typically 50 core years, 50 TB of storage
    - High-level data products: <1 TB of storage (processed data and MC)

- Indirect detection (Dwarf galaxies constraints) [**Subproject 3**]
  - Data production done at the experiment level →moderate space needed for data products at high level, < 1 TB.

- Colliders (ATLAS for now) [**Subproject 1** (initially, then extend to others), **Subproject 4**]
  - Embargoed data:
    - Datasets needed, reduced-information file format: ~2 PB (data) + ~2 PB (simulation) [1] [2]
  - ATLAS Open Data: currently 1/10 of that but not yet ready for analysis
    - By the end of the project, ATLAS will release ~20% of its dataset as Open Data. [3]

- Direct detection (DARKSIDE) [**Subproject 5**]
  - Embargoed data and simulation: 1 TB SSD

- General summary plots [**Overall science outputs**]
  - Minimal data needs (only histograms / curves)

Resources: software for OSSR

- Indirect detection (KM3NeT) [**Subproject 2**]
  - MC analysis pipeline (part of ESCAPE, part KM3NeT-TSP post-doc)
  - Combination of Instrument Response Functions (possible combined effort)

- Indirect detection (Dwarf galaxies constraints) [**Subproject 3**]
  - The software we will produce needs to be stored in the software catalogue, in particular the code, input, and output of the Glory Duck project which aims at combining results from five major gamma experiments.

- Colliders (ATLAS for now) [**Subproject 1** (initially, then extend to others), **Subproject 4**]
  - RECAST/REANA + packaged analysis code to be included in the Software Catalogue
  - Prototypes of machine algorithms software for reconstruction and analysis

- Direct detection (DARKSIDE) [**Subproject 5**]
  - Reconstruction and analysis code (including statistical analysis)

- General summary plots [**Overall science outputs**]
  - GitHub repository with code and instructions to make summary plots starting from experimental curves (standalone macros using ROOT or Matplotlib, ideally can made it to work as Binder)

Resources: analysis platform

- Colliders (ATLAS for now) [**Subproject 1** (initially, then extend to others), **Subproject 4**]
  - An instance of REANA+RECAST operating on the Analysis Platform (part of CERN postdoc)
    - This will need authentication and accounting of time usage, as well as helpdesk(s) for troubleshooting
- Indirect detection (KM3NeT) [**Subproject 2**]
  - Platform to run MC analysis pipeline/combination
- Direct detection (DARKSIDE) [**Subproject 5**]
  - Continuous Integration-like service running on the Analysis Platform
- Indirect detection (Dwarf galaxies constraints) [**Subproject 3**]
  - The code to run will be a mix of C++ and Python code. Python code could be ran from a Jupyter Notebook.
  - The C++ code (eg, glike) could be ran either from a user machine, or from the analysis platform if this is possible.
- General summary plots [**Overall science outputs**]
  - A server running Jupyter notebooks with ROOT and MatPlotLib for overlays of curves in the final summary plots.
  - Ideally this would run on a Binder-like service so that users can contribute to the code after testing it in practice

Recognition of software products

[Lukas Heinrich's talk @ ESCAPE WOSSL 2020]

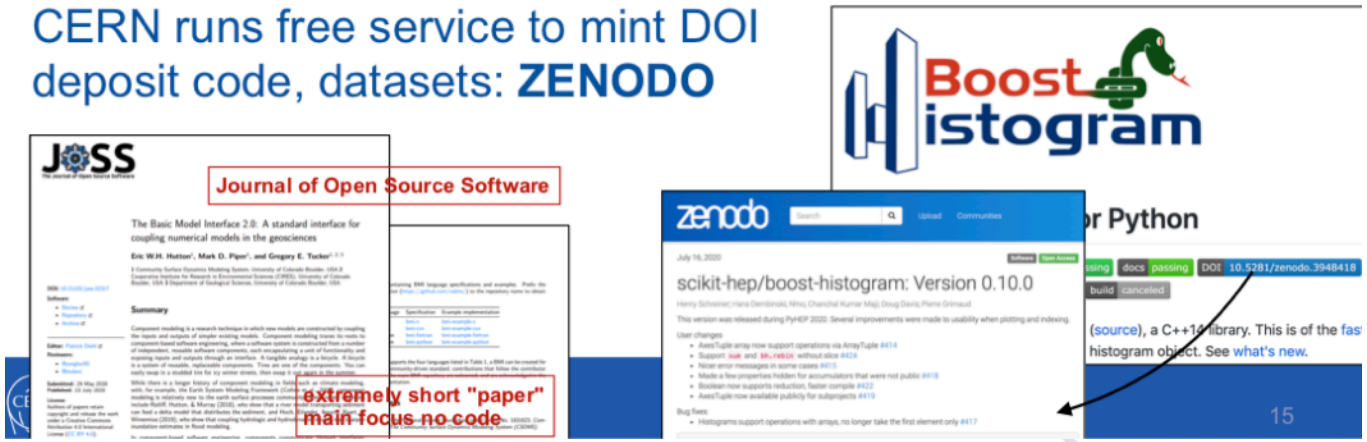
Software Citation:

`CITATION`

Software is often the research product itself. Should be treated as part of the scholarly record.

- cite software directly instead of "software papers" to attirbute proper credit
- if you need a paper consider **JOSS**

CERN runs free service to mint DOI deposit code, datasets: **ZENODO**



Initial ideas: papers **with** code

- Code: Zenodo
  - Future thought: will the Virtual Analysis Platform provide an interface to Zenodo a la Binder?
- Journal of Open Source Software
- Frontiers “Big Data and AI”
- Need discussion with collaborations/ESCAPE: how to credit software curators in large collaborations?

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

Expanding the Dark Matter Science Project

Initial effort focused on first 5 sub-projects *(+1 nuclear physics?)*

- Goals:
  - make progress on science content of the sub-project
  - build know-how to interface experimental software with ESCAPE tools
    - documentation particularly important for onboarding of others
- During this period, always happy to help seek funding for more in-kind resources

Further directions

- Nuclear physics
  - Using ALICE measurements to determine indirect detection backgrounds
- Other DM Direct Detection experiments pipelines (early talks with Xenon1T)
- Ideas for CERN:
  - FCC and future colliders - software pipelines & simulation
  - Other DM models (lighter DM to include e.g. FASER / forward physics facility)

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

Rough, preliminary timescale for DM Science Project

Months 12-24 (April 2022-October 2022)

- Focus of postdocs shared between science and implementation of workflows
  - Documentation needs to be written as we go
- First results from data analyses available
- Onboarding of other experiments can happen at this point

Months 24-30 (April 2021-October 2022)

- Full set of results from data analyses available —> creation of final plots
- Consolidation and dissemination

Planning for next DM TSP meeting: https://lettucemeet.com/l/NrrXX
Add your availability before Friday 07/05

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

Initiative for Dark Matter in Europe and beyond

- Many DM discussions, from **Granada** to the **ApPEC-ECFA-NuPECC JENAS meeting** held in Orsay in October 2019
  - Talk on ESCAPE (G. Lamanna) in plenary programme
  - HEP Software Foundation meeting on possible software synergies

- JENAS prompted a new initiative centered around **dark matter:** https://indico.cern.ch/e/iDMEu, also featured in ESCAPE newsletter
  - *iDMEu* aiming to build a discussion platform to facilitate collaboration of existing groups/efforts
  - *Dark Matter Test Science Project* targeting data, software and tools sharing where necessary/useful
  - Points of contact between *iDMEu* and *TSP*:
    - participation of DM community to software catalogue
    - list and help populate common repositories of data and final results (e.g. versioning)
      - e.g. DMTools, DM Limit Plotter

- Kick-off meeting on May 10-12 - everyone is welcome to register and participate!
  - In this meeting we won't yet talk of how to connect to EOSC-Future as this is more of a meeting to "survey DM communities" and understand what the needs are, bottom-up
  - There are **breakout sessions** that will be of interest to the work we're doing and we can raise new topics

REANA/RECAST in a (high-level) nutshell

[Lukas Heinrich's talk @ ESCAPE WOSSL 2020]

Simple Software Prerservation is not enough for Analysis Preservation: need the full pipeline

capture software	capture commands	capture workflow	data assets
archive analysis code incl. dependencies	what do with the captured software	order of individual steps	input data needed to run the analysis

Ingredients: Container Images, Workflow Languages
- **similar trends in bioinformatics**

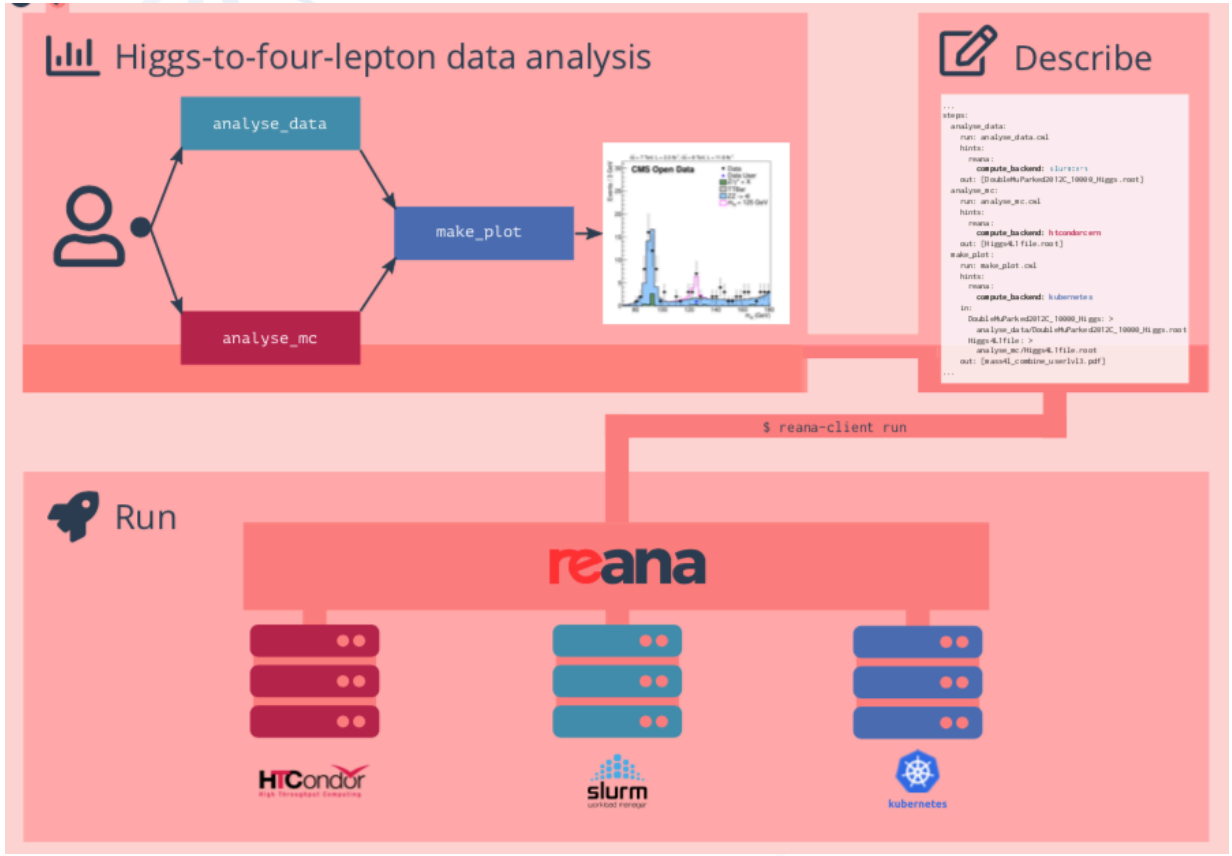
COMMON WORKFLOW LANGUAGE

CERN Working on Cyberinfrastructure to provide Archive of preserved analysis and compute resources to re-execute them (REANA)

CERN Analysis Preservation

reana
Reproducible research data analysis platform

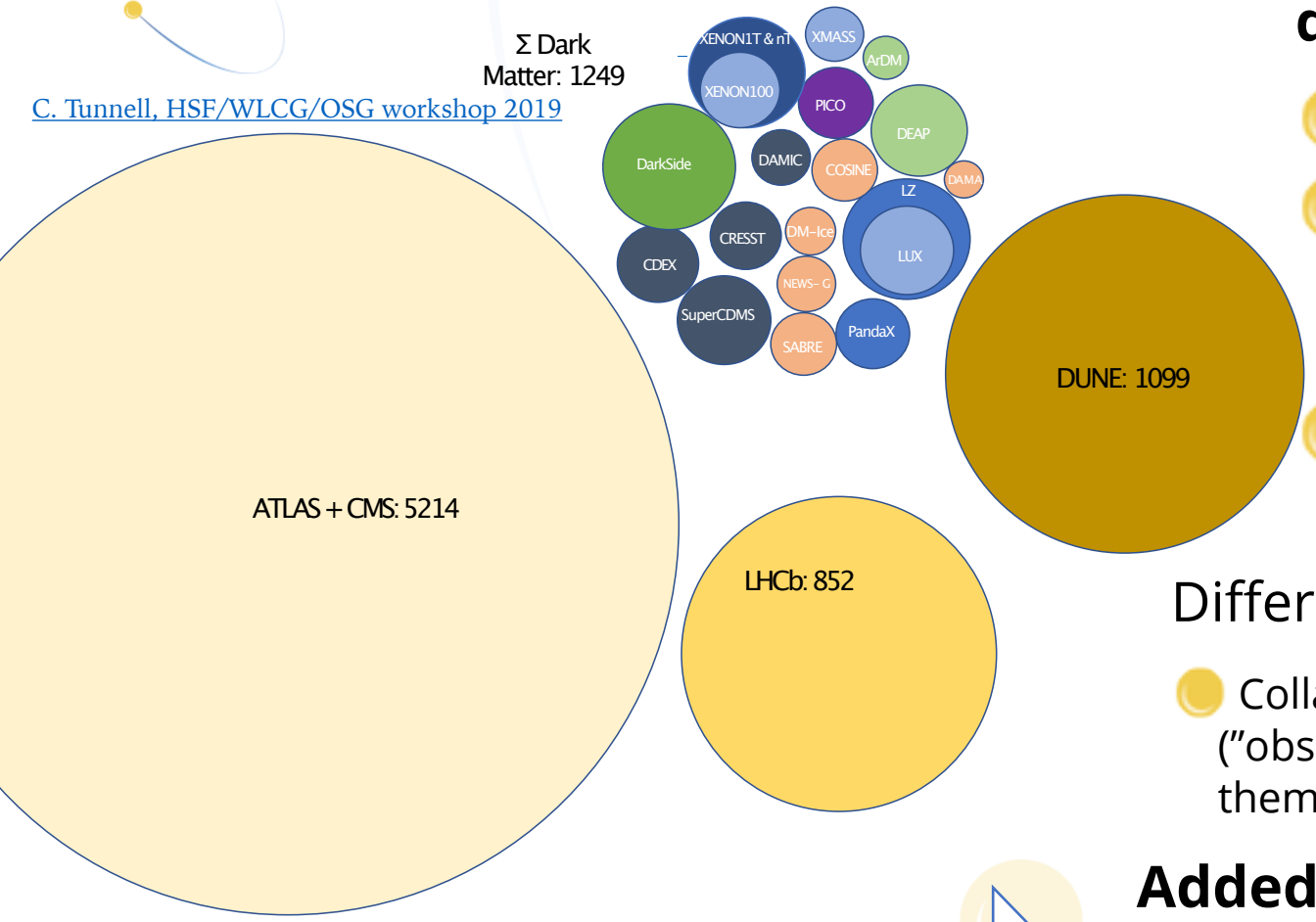
[Poster @ CHEP 2019]



Different kinds of (WIMP) DM communities & data/software needs

(full table uploaded by Ian Bird in Teams)

C. Tunnell, HSF/WLCG/OSG workshop 2019



C. Tunnell: Area corresponds to number of people based on most recent publication from any experiment that has published scientific papers in the last two years. This relied on Inspire-HEP. See gist for calculation notes. 16/March/2019

Diagram only representing **collider and direct detection**

- Differences in collaboration variety and size
- Differences in data volumes:
  - Colliders: ”Big Data” volumes (>> PB)
  - DD: smaller data volumes (~TB/PB)
- Synergies in statistical analysis and interpretation of results

Different modus operandi for **indirect detection**

- Collaborations e.g. Fermi release data for general use (”observatory mode”), but also perform high-profile analyses themselves

Added value of DM-TSP:

Proof that communities with different data needs can use a common platform —> further demonstration of EOSC versatility, encourages future use by the entire scientific community

The TSP, in a nutshell

Experimental method	Partners	EOSC Tools	DA Tools (AI/ML)	Project outcome
Producing DM in the lab (collider): ATLAS @ CERN	CERN LAPP	Data Lake Software Catalogue Analysis Platform	ML algorithms for: 1. Data compression 2. Data reconstruction (e.g. pattern recognition) 3. Background rejection	Constraints(/projections) on dark matter cross-section / DM mass plane and on dark matter velocity-averaged xsection / DM mass plane
Detecting dark matter from the sky (direct detection): DARKSIDE @ INFN	INFN	Data Lake Software Catalogue Analysis Platform	TBC	Constraints(/projections) on dark matter interaction cross-section/mass plane
Detecting interactions of dark matter using neutrinos (indirect detection): KM3NeT	FAU	Data Lake Software Catalogue Analysis Platform	TBC	Constraints(/projections) on dark matter cross-section/mass plane
Detecting interactions of dark matter in space (indirect detection)	LAPP	Data Lake Software Catalogue Analysis Platform	TBC	Constraints(/projections) on dark matter velocity-averaged xsection / DM mass plane
Surveying dark matter in the universe (astrophysical probes)	Open University [not in WP6 in EOSC-Future)	Data Lake Software Catalogue Analysis Platform	TBC	Combination of constraints on different models using simulation + statistical analysis software (Gambit)
[TSP2] Exploiting the gravitational interactions of DM (GW probes)	See GW TSP	Software Catalogue Analysis (multimessenger) Platform	See GW TSP	TBC

- Table, originally designed by ESCAPE-TSP-GW, is still as a work in progress
- Idea of Data Analysis Tools column: algorithms that can be **shared beyond a single infrastructure / field**
- IWAPP was very useful in terms of food for thought on **how to implement these common algorithms** (especially ML)
- How to follow up?

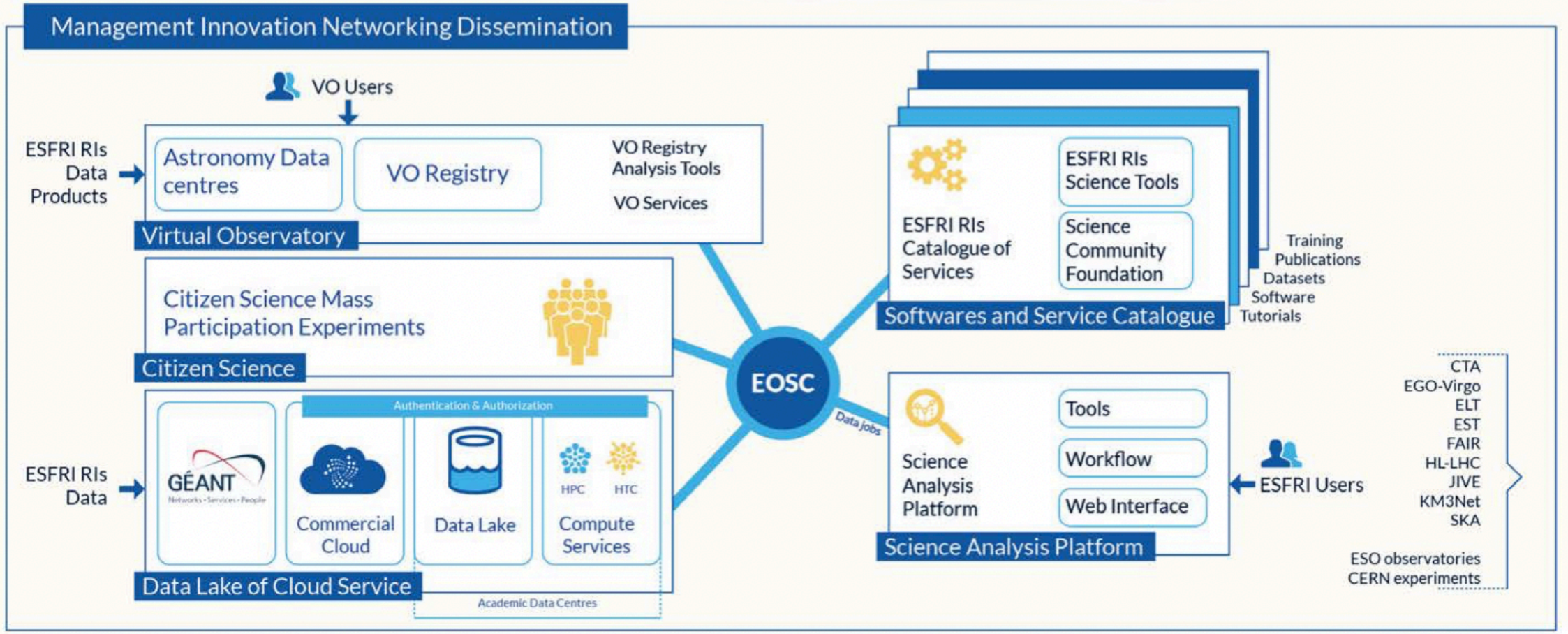
How/why Test Science Projects?

Slide from G. Lamanna

- Propose **Test Science Projects** to demonstrate multi-domain science integration across ESCAPE / EOSC
- **Involve researchers** to **demonstrate** new cutting edge **open science capabilities**, making use of the services implemented within EOSC
  - researchers can give feedback on the capabilities delivered by ESCAPE/EOSC
  - researchers can exploit synergies between the ESFRIs and among the scientific communities of Astrophysics/Astroparticle, accelerator-based Particle and Nuclear Physics
- Supported by consortia of EU member states research agencies and institutes within the **J**oint **E**CFA **N**uPECC **A**PPEC **A**ctivities (JENAA)

Services towards the European Open Science Cloud (EOSC)

Slide from G. Lamanna



ESCAPE services

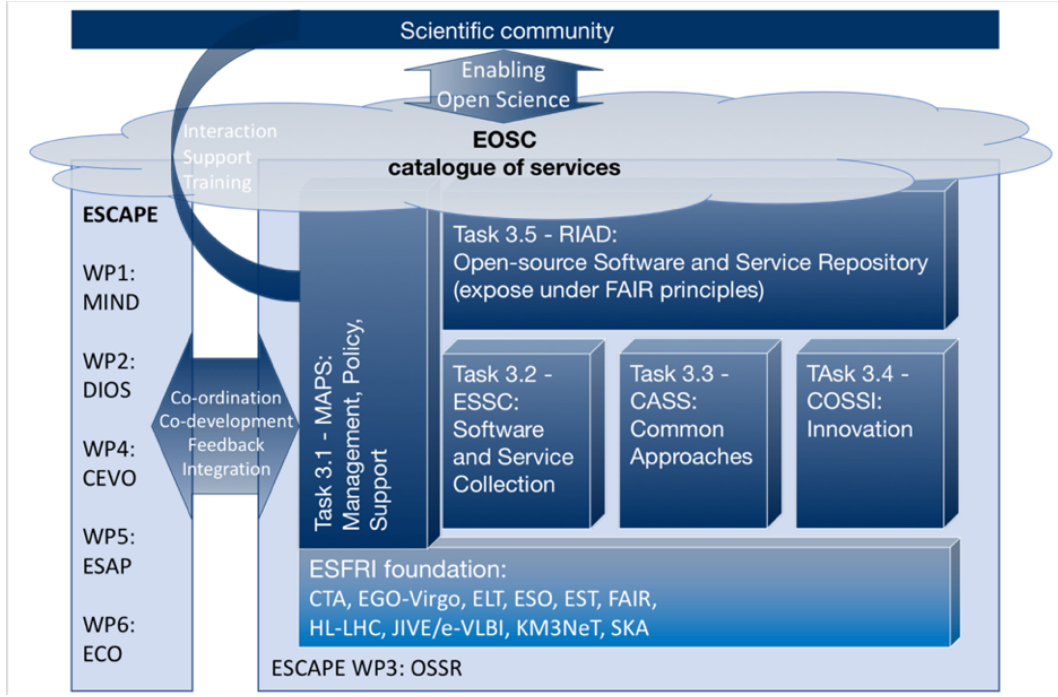
Slide from G. Lamanna

- **Data Lake**:
  - Build a scalable, federated, data infrastructure as the basis of open science for the ESFRI projects within ESCAPE. Enable connection to compute and storage resources.
- **Software Repository**:
  - Repository of "scientific software" as a major component of the “data” to be curated in EOSC. Implementation of a community-based approach for the continuous development of shared software and for training of researchers and data scientists.
- **Virtual Observatory**:
  - Extend FAIR standards, methods, tools of the Virtual Observatory to a broader scientific context; demonstrate EOSC ability to include existing platforms
- **Science Platforms**:
  - Flexible science platforms to enable the analysis of open access data
- **Citizen Science**:
  - Open gateway for citizen science on ESCAPE data archives and ESFRI community CS projects

ESCAPE software catalogue

Slide from K. Graf

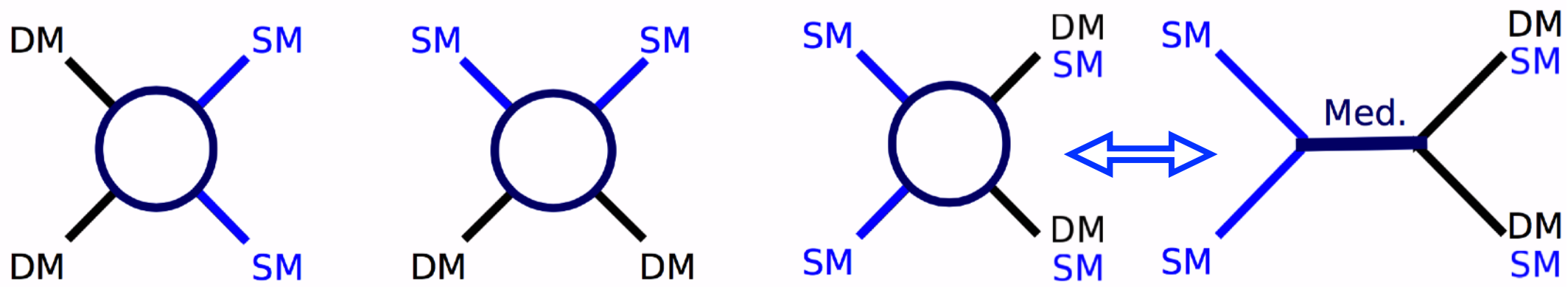
OSSR Overview



ESFRI/RI	Institute/SME
CTA	CNRS-LAPP
CTA	CTAO
CTA	IFAE
CTA	MPG-MPIK
CTA	UCM
EGO-Virgo	EGO
ELT	HITS
EST	AIP
EST	NWO-I-CWI
EST	UNITOV
FAIR	GSI
HL-LHC, CERN	CERN
JIVE	JIVE
KM3NeT	CNRS-CPPM
KM3NeT	FAU
KM3NeT	INFN
KM3NeT	NWO-I-Nikhef
SKA	SKAO
SME	OROBIX
9 ESFRI / RI	**19 Partners**

Dark matter complementarity

- DM discoveries need complementary experiments that involve DM with **cosmological origin** / can **produce DM**
  - Direct detection can **discover DM that interacts** inside the detector
  - Indirect detection can see **annihilating/decaying DM** through its decays
  - Accelerators/colliders can produce DM and **probe the dark interaction**

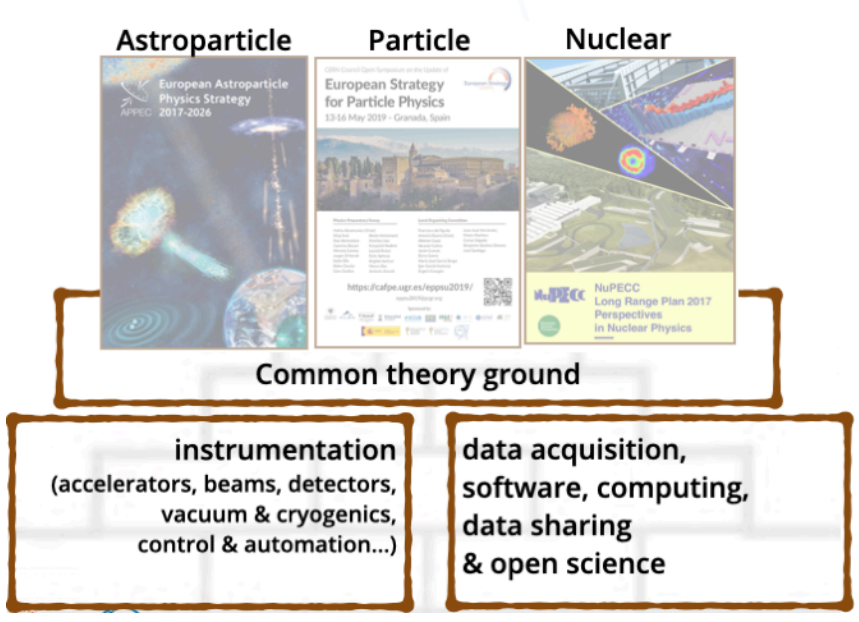


Indirect Detection (+ cosmic surveys) Direct Detection Particle Accelerators (colliders & extracted beam lines)

- Work on "common language / common resources" (plots, scenarios, tools) ongoing in Snowmass / iDMEu JENAA EOI / many other communities

Synergistic initiatives following European Strategy Update

searches & interpretation



JENAS EoI: Initiative for Dark Matter in Europe and beyond: Towards facilitating communication and result sharing in the Dark Matter community (iDMEu)

https://indico.cern.ch/event/869195/
ESCAPE newsletter APPEC newsletter

build a discussion platform to facilitate collaboration of existing groups/efforts on **dark matter searches** and **interpretation**

*provides a discussion platform for the **comparison of common DM interpretations***

software & data

Towards a Dark Matter Test Science Project

ESCAPE Progress Meeting, 2020

compare **end-to-end analysis workflows** for WIMP searches, towards their implementation in a common **Software Catalogue** and as input to the design of the **European Open Science Cloud**

*allows to **create experimental curves** by **example ESCAPE experiments,** comparing and contrasting analysis pipelines that use ESCAPE / EOSC tools*

More initiatives and links in backup slides