

# WP3 - OBELICS

## OBservatory E-environments LIinked by common Challenges

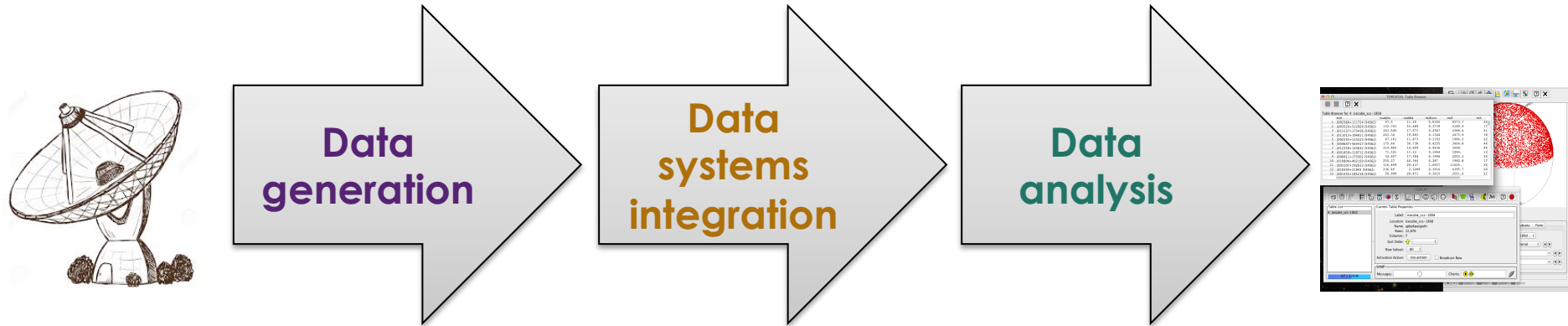
M18 Midterm Review

Giovanni Lamanna



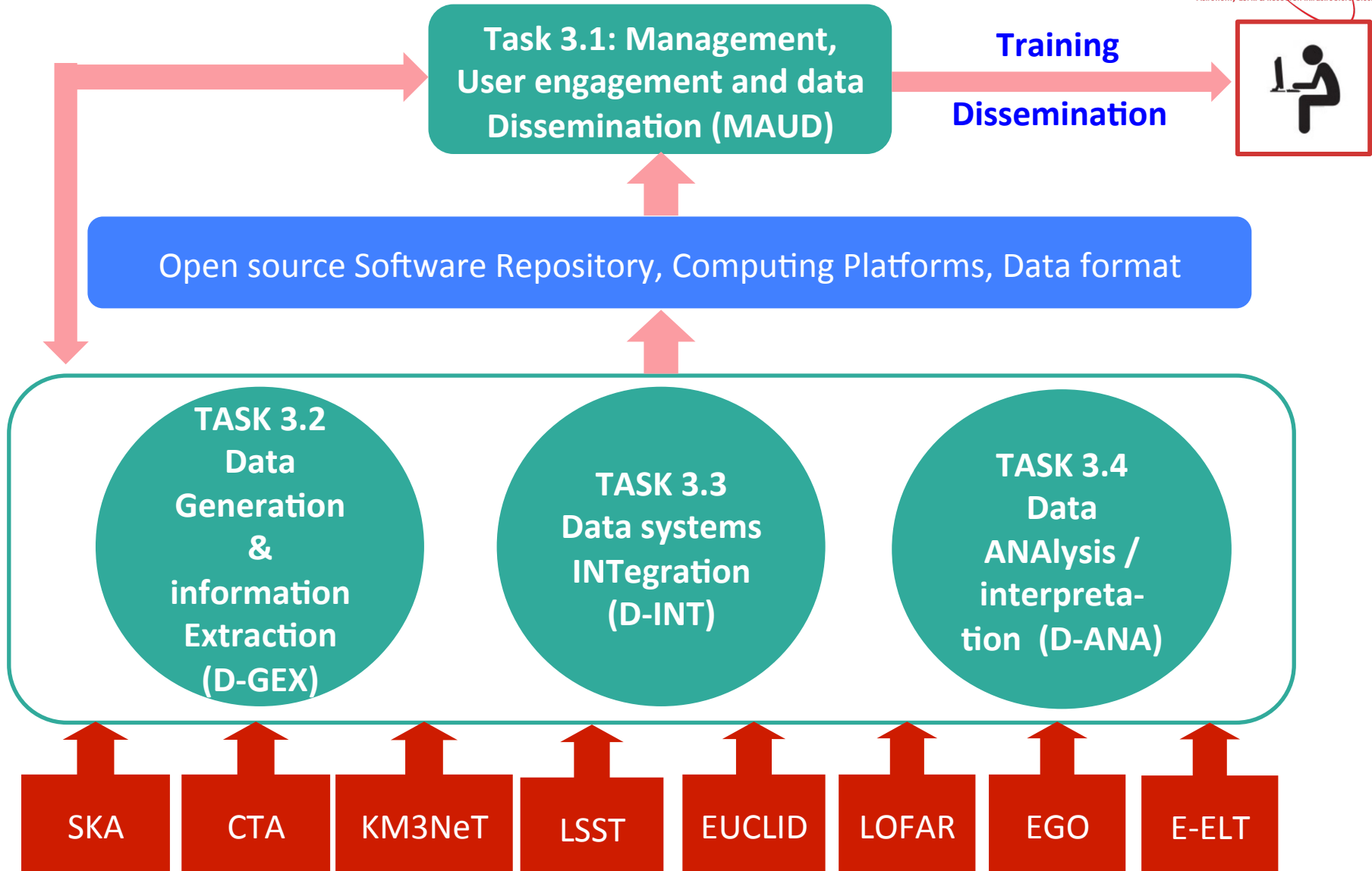
- The ASTERICS largest work package.
- Targeting common ESFRI-projects « Data Challenges ».
  - Generate large sets of data.
  - Need computing resources, intensive simulation and large storage space.
- Scopes:
  - Enable interoperability and software re-use.
  - Enable open standards and software libraries for multi-messenger data.
  - Develop common solutions, share prototypes, exchange experience.
- Objectives:
  - Create an open innovation environment across ESFRI facilities.
- Expected impact:
  - Economies of scale and saving resources.
  - Contribute to the construction and operation of ESFRI projects.

Working on commons along the “data flow”:



Twelve international partners cooperating around three main steps of data pipelines of major ESFRI projects in Astronomy.







## Internal Coordination

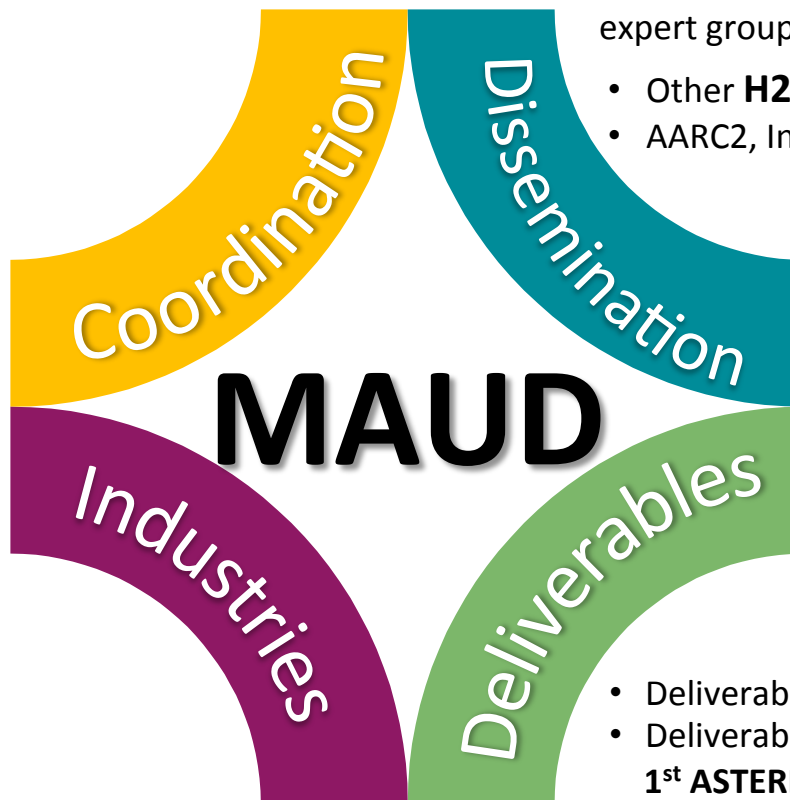
- Feb 2016: 1<sup>st</sup> face to face meeting in Rome,
- Sept 2016: 2<sup>nd</sup> face to face meeting in Madrid,
- Follow up with WP3 partner institutes on the work plan.

## Industrial Participation

- **Expression of interest** for Industrial Participation.
- Budget of €300000.
- 5-6 projects can be funded.
- Contacts made with **IBM, NVIDIA, ATOS, ORACLE, E4.**

## Dissemination to external community

- Preliminary contacts with **e-infrastructures** (EGI, PRACE)
- **EU consortia:** APPEC, EU-T0, RDA.
- Meeting with **European Open Science Cloud** expert group.
  - Other **H2020 projects:** HNSciCloud, AARC2, Indigo DataCloud
  - 4<sup>th</sup> LSDMA Symposium 2016; ESA-ESO SciOps 2015, SPIE 2016, IHPSS 2016.
  - Poster presentations at ADASS 2016 Trieste, GAMMA 2016 Heidelberg.



## Deliverables

- Deliverable 3.1: WP3 project plan
- Deliverable 3.2 + Deliverable 3.6 = **1<sup>st</sup> ASTERICS-OBELICS Workshop, Rome 12-14 December 2016.**

# 1<sup>st</sup> ASTERICS-OBELICS Workshop

*Science Data Cloud & Computing  
models in Astronomy and  
Astroparticle physics, Rome 12-14  
December 2016.*

Astronomy ESFRI & Research Infrastructure Cluster  
ASTERICS - 653477



- Merge of the first thematic training event (D3.2, M12) and the first general workshop (D3.6, M18) to formulize the scientific content and organization of the event.
- **7 tutorials with live demos, 20 invited talks and 5 panel discussions** addressing Authorization and Authentication, Data Storage, Transfer & Preservation, Large Databases, Workflow management & Interoperability.
- **87** scientists and technologists amongst the participants.
- Participants from **CERN & LIGO**.
- Call for **expressions of interest** for industrial cooperation was launched.
- Outreach on social media: **6450 tweet impressions**.
- **Positive impact** of deviation: Bolster networking with industries and other projects to address innovation related aspects.



- **1<sup>st</sup> ASTERICS-OBELICS International School (D3.7, M24)** , 6-9 June 2017, LAPP-Annecy, France
  - Specifically intended for a large audience including **PhD students, postdocs, senior researchers** from the domain of astrophysics and astroparticle physics .
  - The school will include theory and hands-on sessions on **Efficient code writing, Parallel & GPU programming & Python libraries for astronomy & astrophysics.**



- Post-workshop follow up
- Collaboration with industries through internal **call for expression of interests.**



## Deliverables Produced by MAUD

- D 3.1: Detailed WP3 Project Plan ✓
- D 3.2: First main thematic training event ✓
- D 3.6: First WP3 general workshop ✓





## Task 3.2 D-GEX Highlights

- **Low Power Computing platforms**
- **Efficient algorithms** for data reduction
- Development of **high performance computing (HPC) solutions**
- **A Data format survey (Deliverable D3.3)**
- Contribution to define specifications in an **open data format** within the gamma-ray community
- Adoption of the **open FITS format**,
- **HDF5** for data products other than images,
- **Data streaming and architecture** of the data processing unit in SKA,
- **Apache-Spark** for handle streaming, iterative streaming, quality assessment and databases





## D-GEX Impact

- The work conducted on Open Data Format will be particularly useful for **ESFRI Project KM3NeT to open source its data.**

## Future Plans

- **Low power architectures benchmarking** for building innovative computing, storage and data handling platforms will be explored.
- Exploitation of LPC architectures **for map/reduce applications.**
- The validation of the format gamma ray with current open science tools (e.g., ctools or Gammapy) as well as in studying the viability for the adaptation of this standard to other event-based experiments, like neutrino and cosmic-ray observatories.
- The optimisation and parallelization of algorithms for data reduction and analysis in order to exploit the low-power computing facilities in progress.

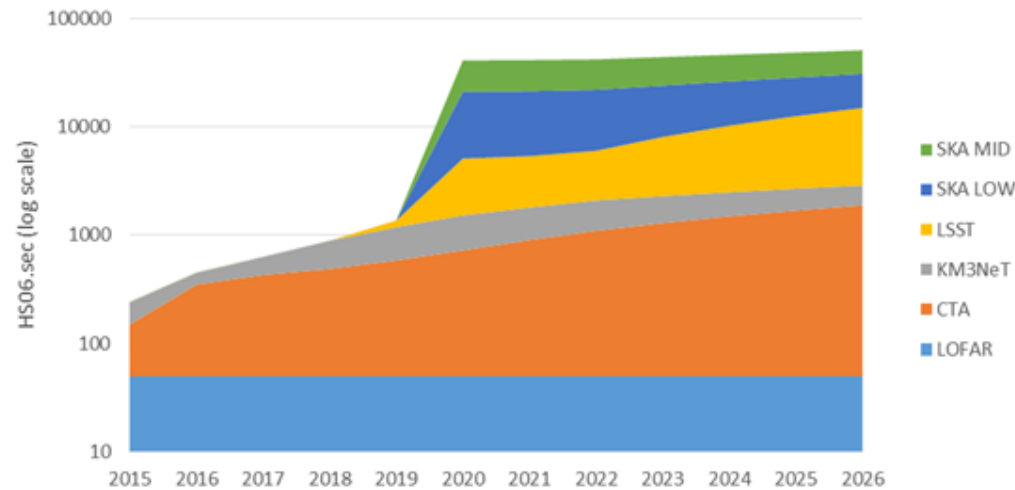
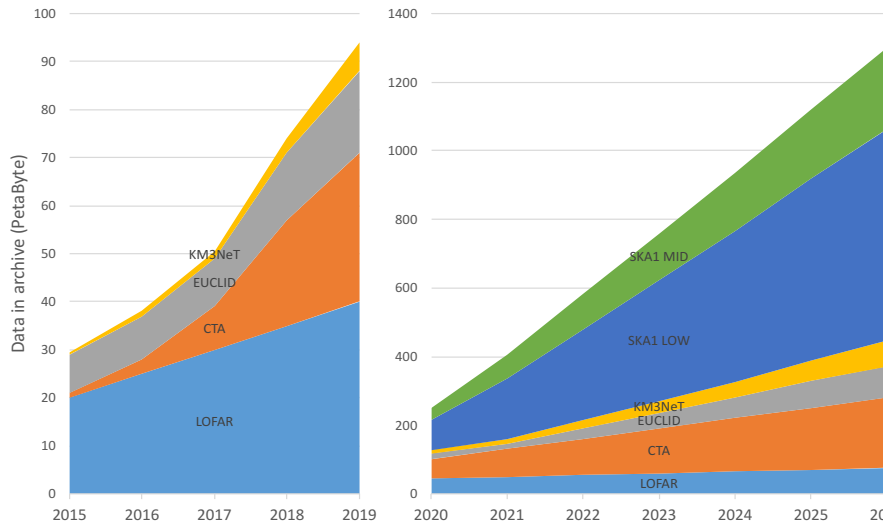
## Deliverables

D 3.3: Analysis Report on Standards and Libraries



# Task 3.3 D-INT Highlights

- Independent evaluation of all RI requirements:
  - Evolution of **storage needs and computing needs** for the ESFRI Projects.
  - Helping the collection of data management plan and requirements of **KM3NeT**
- An innovative algorithm of **lossless compression** for digitalized signals



Evolution of storage needs for the ESFRI Projects

Evolution of computing needs for the ESFRI Projects



## D-INT Impact

- Innovative application of the **docker** (<https://www.docker.com>) **containerisation** was identified.
- **Lossless compression algorithms** produced within one RI context now under discussion and tests in more infrastructures.
  - In the best case scenario, these will be adopted by the ESFRI projects.
- In most of the cases, these results have triggered discussions within collaborations and therefore improved the current development of the ESFRI projects.

## Future Plans

- Exploit the different database solutions particularly, **Qserv** solution developed within LSST and solutions based on **Spark** developed in EUCLID.
- Define a data model based on **HDF5** for CTA raw data and building an API to write appropriate HDF5 files.

## Deliverables

D 3.5: Analysis Report on Resource Requirements





## Task 3.4 D-ANA Highlights

- Prototyping of StatPLAN & MW-INFERENCE
- Development innovative algorithms for **calibration and imaging**
- Library for **fast, vectorised, array reductions and moment calculations** written in C++ with Python binding
- Initial general-purpose library of **A&A** and **workflow management** systems have been delivered as part of deliverable D3.4 + Pilot implementation
  - Work on interoperability between clouds (EGI-ComputeCanada) based on an IVOA standard (VOSpace) led to the agreement on an IVOA Authorization standard proposal, in coordination with WP2.
- Cooperation for the end-to-end simulation of the second **LIGO/Virgo science run**.
- **Wavelet denoising** for discriminating a Cherenkov signal versus the sky background.
- Robust direction and energy reconstruction algorithms which are crucial for the interpretation and **analysis of astroparticle signatures**





## D-ANA Impact

- The innovative libraries produced under D-ANA task for fast computation (PLIBS, PLISA) can potentially have **an impact outside of astronomy** since they implement fairly generic operations with wider usage (a scientific publication to a computing research related journal will be submitted).
- In addition, the notebook interface (VLBI in the Cloud) can be of use for **all radio interferometric data analysis**.
- The multi-wavelength inference library (MW-INFERENCE) is intended for the start to allow combination from **multiple astronomical facilities**.

## Future Plans

- The ongoing work is looking for integration possibilities between the modules being produced under this task.
- The implementation and completion of the libraries and tools that have been defined.

## Deliverables

D 3.4: Software Libraries repository (initial)





## Deliverables produced by WP3-OBELICS

- D 3.1: Detailed WP3 Project Plan ✓
- D 3.2: First main thematic training event ✓
- D 3.3: Analysis Report on Standards and Libraries ✓
- D 3.4: Software Libraries (initial) ✓
- D 3.5: Analysis Report on Resource Requirements ✓
- D 3.6: First WP3 general workshop ✓

## MAUD: Self Evaluation

- To fulfill the cross collaboration with e-infrastructures and industries through the **combination of tutorial and talks** was extremely successful and we intend to keep the similar format for the upcoming workshop in October 2017.
- **Call for Expressions of Interest for industrial cooperation** will close tomorrow, and will help address innovation related objectives.



# Technical Tasks: Self Evaluation

- **Interactions among subtasks** is required as simultaneous evolution of ASTERICS as well as member ESFRI projects, moving further towards their implementation, has been observed.

