



## ASTERICS - H2020 - 653477

# Third ASTERICS- OBELICS workshop

### ASTERICS GA DELIVERABLE: D3.16

Document identifier:	ASTERICS-D3.16.doc
Date:	<b>6 May 2019</b>
Work Package:	<b>WP3 Obelics</b>
Lead Partner:	<b>LAPP</b>
Document Status:	<b>Final</b>
Dissemination level:	Public
Document Link:	<a href="http://www.asterics2020.eu/documents/ASTERICS-D3.16.pdf">www.asterics2020.eu/documents/ASTERICS-D3.16.pdf</a>

### Abstract

This report summarizes the discussions that took place at the third ASTERICS – OBELICS workshop. This workshop was dedicated to New paths in data analysis and open data provision in Astronomy and Astroparticle Physics and it took place at the University of Cambridge, UK from 23-25 October 2018. This three-day workshop discussed common solutions developed by WP3 OBELICS, novel approaches in machine learning for data analysis and follow up on the European Open Science Cloud roadmap implementation to discuss potential contribution from astronomy, astroparticle and particle physics community.

## I. COPYRIGHT NOTICE

Copyright © Members of the ASTERICS Collaboration, 2015. See [www.asterics2020.eu](http://www.asterics2020.eu) for details of the ASTERICS project and the collaboration. ASTERICS (Astronomy ESFRI & Research Infrastructure Cluster) is a project funded by the European Commission as a Research and Innovation Actions (RIA) within the H2020 Framework Programme. ASTERICS began in May 2015 and will run for 4 years.

This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/3.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, and USA. The work must be attributed by attaching the following reference to the copied elements: “Copyright © Members of the ASTERICS Collaboration, 2015. See [www.asterics2020.eu](http://www.asterics2020.eu) for details of the ASTERICS project and the collaboration”. Using this document in a way and/or for purposes not foreseen in the license, requires the prior written permission of the copyright holders. The information contained in this document represents the views of the copyright holders as of the date such views are published.

## II. DELIVERY SLIP

	Name	Partner/WP	Date
Author(s)	Jayesh Wagh	LAPP	06/05/2019
Reviewed by	Rob van der Meer	ASTRON	06/05/2019
Approved by	AMST		10/06/2019

## III. DOCUMENT LOG

Issue	Date	Comment	Author/Partner
1	06/05/2019	First draft	Jayesh Wagh, LAPP
2	20/05/2019	Second draft	Jayesh Wagh, LAPP
3			

## IV. APPLICATION AREA

This document is a formal deliverable for the GA of the project, applicable to all members of the ASTERICS project, beneficiaries and third parties, as well as its collaborating projects.

## V. TERMINOLOGY

AENEAS	Advanced European Network of E-infrastructures for Astronomy with the SKA)
ARCA and ORCA	Astroparticle and Oscillations Research with Cosmics in the Abyss
ASTERICS	Astronomy ESFRI & Research Infrastructure Cluster

CTA	Cherenkov Telescope Array
EOSC	European Open Science Cloud
ESFRI	European Strategy Forum on Research Infrastructures
HNSciCloud	Helix Nebula Science Cloud
KM3NeT	Cubic Kilometre Neutrino Telescope
LSST	The Large Synoptic Survey Telescope
OBELICS	Observatory E-environments Linked by common ChallengeS
SKA	The Square Kilometre Array

A complete project glossary is provided at: <http://www.asterics2020.eu/glossary/>.

## VI. PROJECT SUMMARY

ASTERICS (Astronomy ESFRI & Research Infrastructure Cluster) aims to address the cross-cutting synergies and common challenges shared by the various Astronomy ESFRI facilities (SKA, CTA, KM3NeT & E-ELT). It brings together for the first time, the astronomy, astrophysics and particle astrophysics communities, in addition to other related research infrastructures. The major objectives of ASTERICS are to support and accelerate the implementation of the ESFRI telescopes, to enhance their performance beyond the current state-of-the-art, and to see them interoperate as an integrated, multi-wavelength and multi-messenger facility. An important focal point is the management, processing and scientific exploitation of the huge datasets the ESFRI facilities will generate. ASTERICS will seek solutions to these problems outside of the traditional channels by directly engaging and collaborating with industry and specialised SMEs. The various ESFRI pathfinders and precursors will present the perfect proving ground for new methodologies and prototype systems. In addition, ASTERICS will enable astronomers from across the member states to have broad access to the reduced data products of the ESFRI telescopes via a seamless interface to the Virtual Observatory framework. This will massively increase the scientific impact of the telescopes, and greatly encourage use (and re-use) of the data in new and novel ways, typically not foreseen in the original proposals. By demonstrating cross-facility synchronicity, and by harmonising various policy aspects, ASTERICS will realise a distributed and interoperable approach that ushers in a new multi-messenger era for astronomy. Through an active dissemination programme, including direct engagement with all relevant stakeholders, and via the development of citizen

scientist mass participation experiments, ASTERICS has the ambition to be a flagship for the scientific, industrial and societal impact ESFRI projects can deliver.

## VII. EXECUTIVE SUMMARY

The third ASTERICS-OBELICS workshop took place at the Postdoc Centre-Eddington, University of Cambridge, Cambridge, UK on 23-25 October 2018. It focused on dissemination of OBELICS achievements and plan for the final OBELICS deliverables considering the H2020-ASTERICS project conclusion in April 2019. In addition, this workshop explored novel use cases and approaches in machine learning in astronomy, astroparticle physics, as well as particle physics. The workshop also discussed the current progress in the setting up of EOSC with presentations from various key players involved, such as the projects EOSCpilot and EOSC-hub. Unlike last edition, this year the workshop focused on more technical aspects and implementation roadmap of EOSC, which was of interests to the scientific community.

# Table of Contents

<b>I. COPYRIGHT NOTICE</b>	<b>1</b>
<b>II. DELIVERY SLIP</b>	<b>2</b>
<b>III. DOCUMENT LOG</b>	<b>2</b>
<b>IV. APPLICATION AREA</b>	<b>2</b>
<b>V. TERMINOLOGY</b>	<b>2</b>
<b>VI. PROJECT SUMMARY</b>	<b>3</b>
<b>VII. EXECUTIVE SUMMARY</b>	<b>4</b>
<b>Table of Contents</b>	<b>5</b>
<b>1. Introduction</b>	<b>6</b>
<b>2. Structure of the meeting</b>	<b>7</b>
<i>2.1 Day 1: OBELICS presentations</i>	<i>7</i>
<i>2.2 Day 2: Machine learning presentations</i>	<i>7</i>
<i>2.3 Day 3: EOSC presentations</i>	<i>7</i>
<b>3. Participation to the meeting</b>	<b>9</b>
<b>4. Results of the meeting</b>	<b>10</b>
<b>5. Evaluation of the meeting</b>	<b>14</b>

# 1. Introduction

Out of the five workpackages of H2020-ASTERICS, WP3 - OBservatory E-environments Linked by common ChallengeS (OBELICS) is the largest workpackage. WP3 OBELICS aims at enabling interoperability and software reuse, software libraries for multimessenger data and developing common solutions, sharing prototypes and best practices.

The third ASTERICS-OBELICS workshop took place at the Postdoc Centre-Eddington, University of Cambridge, Cambridge, UK on 23-25 October 2018. It was entitled as *New paths in data analysis and open data provision in Astronomy and Astroparticle Physics*. The event was organized by CNRS-LAPP and locally hosted and supported by the University of Cambridge. For this third edition of the workshop, we decided to focus on following priorities concerning astronomy ESFRI facilities.

- As H2020-ASTERICS project will conclude in April 2019, we organised the workshop to present an overview of software and services produced by WP3-OBELICS for astronomy ESFRI projects over the last four years to external participants from EOSC, other H2020 projects, industrial participants.
- In this edition of the workshop, we also decided to focus on the technical aspects of the European Open Science Cloud to update the astronomy and astroparticle physics community on the implementation roadmap.
- Machine learning is rapidly changing data analysis in many domains and has entered the Astronomy and Astroparticle physics. Following the positive response to the Machine Learning session in the second edition of the OBELICS workshop, we decided to continue to explore novel approaches in Machine Learning adopted by the astronomy and astroparticle physics projects since then.

## 2. Structure of the meeting

It was a three-day workshop. All the three days were open to all the registrants.

### 2.1 Day 1: OBELICS presentations

On the first day we introduced the participants with H2020-ASTERICS and its objectives and presented numerous WP3-OBELICS achievements in data generation and information extraction, data systems integration and data analysis and interpretation. Overall 19 presentations (10 min each) provided a comprehensive overview of OBELICS activities in all the task groups. The first day session concluded with a panel discussion to discuss the blueprint OBELICS report that can be referenced as an OBELICS contribution to the astronomy ESFRI projects after the end of H2020-ASTERICS. This report is being drafted and expected to be made public by June 2019.

### 2.2 Day 2: Machine learning presentations

The second day of the workshop was dedicated to present updates, experience and lessons learnt from ESFRI partners during the adoption of Machine Learning techniques. Latest developments on Bayesian statistics methods applied to astronomy were presented. ASETRICS developments and partnerships with industries were also presented. The ESFRI and other world-class research projects supported by ASTERICS attended the event and presented their own work on this matter. Invited talk from other scientific communities is expected. This session provided an opportunity to discuss new paths/common approaches to minimize the explorative efforts from all and optimize a shared know-how to advance in deep-learning analysis methods in a multi-messengers context. The workshop provided a platform for the scientists and technologists to share competences and to discuss new paths/common approaches to minimize the explorative efforts from all and optimize a shared know-how to advance in deep-learning analysis and Bayesian methods for astronomical and particle physics data.

### 2.3 Day 3: EOSC presentations

The third day was dedicated to the discussion about the current progress in the **setting up of EOSC** with presentations from various key players involved such as **EOSCpilot**, **EOSC-hub**, **HNSCiCloud** and **AENEAS**, while another important topic was the evolution of the ESFRI in terms of computing and services to foster future cooperation. It included presentations from

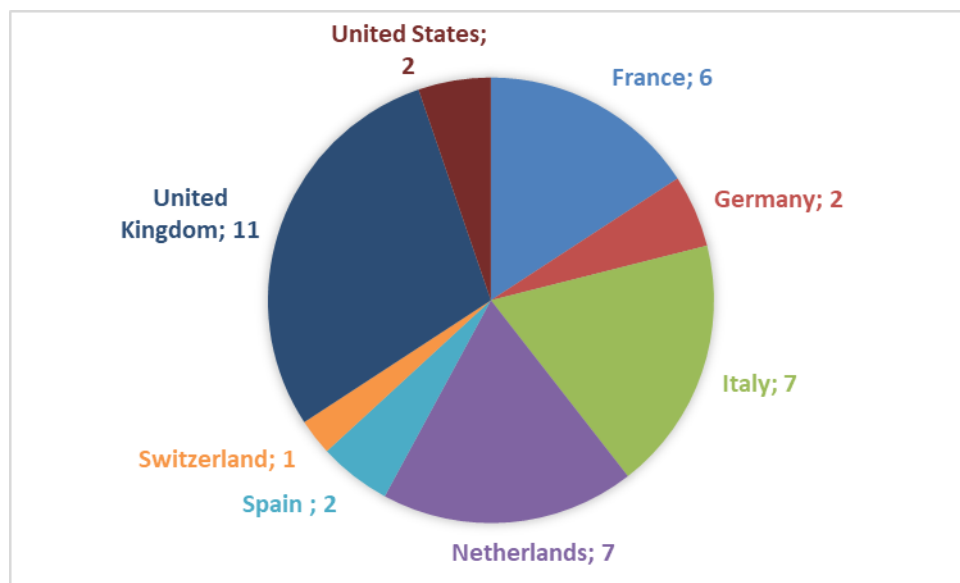


EOSC-hub providing an overview on the project status and plans, presentation of service catalogue. Rules of participation for service providers will be discussed through EOSCpilot outcomes and EOSC-hub experience. These discussions also included the long-term sustainability of the OBELICS scientific software repository. The added value of partnerships with industries were reviewed and discussed.

The detailed programme of the workshop can be found on the meeting website, <https://indico.astron.nl/conferenceTimeTable.py?confId=188#20181023>

### 3. Participation to the meeting

The workshop brought together 38 scientist and technologist from ASTERICS partner institutions as well as representatives from **EOSCpilot**, **EOSC-hub**, **HNSCiCloud**, **AENEAS**. In addition, experts from CERN as well as EGO were amongst the participants. The complete list of participants, speakers and the presentations are available on workshop website <https://indico.astron.nl/confRegistrantsDisplay.py/list?confId=188>



*Figure 1 Third edition of the workshop had 38 registrants.*

## 4. Results of the meeting

Overall, the workshop was well received by the scientific community with active participation in discussions and talks. The event also produced significant impact on social media outreach of H2020-ASTERICS.

Dissemination of WP3-OBELICS activities and knowledge exchange was one of the result of the workshop. The workshop gave floor to the WP3-OBELICS members to demonstrate the common solutions developed for astronomy ESFRI projects. On this occasion, results from industrial collaboration were presented. GammaLearner, a Deep Learning framework for CTA was presented. This deep learning framework was developed by CNRS-LAPP in collaboration with OROBIX, an Italian SME. In addition, H5 data format for Atmospheric Cherenkov Astronomy was also presented. This work was carried out by UCM in collaboration with Quasar SR, an enterprise born from ESA incubator and with expertise in space missions support software as well as astronomical software support.

The second day of the workshop started with the application of Machine Learning in Gravitational Wave research in the context of the **Machine Learning LIGO/Virgo group**. The group is actively adopting Machine Learning and Deep Learning to improve techniques for rapid analysis of data such as glitches classification, noise removal, data analysis and data conditioning.

Specific Machine Learning algorithms and software tools have been applied in the development of the **Square Kilometre Array (SKA)** with the aim of facilitating the execution of automated learning methods processing the vast quantities of data produced by the SKA. From the experience of the **Cherenkov Telescope Array**, supervised machine learning algorithms, like random forest or boosted decision trees, have been successfully applied to the task of **event reconstruction** by current generation of imaging atmospheric Cherenkov technique (CTA). The Cherenkov Telescope Array is now working to enhance its performance with **Deep Learning to reconstruct CTA events from non-parameterized images**, which should produce the possibility to extract and map images with an unprecedented classification accuracy.

In the presentation given by **SLAC** (National Accelerator Laboratory operated by Stanford University for the U.S) representative, the participants learned how new and powerful approaches to analyzing High Energy Physics data have being developed by connecting the challenges in the HEP domain with those in Machine Learning and Deep Learning.

The presentation started with a state-of-the-art analysis on how Machine Learning is currently widely used in different steps in HEP such as analysis, reconstruction, trigger and computing and continued with the quick evolution of **Deep Learning application across the whole data acquisition, simulations, and analysis pipeline** in HEP.



*Figure 2 Sebastian Hoch from European Solar Telescope project presenting machine learning in solar physics*

Machine Learning techniques such as BDTs, Random Forests, Shallow and Deep Networks are an established part of **KM3NeT** analysis toolchain ranging from event classification, particle identification, energy/direction estimation, signal/noise discrimination, source identification and data analysis. During the presentation specific Machine and Deep Learning applications for **ARCA and ORCA projects** highlighted preliminary encouraging results concerning reconstruction phase with stable estimations, comparable and in some cases better than KM3NeT standard reconstruction.

The third and final day opened with the discussion on how the **WLCG community** is adopting Machine Learning to ensure effective changes to the computing models and distributed computing infrastructure that will be needed to meet the challenges of developing a multi-Exabyte process which is manageable in various aspects not just at a computing level but also in terms of costs and resources.

A representative of the **EOSCpilot** showed the progress of the project in terms of engaging with a broad range of different stakeholders to set up the foundation of the EOSC ecosystem in the following aspects:

- *Governance* - EOSCpilot has defined a framework of Governance ensuring a good balance within 3 layers: Advisory, Executive and Institutional board (including EC and EU members representatives) and their interactions.
- *Policy* - A set of high level recommendations have been set up focusing on interactions with industry and research infrastructures, supporting Open Source, Data protection assurance and ownership and Open Science in general.
- *Service portfolio* - Defining what services an organisation will provide to EOSC with standard description and requirements.
- *Services and Data Interoperability* - A dedicate WP in EOSCpilot has investigated and proposed an architecture and solutions to enable this across different e-infrastructures and to ensure maximum re usability across different domains.

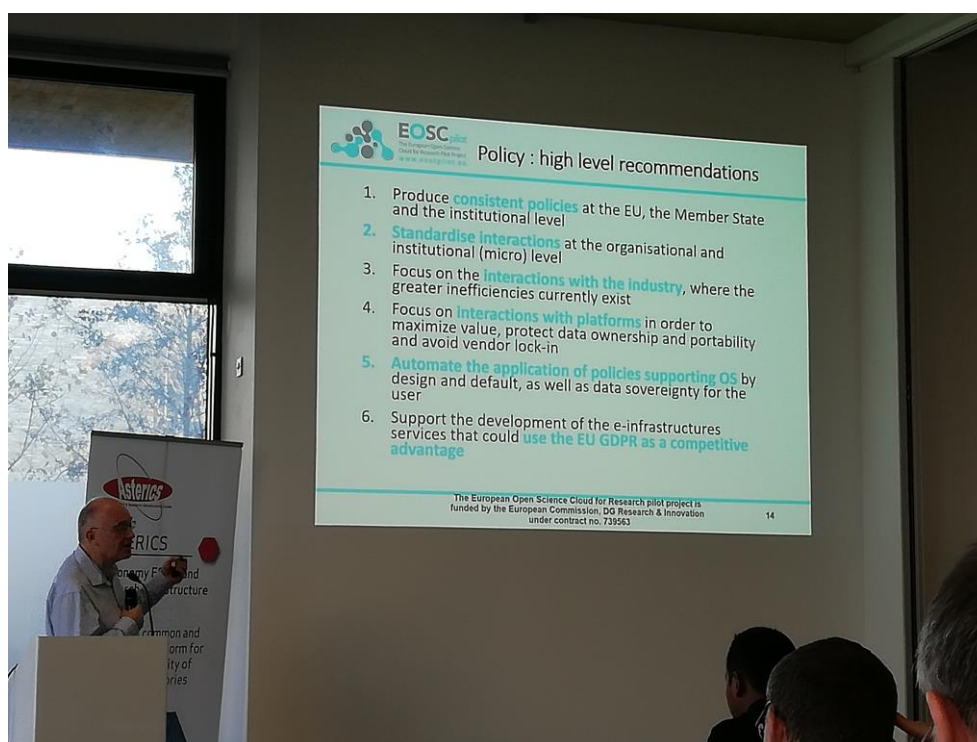


Figure 3: Brian Matthews from STFC presenting the European Open Science Cloud pilot

On the other side, **EOSC-hub** showed the development of a single contact point to access Services, Software and Data and to enhance a coordinated approach for RIs for policies, procedures etc. known as the EOSC-hub Marketplace. The **EOSC-hub Marketplace** is a platform where services can be exploited also in the future whereas EC projects have usually a shorter life.

After the conclusion of the project, the EOSC-hub Marketplace will also be directly adopted by EOSC and rebranded as EOSC Marketplace effectively becoming a tool of the future EOSC. Tiziana Ferrari, representative of EOSC-hub, strongly highlighted the need of the inclusion in the EOSC-hub Marketplace of the OBELICS Software and Service Repository to keep the collaboration ongoing and ensure a concrete and proper representation of the astronomy and astrophysics community into EOSC.

The work carried out in **HNSciCloud** is a contribution towards the development of the EOSC as one of the facilitating entities providing innovative services addressing the needs of European research communities through a common hybrid science cloud platform.

HNSciCloud hybrid cloud platform links together commercial cloud service providers and research organisations' in-house IT resources via the GEANT network. The platform offers data management and compute services accessible via eduGAIN and ELIXIR federated identity and access management systems with support services, account management facilities, documentation and training.

Two use cases related to the astronomy community, **MAGIC and CTA**, were also presented as examples of the functional and scalability tests that HNSciCloud will undertake in late 2018.

The **AENEAS project** presented its progress in supporting the pan-European astronomical community in achieving the scientific goals of the SKA.

A first map of data management facilities across Europe that have expressed interest to serve as nodes in the European Science Data Centre for the SKA was presented together with some of the challenges that the project is taking into account such as data transportation, interfaces, tools and techniques required and data hosting.

As next steps, AENEAS is looking to develop a Governance and a Business model for its federated European Science Data Centre also taking into account future strategic collaborations such as EOSC, and will present its preliminary results in the first half of 2019.

## 5. Evaluation of the meeting

We used different channels such as email campaigns and social media to disseminate the event. We managed to have participation from key representatives from ESFRI projects as well as EOSC related projects. Compared to previous editions of the workshops, we observed less registrants for this event. We observed that it is quite important to finalize the workshop programme well in advance to improve the number of participants. Despite the modest number of registrants, we observed quite fruitful discussions and interactions amongst the participants during the workshop sessions.

At the end of the second day, a panel discussion was chaired by Thomas Vuillaume (LAPP) involving Elena Cuoco (LIGO/Virgo), Michael Kagan (SLAC), Sebastian HOCH (Kiepenheuer-Institut für Sonnenphysik) and Fred Moolenkamp (Rider University). The aim was to find common points of collaboration in Machine Learning. Some of the outcomes of this panel discussion are listed below.

- **Data sharing** is perceived as one of the key topics to set a common ground, however this is not an easy path. For example, **a common data format** has not been adopted yet so different communities are actually working with different formats therefore creating barriers to the road towards open data.
- **Defining better metadata** could help making progress towards a more open data but still there are differences in the way communities analyse these data: for example in LIGO/Virgo, the analysis of data is performed on the entire data, while metadata are covering just specific parts of it.
- A need for a **common sharing visualization tool** emerged from the panel, where all panellists use different tools in their research and this can make it difficult to share their work. Also, having so many tools to be used, open up for practical difficulties in remembering keywords, libraries and functionalities and therefore slowing down the research work.
- All panellists agreed also that having a collaborative platform could be useful to **share knowledge**, get input and corrections from the different communities as well giving guidelines with the aim of developing best practices for resolving similar issues that can be encountered during the research work across different domains.

Similarly, an **interactive discussion and a panel** involving both speakers and the workshop participants were organized on the third day of the event to further discuss the relation between EOSC and ESFRI. Some of the discussion points for the steps forward included:

- A majority of the participants indicated that they will likely use EOSC for their research, but also they indicated that its **facility of use should be a mandatory requirement** to ensure proper buy-in from the researchers.

- EOSC should give recommendations on how to **coordinate large investments in hardware for EU science**. For example, would EOSC plan and manage investments to guarantee preservation and access to the large data volume of ESFRI-RI data?
- EOSC is also a good way to anchor certain models for how we want science to be managed in the future. A clear structure is needed **to ensure all voices are respected from a Cluster perspective but also from a Member States perspective and from a RIs perspectives**.
- EOSC should be **open to the continuous development and training** that forms the basis of the scientific approach.
- Data are key to the EOSC concept. Since there is no 'one-size-fits-all' approach for research data management, specific **data management plans and platforms are essential to ensure a proper buy-in** of the research community, as well as ensuring quality and openness.