Earth and Universe Multi-Messenger Observatory technical annex

Call H2020-INFRAEOSC-2018-2020

Topic INFRAEOSC-02-2019

Research and Innovation action

Research and Innovation action

The structure of this template must be followed when preparing your proposal. It has been designed to ensure that the important aspects of your planned work are presented in a way that will enable the experts to make an effective assessment against the evaluation criteria. Sections 1, 2 and 3 each correspond to an evaluation criterion.

Please be aware that proposals will be evaluated as they were submitted, rather than on their potential if certain changes were to be made. This means that only proposals that successfully address all the required aspects will have a chance of being funded. There will be no possibility for significant changes to content, budget and consortium composition during grant preparation.

⚠ **Page limit**: The title, list of participants and sections 1, 2 and 3, together, should not be longer than 70 pages. All tables, figures, references and any other element pertaining to these sections must be included as an integral part of these sections and are thus counted against this page limit.

The page limit will be applied automatically; therefore you must remove this instruction page before submitting.

If you attempt to upload a proposal longer than the specified limit before the deadline, you will receive an automatic warning and will be advised to shorten and re-upload the proposal. After the deadline, excess pages (in over-long proposals/applications) will be automatically made invisible, and will not be taken into consideration by the experts. The proposal is a self-contained document. Experts will be instructed to ignore hyperlinks to information that is specifically designed to expand the proposal, thus circumventing the page limit.

Please, do not consider the page limit as a target! It is in your interest to keep your text as concise as possible, since experts rarely view unnecessarily long proposals in a positive light.

⚠ The following formatting conditions apply.

The reference font for the body text of H2020 proposals is Times New Roman (Windows platforms), Times/Times New Roman (Apple platforms) or Nimbus Roman No. 9 L (Linux distributions).

The use of a different font for the body text is not advised and is subject to the cumulative conditions that the font is legible and that its use does not significantly shorten the representation of the proposal in number of pages compared to using the reference font (for example with a view to bypass the page limit).

The minimum font size allowed is 11 points. Standard character spacing and a minimum of single line spacing is to be used.

Text elements other than the body text, such as headers, foot/end notes, captions, formula's, may deviate, but must be legible.

The page size is A4, and all margins (top, bottom, left, right) should be at least 15 mm (not including any footers or headers).

Contents

EARTH AND UNIVERSE MULTI-MESSENGER OBSERVATORY

List of participants

#	Participant organisation name	Short Name	Country
1	Centre National des Rechereches Scientifiques	CNRS	France
2	University Paris Diderot	UPD	France
3	University of Geneva	UNIGE	Switzelrand
4	University of Leiden	UL	Netherlands
5	Kiev Academic University	KAU	Ukraine
6	Dublin College University	DCU	Ireland
7	Swiss Federal Institute of Technology	EPFL	Switzerland
8	European Gravitational Observatory	EGO	Italy
9	ASI Space Science Data Centre	SSDC	Italy
10	Linkalong Sarl.	LA	Switzerland
11	Parameter Space Ltd.	PS	Ireland

1 Excellence

Your proposal must address a work programme topic for this call for proposals.

Earth and Universe Multi-Messenger Observatory (EU−MMO) proposal responds to the INFRAEOSC-02-2019 call of Horizon-2020 Work Program Prototyping new innovative services for European Open Science Cloud (EOSC).

EU−MMO project will put together partners which are all deeply involved into data-driven research and promotion of **Findable-Accessible-Interoperable-Reusable (FAIR) data management principles** in different science domains to create a network of cloud-based services for

- implementation of the FAIR principle for cloud-based data analysis workflows;
- reproducible research based on-the fly cloud-based FAIR analysis workflows linked to publications;
- Collaborative data analytics for combination of FAIR data analysis tools within multi-messenger, multi-wavelength, multi-probe research in the domains of astronomy, cosmology and environmental sciences.

The project will leverage on pre-existing developments of innovative approaches for data-driven collaborative research by project partners. The project team has accumulated large experience in developing and running paltforms and services for generic data science (Renku platform developed by EPFL/SDSC, ScienceWISE platform developed by LA and EPFL) and for domain specific data analytics for astronomy (ODA: Online Data Analysis service for multi-messenger astronomy developed by Astronomy Department of UNIGE, OpenUniverse.org services developed by collaboration between SSDC and United Nations Outer Space Affairs office, GAVIP data analysis system developed by PS partner in collaboration with European Space Astronomy Center of the European Space Agency, Multi-Mission Archive for space based telescopes developed by SSDC, AstroWISE platform for management of optical band sky survey data developed by UL) and Earth observations (LiMES : Live Monitoring of Earth Surface service developed by partnership between UNIGE and United Nation Environmental Protection agency, UNEP/GRID Geneva). These services and platforms (TRL6 or higher) are actively used in the forefront scientific research and their features are optimized to maximize knowledge extraction from observational big data (in the sence of data variety, vercacity, volume) in the domains of astronomy, cosmology, astroparticle physics, environmental science.

In related domains of **astronomy** and **astroparticle physics**, the preparatory work for the project is motivated by rapid development of multi-messenger astronomy field has just recently been born with detections of first

astronomical sources of gravitational waves and high-energy neutrinos. This new field explores the limits of applicability of laws of Nature via observations and modeling of violent phenomena in astronomical sources which host ultra-strong gravitational and magnetic fields, ultra-dense and ultra-hot matter not encountered in Earth-bound laboratories. This could not be done based on the information collected by single telescope or in single astronomical messenger channel. New collaborative approaches to data management in this field are driven by the need of combining diverse data collected by various types of telescopes and detectors within a coherent multi-messenger picture.

The domain of **cosmology** is experiencing a profound transformation related to the start of operations of massive sky surveys enabled by proliferation of big data technology applications to electromagnetic waves astronomy in different wavelength bands: from radio to optical/infrared and X-ray bands. Sky surveys provide petabytes of data which are invaluable source of information not only about the structure and the origin of our Universe, but also about fundamental laws of Nature. Using a combination of diverse types of cosmological probes within multi-probe framework, cosmologists have been able to draw an intricate picture of the Universe all the way back to the first few moments of its existence. Yet, with every new step, the questions become increasingly difficult to answer. Seemingly independent and unrelated discoveries begin to intertwine, and existing methods to observe and interpret cosmological data become insufficient. This drives the profound change of data management approaches also in this field.

In the domain of **environmental science**, massive proliferation of digitization revolutionizes our understanding of complex environment of the Earth planet and characterization of human impact on it. Deforestation, fish farming, soil erosion, multiple pollution sources, impacts from extractive industries and unplanned urbanization are examples of these cumulative changes, which are causing local environmental impacts and biodiversity losses, but also leading to global environmental change because they are occurring on a worldwide scale. Digital revolution opens a way to continuous monitoring of ecosystems based on multi-wavelength satellite image analysis. Implementation of this approach requires new organization of data management which provides innovative, powerful and cost-efficient alternative solution to conventional conservation monitoring exercises, which are reactive in nature, slow and expensive.

In spite of the difference of the research questions, **knowledge extraction process** from multi-messenger astronomical data, multi-probe cosmological data, multi-wavelength Earth observations is characterized by similar **challenges**:

- it relies on plethora of different data which have to be analyzed simultaneously within the same research problem;
- it involves non-trivial theoretical predictions that are increasingly based on complex numerical simulations rather than analytic models, and
- it has to assure self-consistent tracing of the influence of theoretical assumptions throughout the simulations and multi-probe and multi-messenger data analysis.

Proper answer to these challenges requires transformation of conventional scientific methods and practices of organization of data and access to them, of representation of the results, of the way numerical simulations are connected with observations. The EU−MMO project will address these challenges by building a network of **intelligently linked cloud-based services** which will implement a **"holistic" approach for management of complex multi-messenger, multi-probe, multi-wavelength data** in which all raw data sources, data lineage, all theoretical assumptions and model calculations are self-consistently traced up to the high-level publishable scientific results. This approach will transform the way we carry research, creating a transparent framework in which we go from raw data to results using multi-experiment probes that can be combined under consistent and constrained roof.

The very nature of multi-messenger, multi-wavelength and multi-probe approaches in the domains as distant and astronomy / cosmology and environment protection implies an important role of **sharing of analysis expertise** across different observation facility experts, experts on numerical simulations and theoreticians working on model developments. Data analysis for particular telescopes, detectors, sensors, cameras, numerical simulations needed for data analysis and model predictions all rely on in-depth domain- and instrument-specific

knowledge. At the same time, research questions on the limits of applicability of laws of Nature, reconstruction of the picture of evolution of the Universe, on the challenges of environmental protection require broad expertise in multiple domains and necessity of efficient combination of voluminous / various / variable / veracious data. This is certainly impossible without availability of open and FAIR data. However, even if the data are available and findable, sharing of expertise and extraction of knowledge from FAIR data is not still possible unless the data are served together with **FAIR data analysis workflows**. The project will combine previous efforts of project partners toward implementation of FAIR principles not only to the data but also to the management of data analysis workflows.

Partners of this proposal believe that **opening of data *together* with data analysis workflows** and creating a holistic analysis framework provides the only viable approach for solution of the problem of full traceability / reproducibility of scientific results based on complex theoretical model assumptions and relying on increasingly large variety of data. Such a framework has also huge potential to boost knowledge transfer through massive reuse of data analysis methods. It also opens the way to proliferation of automated knowledge generation based on artificial intelligence technologies, which is critical for full exploitation of the research and discovery potential of scientific big data.

1.1 Objectives

Main objective of the proposed EU−MMO project is to build a system that connect theorists, simulators and observers / experimentalists and allow them to analyze data collaboratively and obtain synthetic results. To this end, we will create an **intelligent network of cloud-based data analysis services** that are

i capable of processing big data

ii fostering proliferation of intelligent cloud-based automated tools of data analysis,

iii learning efficiently from experts using the data analysis services and from publications reporting results obtained via data analysis deployed through the services,

iv enabling on-the-fly multi-messenger and multi-probe data analysis linking multiple messenger and probe specific cloud-based automated tools of data analysis and

v fostering reproducible science by holistically linking high level scientific results to the raw data sources, underlying theoretical model assumptions and analysis parameters, allowing on-the-fly re-generation of results and linking holistic data analysis chains to publications reporting the results.

Our goal is to take important steps in this directions and to demonstrate immediate benefits of this approach by boosting the **data-driven research** in the domains of multi-messenger astronomy, cosmology and environmental science. This innovative approach using cloud-based services to combine theory, data and simulations within holistic end-to-end pipelines based on FAIR data and data analysis workflows will enable us to gather critical mass of scientific users of the EU−MMO service network and promote it as a tool enabling ground-breaking discoveries and deep insights in multiple science domains.

Particular EOSC-related objective of the project will be to **close a number of gaps in Storage and Data and Thematic service offering** of EOSC-hub. The EOSC-hub service offering currently has a range of Storage & Data services which concentrate on data storage services compliant with FAIR principles, via EUDAT and EGI. These services assure the FA parts of the FAIR principles. However, there is no service which explicitly assures the IR aspects of the FAIR principles. Interoperability and Reusability imply the possibility of integration of the data and data analysis tools in data analysis chains of multiple research projects not necessarily identical to the original project which initially generated the data. This is possible only through efficient ”opening” of data analysis workflows and wide sharing of the expert-level expertise on their usage. Yet, there is no service explicitly implementing FAIR-principle-compliant data analysis workflows. The proposed project will close this gap by upgrading generic Storage&Data service offering to Storage & Data & Data-Analysis service offering. Apart from generic services, thematic service offering in EOSC-hub does not currently include astronomy, cosmology and Earth observation domain services. EU−MMO project will also close this gap in EOSC-hub service offering.

Our strategy will be to achieve the objectives via step-by-step implementation of EU−MMO service network that will develop in the bottom-up use-case driven manner and grow to a network of services useful in broader Open Science context. The EU−MMO platform will first develop Thematic services of EOSC-hub for domain-specific applications in three different science domains: multi-messenger astronomy, cosmology and Earth observations. We will use the focused domain specific developments to compare different possibilities and identify optimal, efficient and convenient approaches to tackle generic Open Science challenges:

- The challenge of ***combining narrow domain specific expertise with broad overview*** in the context of interdisciplinary research and inter-domain knowledge transfer. To address this challenge, we will implement universal tools for exploration, understanding, and reuse of cloud-based data analysis workflows.
- The challenge of ***interoperability of data and analysis service providers***. We will implement interfaces and data format adapters and tools for deployment of data analyses specified in standardized way across EOSC-linked computing infrastructures.
- The challenge of ***data and analysis results preservation*** will be addressed by storing detailed provenance and scientific context information encoded in data analysis workflow metadata within EU−MMO platform Knowledge Base which will serve as a basis for the in-built intelligence of the platform services.
- The challenge of ***automatic knowledge generation*** allowing to shift the focus of scientific work from routine data analysis to high-level analytical tasks. It will be addressed by implementing an intelligent service for automated data discovery and data analysis deployment in response to changing scientific context.

The strategy will be implemented via step-by-step development pursuing the following specific objectives:

1. Develop a EU−MMO framework for sharing data analysis workflows and their on-the flight deployment on the Cloud;
2. Link multi-messenger / multi-probe / multi-wavelength astronomical, cosmological and Earth observations data to EU−MMO platform framework;
3. Port large data analysis workflows extracting knowledge from these data into cloud-based EU−MMO platform environment.
4. Implement FAIR principles for management of data analysis workflows by generation, storage and analysis of workflow structure and usage metadata.
5. Generate scientific context metadata for EU−MMO linked data and analysis workflows by extracting it from scientific publications reporting results of data analysis.
6. Use workflow structure, usage and scientific context metadata to create a service which implements the holistic approach to complex multi-instrument data analysis management, traces provenance of scientific results and credits creators of data and analysis methods.
7. Implement a service to link deployable workflows together with workflow structure, usage and scientific context metadata to publications to enable larger scrutiny by scientific community.
8. Use workflow structure, usage and scientific context metadata to build up a Knowledge Base and use this Knowledge Base to power an intellignet Workflow Inference Engine able to advice users of EU−MMO services on the choices of workflows adjusted to scientific context of their research problems.
9. Use the Workflow Inference Engine within an automatic data analysis EU−MMO service responding to changing scientific context (transient events, new data acquisition, new analysis approaches) and alerting scientists upon identification of potentially interesting features in analysis results

1.2 Relation to the work programme

EU-MMO project aims at the Horizon-2020 work program and in particular at the INFRAEOSC-02-2019 call Prototyping New Innovative Services for the European Open Science Cloud (EOSC).

It will address the challenge of developing an ***agile, fit-for-purpose sustainable services*** which will start as ***domain-specific services*** for astronomy, astroparticle physics, cosmology and environmental science exploiting generic services of EOSC and supplementing them with a new ***generic services*** for management of FAIR data analysis workflows. These thematic and generic services will be promoted to powerful open science tools which will extend service offering for ***FAIR data management*** universally useful across a wide range of science domains.

The main innovative element of the EU−MMO services will be in creation of framework for FAIR analysis workflows which will supplement the FAIR data and possess an in-built intelligence which will assist users in finding data analysis workflow matching the needs of research questions. The EOSC-hub service offering currently includes a range of services which support open FAIR data creation and storage, but it lacks services which enable extraction of knowledge from the FAIR data and detailed tracing lineage of the data based on which scientific results are obtained. The data, even findable, accessible, and stored in format which makes them potentially reusable and interoperable could not really be used as long as they are not served together full lineage information and with FAIR data analysis workflows deployable in a way transparent to potential users.

The project will ***close a gap in service offering*** of the EOSC-hub by creation of an ***intelligent service for support of FAIR analysis workflows*** working on the FAIR data available through the EOSC services and deployable on computing resources provisioned through further EOSC-linked computing services. The findability of the workflows will be assured by the in-built intelligence based on the Workflow Inference Engine which will assist users in identification, composition and deployment of data analysis appropriate for particular science research question contexts. The in-built intelligence of the will be based on machine learning applied to a new type of workflow metadata associated to their structure, usage contexts and statistics as well as scientific contexts in which the workflows are used. The in-built intelligence of EU−MMO services will exploit new innovative techniques of analysis of human- and machine-readable scientific texts to accumulate data analysis expertise not only from human experts, data analysis manuals and online users of EU−MMO services but also from publications in which the results of the data analysis are reported.

The data of experimental and observational facilities have to be curated / analysed at all stages of research data lifecycle, from collection of raw data through the active data analysis stages up to the long-term data preservation upon completion of observations and experiments. The EU−MMO services will ***address all aspects research data lifecycle***. They will start from near-real-time management of e.g. multi-wavelength and multi-messenger astronomical data on transient sky events (starting from seconds after the generation of data by telescopes and astronomical messenger detectors), assist scientists in generation of publishable scientific results with near-real-time and consolidated archival data, assure full traceability / reproducibility of those results by storing and publishing full metadata on the data lineage, use this information within the Workflow Inference Engine to foster sharing and re-use of data and data analysis methods in the context of multi-messenger analysis, assure long-term preservation of data together with data analysis workflows deployable on the cloud in a way transparent to users and preserving FAIRness of both data and data analysis tools on long time scale.

The Workflow Inference Engine will implement ***intelligent linking and discovering of all research artefacts*** via implementation of holistic data analysis approaches in a variety of research problems, from multi-messenger astronomy of fast transients on the sky to multi-probe studies of evolution of protected ecosystems on Earth. The EU−MMO services will ***initially respond the urgent needs of communities of astronomers, astroparticle physicists, cosmologists and environmental scientists***. Intelligent linking and discovering of all multi-messenger data on new types of transient and persistent astronomical sources such as gravitational wave sources is key for understanding violent physical processes involved in their activity. Intelligent linking of data and simulations providing diverse cosmological probes is key for understanding the nature of elusive dark matter and dark energy substances which are most important scientific challenges of cosmology. The EU−MMO services will also respond the needs of environmental protection community by providing an innovative, powerful and cost-efficient alternative solution to conventional conservation monitoring exercises, which are reactive in nature, slow and often expensive.

Parallel development of the astronomy / astroparticle physics, cosmology and environmental science use cases will demonstrate the ***scalability of the novel services*** which will be developed initially for a particular science domain and then leveraged to other domains. This approach assures that development of tools for data discovery and its linking will finally yield ***a service will be universally useful for EOSC based data and data analysis workflows*** for different science domains.

The development of the EU-MMO service will build upon technologies which are at ***Technical Rediness Level*** $\geq$ ***6***, starting from already developed service prototypes. This includes the operational collaborative data science platform Renku developed by EPFL/SDSC partner. This platform has in-built capabilities for on-the-fly deployment, sharing and extraction of metadata of data analysis workflows from different science domains. The prototyping activities also include online open data access and data analysis systems in the domains of astronomy and Earth observations: the ODA platform and data archive systems for space-based telescopes developed by UNIGE / Astronomy department, the LiMES platform for on-the-flight analysis of satellite images of the Earth developed by the UNIGE GRID-Geneva unit, the openuniverse.org platform developed by SSDC, AstroWISE platform for data analysis of VLST telescope data at UL, ScienceWISE platform for text analysis and mining of information from arXiv.org preprint server. These already operational services use well tested technologies, such as that of containers deployable on demand on cloud infrastructures or running in service mode, machine learning algorithms based on word2vec etc. The project will leverage on these prototypes and integrate them into interconnected service network linked to other EOSC services. This service network will be then promoted to a complete and fully operational EU-MMO services which will, by construction, serve different user communities: astronomers, astroparticle physicists, cosmologists and environmental scientists. This level of development will correspond to the ***Technical Rediness Level*** $\geq$***8***. The EU-MMO service will be integrated into EOSC-hub at ***medium integration level, complying with the FitSM service management standard***.

The EU-MMO services will ***have strong user orientation*** building up in a bottom-up way and driven by the urgent need of particular research communities of astronomers, astroparticle physisists, cosmologists and environmental scientists. It will ***support the objectives of Open Science*** by supplementing open data by operational open data analysis workflows indispensable for the extraction of knowledge from the open data. It will also foster reusability of open data powered by an intelligent Workflow Inference Engine which will use structure, usage and science context metadata attached to the workflows to help finding, using and composing workflows. The service will involve ***new communities of innovative actors*** (astronomers, astroparticle physicists, cosmologists, environmental scientists) into the use of EOSC.

1.3 Concept and methodology

(a) Concept

Describe and explain the overall concept underpinning the project. Describe the main ideas, models or assumptions involved. Identify any considerations of interdisciplinary approaches and use of stakeholder knowledge to enrich the existing EOSC service offering.

Massive digitization of scientific research leads to increasingly complex concepts and employs ever more sophisticated analysis methods. In order to tackle new scientific problems, researchers must possess specialized in-depth knowledge. These problems, on the other hand, demand broad expertise, covering many different aspects, methods, techniques of data analysis, sets of theoretical models based on sophisticated analytical or numerical calculations involving uncertain model parameters. Competition between narrow specialization and broad expertise creates a problem of traceability, reproducibility and re-usability of results reported in research papers, as details of complex numerical modelling, data analysis chains and data sets used are often omitted and thus, sharing of data analysis and theoretical / numerical modelling expertise is restricted.

With the proposed project we want to implement a concept of collaborative "holistic" data management platform that connects scientists with different research profiles: theorists, simulators and observers and working in different science domains, so that they could efficiently combine their expertise for most efficient extraction of knowledge from observational, experimental and simulation data and obtain synthetic results by combining narrow domain specific expertise, without loosing full control of data provenance, lineage and theoretical hypotheses involved in analysis and intepretation of results.

This platform will be based on a network of online services connected to the EOSC and providing access to FAIR data and data analysis workflows adjustable and transformable depending on particular scientific contexts. To this end, we will create an "Intelligent Interface" capable of

- processing big data and retaining full data lineage and provenance information for all data analysis results.
- absorbing efficiently the narrow domain and instrument specific expertise from the experts in each field,
- transferring expert knowledge to automated tools of data analysis working on open data distributed across EOSC-connected infrastructure,
- using the accumulated expert knowledge to assist users in identification of analysis tolls and composition of workflows suitable for particular research questions in particular science contexts.
- using the accumulated expert knowledge to automatically perform data analysis in response to changing context / environment and alert users on potentially interesting results

We will demonstrate immediate benefits of this approach by using the platform for particular applications such as multi-messenger studies of transient sources, tackling the challenging cosmological problems, such as that of the nature of the dark matter, getting new insights into evolution of protected natural sites on Earth. The power of this innovative approach of bringing together theory, data and simulations within holistic end-to-end cloud-based data analysis pipelines will enable us to gather the critical mass of information required to explore these questions at great depth, with the potential to make ground-breaking discoveries in different science domains.

The project will leverage on pre-existing developments by the project partners, which include domain specialists in the relevant fields of astronomy, cosmology, astroparticle physics, environmental scientists, data scientists and software developers in different fields. Different as they are, all these fields face the same needs of efficient combination of different types of narrow instrument-specific data analysis expertise within holistic multi-messenger, multi-probe, multi-wavelength research tasks described below.

(a).1 Multi-Messenger Astronomy and multi-probe cosmology Recent detections of first astronomical sources of gravitational waves and high-energy neutrinos marked the birth of a new field of multi-messenger astronomy which provides new tools to explore violent phenomena in astronomical sources (black holes, neutron stars, supernovae etc.) with the goal to understand limits of applicability of known laws of Nature.

The new multi-messenger astronomy field combines data and analysis methods from different science domains (astronomy, particle physics, precision metrology). Analysis of each type of data relies on in-depth domain- and instrument-specific knowledge. At the same time, research questions of multi-messenger astronomy involve broad expertise in multiple domains and necessity of efficient combination of voluminous / various / variable / veracious observational data and numerical simulations which rely on theoretical models to be tested, on assumptions about properties of astronomical sources emitting astronomical messengers and detectors and telescopes registering them. The pace of progress of the new astronomy field depends on response to the challenge of combining narrow domain specific data analysis expertise within a broader context of multi-instrument multi-messenger analysis techniques and of ever more sophisticated theoretical modelling which relies on extensive numerical simulations.

The field of cosmology is about to be revolutionised by deployment of massive automated sky surveys enabled by the big data techologies. Using a combination of multiple survey data and powerful numerical simulations which model yet uncertain physical processes governing the Universe evolution cosmologists are now able to infer the entire picture of the Universe all the way back to the Big Bang moment. Within this picture, important role is played by dark matter and dark energy, the invisible substances which are not described by the laws of Nature as they are known to us. Further progress in the field depends on the efficiency of combination of narrow data analysis expertise related to each individual type of observational data and simulation analysis within a broad holistic multi-probe framework.

In both multi-messenger astronomy and cosmology such combination of expertise is possible if data and data analysis tools are widely available for sharing and reuse across the research community. Progress in this direction is already underway. Instrument specific (e.g. optical, X-ray and gamma-ray telescope data archives

maintained by the project partners UNIGE[1], SSDC[2], PS[3]) and multi-instrument astronomical data archives (like e.g. HEASARC[4] high-energy astronomy data archive maintained by NASA) are serving data in common interoperable formats like FITS or VOtable commonly accepted by astronomical data analysis systems. The end-products of data analysis are made FAIR by numerous services employing standards developed by Virtual Observatory[5] consortium. In the particular case of transient multi-messenger sources, data and results of data analysis are made available to community of astronomers via rapid nano-publication services like Astronomer Telegrams (ATEL), Gamma-ray Coordinates Network (GCN) and Virtual Observatory Events (VOevent) which are important part of information and expertise sharing in the field.

Contrary to the data sharing, efficient sharing of in-depth data analysis expertise is not yet fully assured. Even if the data analysis toolkits are most often freely available via public software distribution systems, instrument-specific data analysis still requires specific computing environment, knowledge of instrument calibrations, analysis parameter adjustments which are not universally shared. Project partners are developing approaches to make data analysis workflows compliant with FAIR principles and enable their reuse by scientists who are not necessarily specialists in details of data analysis of particular instruments, but need to use these instrument-specific analysis within broader multi-wavelength astronomy research problems. Prototypes open FAIR workflow platforms are implemented by UNIGE partner (Online Data Analysis system[6]) SSDC (multi-mission archive4, OpenUniverse[7]), PS (GAVIP[8] data analysis system for GAIA telescope).

These platforms use cloud computing approach which enables implementation of the FAIR principles for astronomical data analysis workflows via online services within which analysis workflows are deployed in a way transparent to the user. These services accumulate necessary narrow instrument-specific expertise provided by teams running the services. EU-MMO platform will aggregate existing and deploy new online instrument specific data analysis services and integrate them into a platform which will enable efficient finding, accessing, composing, deploying and preserving instrument specific and complex multi-messenger data analysis workflows.

(a).2 Earth observations. Human activities such as extractive industry, food and energy production, transportation, tourism, urbanization, with their corresponding use of resources and pollutions, are rapidly transforming the environment. One of the urgent challenges is assessment of human impact on different ecosystems and their protection. International organisations including United Nations Environment Program (UNEP), the Ramsar Convention and the International Union for Conservation of Nature (IUCN) World Heritage, World Wildlife Foundation (WWF) monitor thousands of sites worldwide, covering hundreds of millions of hectares. The scattered distribution, the size and the number of sites is posing a monitoring challenge for organizations in charge of their monitoring, leading to low frequency of updates on the status of these sites. More regular tracking is essential.

The data revolution generated by new technologies (e.g., satellites, mobile devices, cloud computing, crowdsourcing, physical and chemical sensors, linked data) provides an exponentially increasing volume and variety of data which could be used for monitoring of ecosystems. Qualitatively new insight could be obtained by combining these multi-messenger data on Earth observations in the context of physical, chemical and biological conditions of the Earth system.

Remote sensing techniques allow monitoring environmental changes over large areas. The adoption of broad open data policies allows users to freely access low to medium-resolution imagery such as Landsat, MODIS, ASTER, Sentinels satellite data. Similarly to the astronomy domain, the data are made FAIR-principle-compliant via common standards developed e.g. by the Open Geospatial Consortium (OGS) . Sharing of data instrument-specific analysis expertise is possible via implementation of FIAR-principle-compliant data

[1] https://www.astro.unige.ch/cdci/integral-archive , https://www.astro.unige.ch/cdci/cta-sst1m-archive
[2] https://www.ssdc.asi.it/mma.html
[3] http://docs.gavip.science
[4] https://heasarc.gsfc.nasa.gov
[5] http://www.ivoa.net
[6] https://www.astro.unige.ch/cdci/astrooda_
[7] http://openuniverse.asi.it
[8] http://docs.gavip.science

analysis workflows. UNIGE project partner, through its UNEP/GRID-Geneva partnership , has developed Live Monitoring of Earth Surface (LiMES) system which for assessing human impacts in protected areas, which implements online data analysis of satellite images. Similar developments for online data analysis services in the Earth observation domain are taking place worldwide.

Development of multi-wavelength and multi-probe approaches to analysis of diverse environmental data sets to efficiently extract relevant information on the Earth ecosystem parameters for answering scientific questions as well as for supporting decision making processes represents a major challenge for the Earth observation community. Even though numerous services are providing data products to expert users, they remain still difficult to handle for tracking, monitoring, visualizing, understanding, and communicating within a broad multi-messenger assessment of environmental changes. EU−MMO platform will provide solution to this challenge by using cloud technology to integrate online data analysis services which and aggregate them within an intelligent interface allowing efficient knowledge transfer.

(a).3 FAIR data analysis workflow analytics Progress in different fields introduced above depends on efficient sharing and combination of data analysis expertise within multi-messenger, multi-probe, multi-wavelength research contexts. This is possible within a framework implementing "holistic" approach in which full information on the data provenance, transformations, on model assumptions and analysis parameters is consistently traced in a transparent way from the raw data up to the multi-messenger / -probe / -wavelength scientific results ready for publications and presentable to public.

A necessary condition for implementation of the holistic approach is that large number of instrument-specific workflows and generic data analytics tools should be available, findable and executable preferentially within a federated environment which provides interoperability between different analysis tools.

However, availability of deployable and executable analysis workflows for many instruments could not by itself provide solution to the problem of sharing and combination of data analysis expertise. Data analysis for each instrument still typically requires a dedicated expert who knows details about calibration procedures, keeps track of anomalies, knows limitations of data taking etc. The instrument-specific expertise confined to research groups running the detector or telescope is not easily shareable. External users would not master refined analysis details, even if they are provided with access to the data analysis tools. This limits the potential of the multi-messenger, multi-probe, multi-wavelength approaches in all the three domains described above.

Development of tools for efficient sharing of the specialized in-depth data analysis expertise via a combination of human expertise with machine intelligence is the way to solve the narrow analysis expertise transfer challenge. Such tools are prone to boost research in the problems requiring holistic approaches, be it multi-messenger view on astronomical sources, understanding of the history of evolution of the Universe or understanding of complex environments in Earth ecosystem.

Use of machine intelligence augmented system of sharing of narrow instrument specific expertise is also prone to provide a breakthrough in the issue of larger scrutiny of scientific results by broad scientific community and by society. This is an essential part of scientific research on the limits of applicability of the laws of Nature and an equally important element of trust in results analysis of the implications of human activities on the our planet.

EU−MMO will accumulate narrow instrument-specific data analysis expertise via collection of metadata on workflow usage by the experts. The usage cases include the online users deploying the workflows in the federated environment of EU−MMO platform as well as use cases reported in scientific publications giving the details of data analysis in the context of specific research questions. Systematic extraction of metadata on the workflow usage and its scientific context has to be done within appropriate ontology system structuring metadata. Project partner LA has accumulated significant experience in knowledge extraction from publications and classification of metadata within ontology frameworks by conceiving and running ScienceWISE[9] service which analyzes publications posted on arXiv.org preprint server. The EPFL/SDSC partner team uses the Common Workflow Language[10] (CWL) developed at CERN to annotate analysis workflows embedded into collaborative science

[9] http://sciencewise.info

[10] https://www.commonwl.org

Renku platform[11].

(a).4 The main concept underpinning EU−MMO platform development will be that of ***FAIR analyfsis workflows*** which make the FAIR data "alive" and allow tracing their origin, re-generation of results, re-use of data across multiple applications and automatic knoweldge generation. ***Online data analysis serivces*** will assure accessibility of workflows via their deployment in EOSC-connected federated computing environment. A ***Knowledge Base*** containing the corpus of metadata on workflow structure, usage and science context will assure findability of the workflows. An ***intelligent Workflow Inference Engine*** will assure interoperability of workflows and provide possibilities for construction of holistic multi-messenger, multi-probe, multi-wavelength analysis chains. A network of ***cloud-based services*** using the workflow Knoweldge Base and the Inference engine will provide reusability of diverse workflow components extraction of knowledge from the FAIR data within multiple multi-messenger, multi-probe, multi-wavelength research projects.

We will implement these concepts for the first time and in this way demonstrate the potential of Open Science boosted by massive availability of open data and open data analysis tools.

(b) Methodology

Describe and explain the overall methodology, distinguishing, as appropriate, activities indicated in the relevant section of the work programme, e.g. for research, demonstration, piloting, first market replication, etc.

(b).1 EU−MMO platform. The concepts on which our approach to FAIR analysis workflows is based will be implemented within a network of smart cloud-based services which will address the challenges of multi-messenger and multi-probe research in the domains of astronomy, cosmology, environmental sciences. These services will also provide a viable solutions to the problems of traceability, reproducibility and re-usability of results reported in research papers by exposing the holistic end-to-end data analysis chains behind the results. The project team will pursue a bottom-up approach to service development, responding to specific needs of different science domains: multi-messenger astronomy, cosmology, Earth imagery for environmental monitoring. The domain specific applications will drive the development of common use EU−MMO platform linking domain-specific and generic cloud-based services into a unified framework.

Platform interfaces. The platform will provide a web-based federated work environment adapted for different applications. The environment will include interfaces to human and automatic users.

Online Data Analysis web interfaces (virtual desktop embedded into web browser) will assist expert and non-expert level users in creation and deployment of valid instrument-specific and multi-instrument data analysis workflows. The virtual desktop will provide tools of visualisation and manipulation of high-level data products which are outputs of data analysis workflows.

Application Programming Interfaces (API) will provide access the workflows remotely and integrate them into analysis activities of users worldwide and with other astronomical and environmental science data services.

Online Publication Analysis interfaces to scientific publication services (preprints of research papers posted on arXiv.org, nano-publication services such as Astronomers Telegram (ATEL), Gamma-Ray Coordinates Network (GCN), Virtual Observatory Event (VOEvent)) will allow to link data analysis workflows to publications or extract data analysis workflows from publications.

Analysis Knowledge Base. The user interfaces will be connected to the platform core which will capture human and automatic user activity. It will collect metadata on each user-defined workflow and generate representation of the workflow in the EU−MMO Knowledge Base. The workflow elements metadata will be generated following certain ontology of data analysis methods and elements within the Knowledge Base of EU−MMO.

Such metadata organisaiton will be suitable for efficient finding and sharing of the analysis instances with different parameter choices captured during execution within EU−MMO environment. The generic Common Workflow Language (CWL) annotation of workflow elements, input and output data will be supplemented

[11] https://renkulab.io

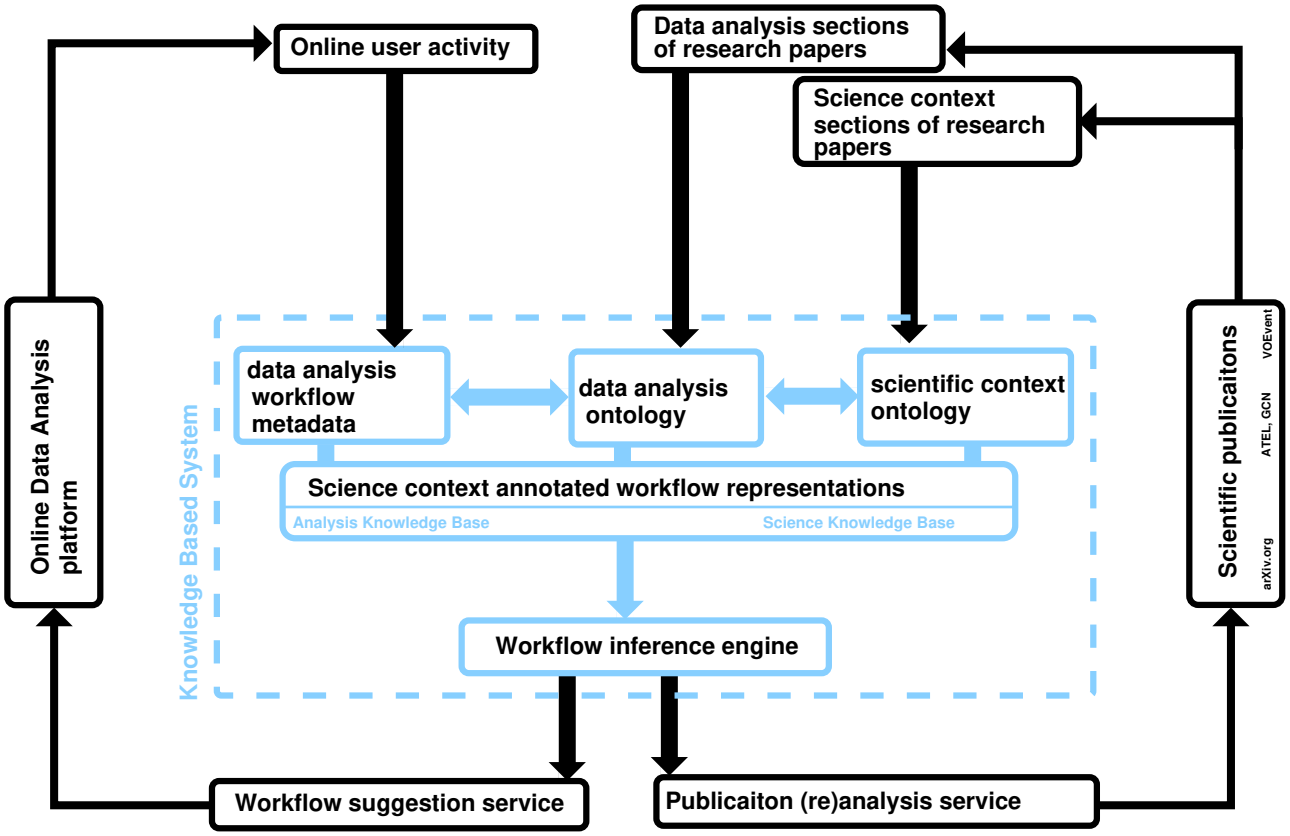

Figure 1: Elements and interfaces of the EU−MMO Knowledge Based System

domain-specific metadata (e.g. input corresponding to time interval / energy bounds, nodes performing imaging, spectral extraction etc).
Workflow metatata will also include analysis parameters (for example, exposure time) along with information on studied objects (for example, astronomical source flux listed in a catalog). This form of knowledge representation could be used e.g. for the search of patterns which encode types of analysis appropriate for particular types of objects.

Science Knowledge Base. The interface to scientific publications will allow users to explicitly link workflows and results of data analysis to the publications. Users will be encouraged to do this to assure full reproducibility of the results, their traceability to the raw data and correct referencing of data sources and creators of data analysis tools. EU−MMO services will capture and analyze each instance of workflow-results-publication link to generate science context metadata on the workflows.

To this end, we will use keyword extraction technique to analyze the linked publications to extract information on the descriptions of data analysis techniques and their relation to the scientific contexts considered in publications. These linked science context and analysis description metadata will be added to the Knowledge Base.

Intelligent Workflow Inference Engine will use the expertise accumulated in the Knowledge Base to develop in-built intelligence which allow scientists to identify and compose data analysis workflows relevant for a specific scientific context.

We will apply machine learning techniques to find patterns in the corpus of workflow metadata harvested from the online usage and publications. These patterns will be classified and ranked. Ranking and scoring of particular workflows, particular parameter choices will depend on the science context.

The classified and ranked content of the Knoelwdge Base will be used to power an engine which will find workflow elements in response to requests of human and automatic users and help users in composition of viable workflows adjusted to the particular scientific context of their research. It will

- provide interactive help / suggestions / validation environment for creation of new workflows based on

learning from best practices of usage of known workflows, consistency with data analysis manuals and relevance to scientific context;

- Assist in re-use of existing analysis blocks for composition of novel workflows in multi-messenger context, to facilitate interdisciplinary knowledge transfer;
- Assist in annotation of publication linked workflows with scientific context metadata to generate knowledge base of data analysis workflows via crowd sourcing approach.

The EU−MMO Knowledge Base and Workflow Inference Engine will be used by the domain specific and generic services integrated with the platform.

(b).2 Multi-Messenger astronomy services. Integration of astronomical data analysis into EU−MMO will leverage on existing developments by the project partners: ODA platform developed by UNIGE, OpenUniverse platform developed by SSDC, GAVIP interface by PS, AstroWISE system developed by UL. to complete... We will integrate these systems with EU−MMO platform and enable their interoperability by standardising the interfaces and providing tools to annotate instrument-specific data analysis workflows with the usage and science context information to generate workflow metadata which will be included in the analysis Knowledge Base.

We will use the iWIE to power *Online Data Analysis services* of EU−MMO. It will help to suggest data analysis workflows suitable for particular needs of ODA users and help them in composition of new multi-messenger and multi-probe data analysis chains in the following benchmark use cases:

UC1 Exploration mode of EU−MMO web interface:

- User provides a name of an astronomical source or specifies a sky region in EU−MMO web interface.
- A source or region of interest are found in the Knowledge Base. A Knowledge graph of all known source / region related information is generated. This knowledge graph is presented to the user.
- The iWIE finds data and data analysis workflows compatible with the source / region type and location on the sky / Earth. This information is presented to the user.
- The user has a possibility to deploy presented workflows, repeat the analysis with modified parameters, work on further analysis of the presented data or refine the query by entering additional parameters.

UC2 Help mode of the EU−MMO online data analysis interfaces:

- User provides complete set of parameters for the ODA web interface for astronomical data analysis, for particular source and particular telescope / detector.
- The workflow defined by these parameters is mapped into the Knowledge Base.
- The iWIE checks if the workflow mapping into the Knowledge base is within the pattern of workflows used for the specific source type provided in the input and signals if the parameter choice is an outlier.
- If the parameter choice is an outlier, the iWIE suggests viable workflows closest to the user input and exposes the user to the choice of the workflows together with their scientific contexts.
- The user has a possibility to adjust the input parameters depending on the science context of his/her analysis.

We will also integrate the Knowledge Base and iWIE into Online Publication Analysis services.

A *service for transient sources* will use analysis of publications to harvest the data analysis and scientific context information from well-structured ATEL and GCN Circular human-readable, as well as GCN Notice and VOEvent machine readable publications, which typically are short notes on remarkable events, like gamma-ray or gravitational wave bursts, flaring activity of black holes and neutron stars etc. We will process these publications to extract science concepts and data analysis information from them and express it in terms of

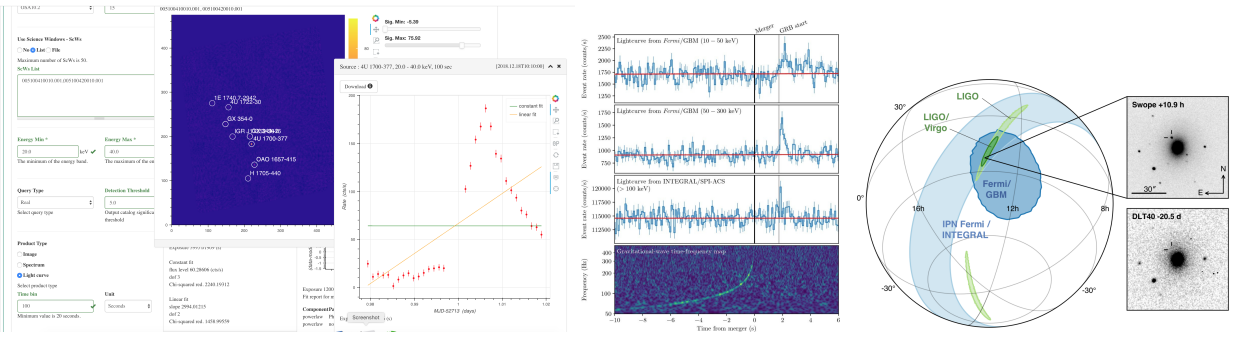

Figure 2: Left: the workspace of ODA web interface. Middle and right: example of "prompt" observational data on transient event GW170817 (neutron star merger) reported via astronomical nano-publication services. EU−MMO transient service will provide on-the-flight analysis and ODA web interface will provide an interactive web browser based analysis and visualisation environment for such data.

specific Knowledge Base representation, use iWIE to construct the matching executable workflow and deploy this workflow to reproduce the results reported in the nano-publication. The results (astronomical source detections, their spectral characteristics, variability properties) reported in the nano-publications will be compared with the re-analysis derived from the combined EU−MMO Knowledge Based System. Compatibility of the results will validate the performance of the iWIE ability to suggest data analysis methods suitable for particular source classes and analysis goals. At the next stage, we will set up a service which will respond to new ATEL/GCN/VOEvent publications by interpreting the science context of the reported events, verifying the reported results with re-analysis of the observational data (if the data is publicly available), launching additional data analysis relevant in the scientific context of the reported event (e.g. characterising historical activity of a bursting / transient source) and publishing the outcomes of this additional analysis in human and machine-readable ATEL/GCN/VOevent nano-publications.

The data extracted from the publications and the results of eummo data analysis will be visualised via an interactive web interface of ODA on event-by-event basis (Fig. 2), to allow further refinement of analysis parameters and extraction of multi-messenger science results.

A *service for reproducible science* will use iWIE to extract the details of data analysis from publications and compose data analysis workflows based on the inferred details. Scientific publication (online nano-publication, arXiv preprint, journal article) will be scanned to identify the publications reporting results of data analysis of telescopes / detectors known to EU−MMO. The service will extract science context keywords and analysis details the from publication. This information is mapped into the Knowledge Base. The iWIE ranks the known workflows and identifies the workflow most closely matching the workflow described in the publication. The best ranked workflow is deployed in EU−MMO federated computing environment. The results of the data analysis will be compared with those reported in the publication. Comparison of EU−MMO based data analysis results with the published results will be used for training of the iWIE. If the re-analysis results by EU−MMO are different from published results, further investigation will be initiated to to determine the reasons for discrepancies. If they are due to incomplete or erroneous information present in the article, an alert will be issued.

(b).3 Earth Observation services The integration of Earth Observations data analysis into EU−MMO will leverage on the LiMES platform developed by GRID-Geneva. LiMES pre-processes raw satellite images to a level where even people who are not remote sensing experts can use these images for accurately assessing changes. It performs automated remote sensing processing (accessing the data, atmospheric corrections, conversion to reflectance, pan-sharpening as well as indexes, see Figure 3) thus allowing to process larger amount of data and to provide new products to non-remote sensing experts.

With increased number of freely accessible High Resolution satellite sensors, we can obtain more data offering higher spatial, spectral and temporal resolutions. The Landsat archives offer free access and global coverage since 1972. The recent launch of the ESA Sentinel satellites, as well as the opening of free access to ASTER imagery, are providing additional free monitoring capabilities. Dealing with such a wealth of data is now

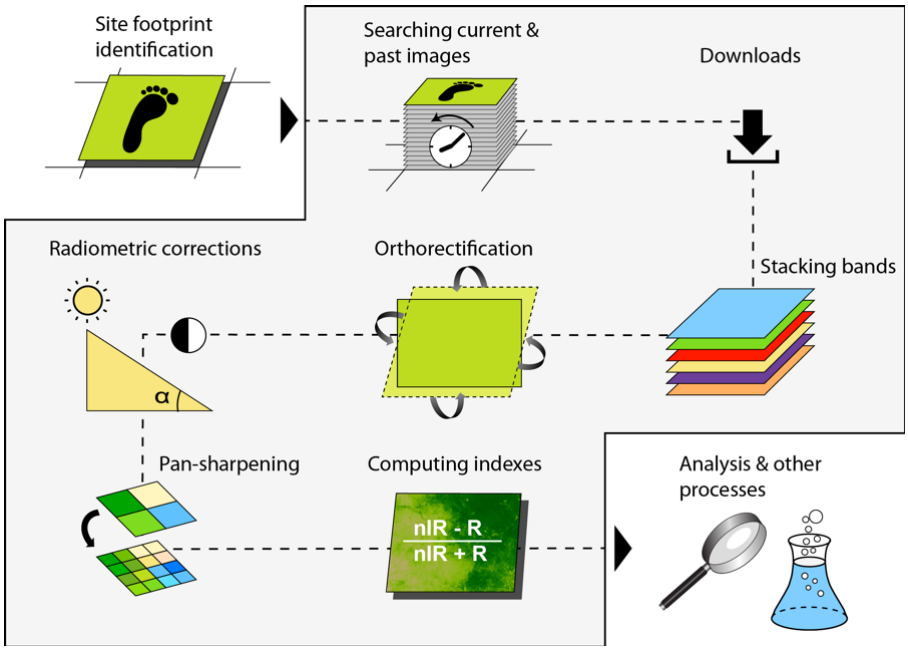

Figure 3: LiMES platform scheme.

technically feasible, but requires extensive automation. Indeed, the increasing spatial resolution of remote sensing images as well as the effort and cost required to convert Earth Observations data into meaningful geophysical variables has hampered a systemic analysis to monitor changes from this Big Data archives.

During the first phase of the project, the available sattelite imagery data and the analysis workflows of the LiMES platform will be integrated into the EU−MMO by standardising the interfaces and providing tools to annotate instrument-specific data analysis workflows with the usage and science context information to generate workflow metadata which will be included in the analysis Knowledge Base.

These metadata will be used within the Knowledge Based System to identify and deploy analysis workflows adjusted to the specific science context of user requests through LiMES service integrated into EU−MMO. This will enable efficient extraction of relevant information to answer scientific questions as well as sustaining decisions making processes.

The user interface will allow users to define a study area the service will obtain the relevant satellite images, do the atmospheric correction, transform the signal into reflectance, and generate the requested products (e.g., pan-sharpened images, NDVI, and various other indexes). The output is a time series of images and statistics, which can be used to assess trends in status and level of protection of a selected site (such as given in Fig. 4).

The services will enable people without high expertise in remote sensing to produce an assessment of the status of the 2,240 Ramsar Sites the 229 World Heritage Sites and sites of interest for UN-REDD. The results will be integrated into the existing websites of these organizations, such as the UNEP World Environment Situation Room, Ramsar Sites Information Service, IUCN World heritage websites. The results will also be shared through webservices so it can be widely accessible, e.g. via the Group on Earth Observations (GEO).

Gender considerations TBC

The EU−MMO platform will develop services within the open framework of EOSC. These services will respond to the needs of the international research community in its *entirety*, and it is therefore absolutely essential that the project embraces and promotes the principles of gender equity. The project team will adopt the London Code of Conduct its meetings (https://github.com/apontzen/london_cc) which is adapted by many facilities contributing to the project.

Could there be some activities explicitly promoting gender balance?

Please note that this question does not refer to gender balance in the teams in charge of carrying out the

Figure 4: Extract from LiMES: automatic images processing (Normalized Difference Vegetation Index) over the Ramsar Site: Gulf of Guayaquil (Ecuador) where mangroves have been steadily converted to shrimp aquaculture ponds.

project but to the content of the planned research and innovation activities . Sex and gender analysis refers to biological characteristics and social/cultural factors respectively. For guidance on methods of sex / gender analysis and the issues to be taken into account, please refer to http://ec.europa.eu/research/swafs/gendered-innovations/index_en.cfm?pg=home

1.4 Ambition

Describe the advance your proposal would provide beyond the state-of-the-art, and the extent the proposed work is ambitious.

Explain how the proposed work is beyond the state of the art, builds on technologies at TRL 6 or above, and demonstrates innovation potential (e.g. ground-breaking objectives, novel concepts and approaches, new products, services or business and organisational models).

EU−MMO platform will engage large community of astronomers, cosmologists and environmental scientists into innovative approach to research using intelligent cloud based services for data analysis tasks. This will open the way to wide spreading and sharing of data analysis expertise between researchers working in the same science domain and also for the inter-domain expertise transfer. It is prone to become the dominant data analysis tool for the new generation of researchers.

The project will address an important challenge of the open science based on FAIR data: that of efficient extraction of knowledge from the open data which is only possible with open data analysis workflows which are aware of refined detail and peculiarities of data analysis. It will do this by creating for the first time a framework for implementation of the FAIR principles for the analysis workflows. Such ambition could be realised only thanks to the development of cloud computing technologies and the project aims at demonstrating the power of this technology. Establishing a network of online cloud-based services which provide possibilities for on-the-fly data analysis of large variety of instruments from different science domains and possibilities for combining instrument specific data analyses within holistic multi-messenger and multi-probe research workflows will provide an important contribution to the open science research environment promoted by EOSC.

Domain-specific ambitions

Multi-messenger astronomy and cosmology

Environmental sciences

Data science

Generic science ambitions. describe ambiiton for reproducible research linking analysis to publications, holistic approach etc

Social ambitions. EU−MMO platform has an ambition to introduce new cloud-based collaborative research culture to specific research domains. It will do so by training and educating users about innovative approaches to data stewardship and demonstrating immediate benefits of these approaches which will boost the discovery potential of observational, experimental and simulaiton data combined within multi-messenger and multi-probe contexts. The EU−MMO team will also give important consideration to communicating benefits of these new innovative approaches also outside the project to industry, SME and general public. Its services will be open to citizen science, especially important in the environmental preservation context.

Innovative potential. Exploring new approaches to FAIR data management and open science, consolidating standards of interoperability of data and service providers, combining hardware, software and middleware within a platform running agile, fit-for-purpose services boosting the discovery potential of different science fields naturally yellds large terrain for innovation potential important in the general context of on-going digitization of society in particular context of EOSC development and of the open data driven Open Science.

2 Impact

2.1 Expected impacts

Please be specific, and provide only information that applies to the proposal and its objectives. Wherever possible, use quantified indicators and targets.

Describe how your project will contribute to o each of the expected impacts mentioned in the work programme, under the relevant topic;

Describe how your project will enhance capacity for innovation and production knowledge.

Explain how you will the exploit the results of the project (including IPR), and more specifically how you will ensure the integration of services in the EOSC for the benefit of a broader set of user communities. Also, explain how you will enhance EOSC's potential to support multidisciplinary research, manage research data where relevant and effectively disseminate and communicate the project activities and results to appropriate target audiences.

Describe any barriers/obstacles, and any framework conditions (such as regulation, standards, public acceptance, workforce considerations, financing of follow-up steps, cooperation of other links in the value chain), that may determine whether and to what extent the expected impacts will be achieved. (This should not include any risk factors concerning implementation, as covered in section 3.2.)

Integrating co-design into research and development of new services to better support scientific, industrial and societal applications benefiting from a strong user orientation.

The main impact of the EU−MMO project will be on the research domains of multi-messenger astronomy and environmental science. It will bring significant reduction in the fragmentation of data analysis expertise skills to which researchers are confronted and which limits the data driven scientific knowledge generation in these fields. Development of FAIR cloud-based analysis services, of interoperability between different data analysis service providers and of collaborative platform allowing to combine different data analyses coming from different providers within holistic multi-messenger and multi probe approaches will remove the limitations of fragmented data analysis experise and uncover new research potential of the data pushing the limits of our knowledge of the laws of Nature and of the Earth environment.

Beyond this domain-specific impact, EU−MMO services will have a broader reach in the context of on-going digitization of all sides of human activities. This impact will be through creation of innovative tools which supplement FAIR open data with FAIR open analysis workflows, without which extraction of knowledge from the open data is not possible.

Supporting the objectives of Open Science by improving access to content and resources, and facilitating interdisciplinary collaborations.

Details...

Fostering the innovation potential by opening up the EOSC ecosystem of e-infrastructure service providers to new innovative actors.

Details...

2.2 Measures to maximize impact

(a) Dissemination and exploitation of results

Provide a draft plan for the dissemination and exploitation of the project's results. Please note that such a draft plan is an admissibility condition, unless the work programme topic explicitly states that such a plan is not required.

Show how the proposed measures will help to achieve the expected impact of the project.

The plan, should be proportionate to the scale of the project, and should contain measures to be implemented both during and after the end of the project. For innovation actions, in particular, please describe a credible path to deliver these innovations to the market.

Your plan for the dissemination and exploitation of the project's results is key to maximising their impact. This plan should describe, in a concrete and comprehensive manner, the area in which you expect to make an impact and who are the potential users of your results. Your plan should also describe how you intend to use the appropriate channels of dissemination and interaction with potential users.

Consider the full range of potential users and uses, including research, commercial, investment, social, environmental, policy-making, setting standards, skills and educational training where relevant.

Your plan should give due consideration to the possible follow-up of your project, once it is finished. Its exploitation could require additional investments, wider testing or scaling up. Its exploitation could also require other pre-conditions like regulation to be adapted, or value chains to adopt the results, or the public at large being receptive to your results.

Include a business plan where relevant.

As relevant, include information on how the participants will manage the research data generated and/or collected during the project, in particular addressing the following issues:

- What types of data will the project generate/collect?
- What standards will be used?
- How will this data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why.
- How will this data be curated and preserved?
- How will the costs for data curation and preservation be covered?

What types of data will the project generate / collect? What standards will be used? First source of data which will be used by EU−MMO services are astronomical, cosmological and Earth observational data and data analysis software. These data and software systems are distributed by the respective Data Centers (for example, INTEGRAL Science Data Centre operated by UNIGE partner, AstroWISE system operated by UL partner). The observational data are typically available in structured according to the data sharing standards, such as those defined by Virtual Observatory, and can also be queried through unified interfaces, for example through Centre de Donnes astronomiques de Strasbourg (CDS) or HEASARC (for NASA missions) and ESAC (for ESA missions). Exploiting these freely available astronomical data, diverse yet uniformly standardized, is at the foundation for success of our proposal.

The second essential source of data for this work are publications. The full-scale scientific articles posted on arXiv.org are publicly available and could be freely used as training corpus within the Knowledge Base. Nanopublications which are used for rapid communications about astronomical events, ATEL, GCN, VOEvent, are

also publicly distributed. An essential fraction of these nano-publications are fully machine readable, conforming, for example, to the Virtual Observatory standards. These nano-publications are structured conventiently for automated extraction of the relevant information. Automating sharing of the structured scientific knowledge has become a common practice in the transient astronomy domain, making this domain particularly suitable for the purposes of this proposal. Authors of the present proposal have especially extensive experience in this field.

Software for specialized astronomical analysis is usually public, and is distributed by the corresponding Data Centers. However, up to now there is no common standard for describing the data analysis workflow. All the relevant scientific software is designed primarily for Linux OS, making the the transition to portable containers (e.g. Docker [19]) fairly straightforward. Our team has extensive expertise in exploiting this software, proven by the track of scientific publications. Although data analysis systems of different telescopes / detectors / Earth imagery cameras use quite different approaches, we plan to consider standardization of scientific data analysis workflow description using e.g. CWL. These descriptions of data analysis workflows, once available, can be readily related to each other and to the Knowledge Based part generated from publications.

How will this data be exploited and/or shared/made accessible for verification and re-use? The project will create a range of tools which will implement FAIR principle not only to the open data available through EOSC services, but also to the data analysis workflows deployable through the EOSC-connected services. provide further detals

The appropriate structure of the consortium to support exploitation is addressed in section 3.3.

Outline the strategy for knowledge management and protection. Include measures to provide open access (free on-line access, such as the green or gold model) to peer-reviewed scientific publications which might result from the project .

Open access publishing (also called 'gold' open access) means that an article is immediately provided in open access mode by the scientific publisher. The associated costs are usually shifted away from readers, and instead (for example) to the university or research institute to which the researcher is affiliated, or to the funding agency supporting the research. Gold open access costs are fully eligible as part of the grant. Note that if the gold route is chosen, a copy of the publication has to be deposited in a repository as well.

Self-archiving (also called 'green' open access) means that the published article or the final peer-reviewed manuscript is archived by the researcher - or a representative - in an online repository before, after or alongside its publication. Access to this article is often - but not necessarily - delayed (embargo period), as some scientific publishers may wish to recoup their investment by selling subscriptions and charging pay-per-download/view fees during an exclusivity period

(b) Communication activities

Describe the proposed communication measures for promoting the project and its findings during the period of the grant. Measures should be proportionate to the scale of the project, with clear objectives. They should be tailored to the needs of different target audiences, including groups beyond the project's own community.

3 Implementation

3.1 Work plan – Work packages, deliverables

Please provide the following:

- detailed work description, i.e.
  - a list of work packages (table 3.1a);
  - a description of each work package (table 3.1b);
  - a list of major deliverables (table 3.1c);
- graphical presentation of the components showing how they inter-relate (Pert chart or similar).

Give full details. Base your account on the logical structure of the project and the stages in which it is to be carried out. The number of work packages should be proportionate to the scale and complexity of the project.

You should give enough detail in each work package to justify the proposed resources to be allocated and also quantified information so that progress can be monitored, including by the Commission

Resources assigned to work packages should be in line with their objectives and deliverables. You are advised to include a distinct work package on management (see section 3.2) and to give due visibility in the work plan to dissemination and exploitation and communication activities, either with distinct tasks or distinct work packages.

You will be required to include an updated (or confirmed) plan for the dissemination and exploitation of results in both the periodic and final reports. (This does not apply to topics where a draft plan was not required.) This should include a record of activities related to dissemination and exploitation that have been undertaken and those still planned. A report of completed and planned communication activities will also be required.

If your project is taking part in the Pilot on Open Research Data, you must include a 'data management plan' as a distinct deliverable within the first 6 months of the project. A template for such a plan is given in the guidelines on data management in the H2020 Online Manual. This deliverable will evolve during the lifetime of the project in order to present the status of the project's reflections on data management.

(a) Overall structure

brief presentation of the overall structure of the work plan; timing of the different work packages and their components (Gantt chart or similar);

EU-MMO project work will be divided onto 7 work packages (WP). The WP1 deals with the management, coordination, networking and dissemination tasks. WP2 and WP7 are devoted to the generic tasks of organisation and maintanance of computing environment of the EU-MMO platform and its integration with EOSC and to the mainanance of operational EOSC connected services created by the project. WP3-WP6 are devoted to the core innovative developments of the project: setup of online data and publication analysis services, their interoperability, combination and composition as well as implementation of the FAIR principles for the cloud-based analysis workflows. The project will be managed by the Management Team (MT) which will oversight progress of all WPs. . to be completed.

WP1 **Management, coordination, results dissemination (Management)** will perform the overall management of the project tasks, planning of work, coordination between project teams and will take care of dissemination of the project results.

Key Outputs: EU-MMO coordination and management. Implementation of interaction between MMO service and EOSC

WP2 **Development of MMO platform (Platform).** The task of this workpackage is to develop and integrate components of the MMO platform which will be the "backbone" of MMO activities. The development will build upon existing prototypes: generic Renku collaborative science platform developed by SDSC, ODA interfaces for astronomical data analysis developed by UNIGE/ASTRO and LiMES platform for Earth observaitons data developed by UNIGE/GRID.

Key Outputs: MMO platform elements:

- federated computing environment integrated with EOSC data storage and computing services.
- tools for standardization, interoperability and portability of data analysis workflows for different telescopes, astronomical messenger detectors, space-based Earth observation cameras
- web portal with Data Discovery and Online Data Analysis (ODA) services,
- services for deployment data analysis via API in response to remote requests by human users, astronomical and Earth observation data and publication services,
- services for automatic discovery and processing of data and in response to transient event alerts, upon discovery of new data or new versions of data analysis workflows

- reproducible analysis service for integration of data analysis workflows into scientific publications

WP3 **Astronomy and cosmology data (Astro-Data)** will be devoted to activities on porting the data and and data analysis tools of multiple telescopes, astronomical messenger detectors and associated numerical simulations

- radio, optical, X-ray, gamma-ray telescopes
- high-energy neutrino detectors
- gravitational wave detectors
- sky surveys analysis tools
- image, spectra and time series analysis tools
- numerical simulations for data and detector / telescope response modelling
- cosmological numerical simulations

into the EU-MMO framework. All the data and analysis tools will be made FAIR via generation of metadata fully describing data provenance and lineage, functionalities and interfaces of data analysis tools etc. This workpackage will use the infastructure developed and maintained by WP2. The workflow metadata information will be organized within a Knowledge Base which will be used in WP6. The Knowledge Base metadata will be organized in a standartized way to favour interoperability of different types of data and data analysis workflows from WP4-6.

Key Outputs:

- service for integration of data and data analysis tools of telescopes and astronomical messenger detectors into EU-MMO;
- EU-MMO online data analysis services for specific telescopes (radio, optical, X-ray, gamma-ray) accessible through EOSC via EU-MMO frontend.
- Analysis Knowledge Base of telescope and astronomical messenger detector data analysis workflows metadata.

WP4 **Earth observations data (Earth-Data)** will be devoted to activities on porting

- satellite imagery of the Earth in different wavelength bands

and the associated data analysis tools into the EU-MMO framework. All the data will be made FAIR via generation of metadata fully describing their provenance, lineage and scientific context which will be organized within a the Knowledge Base in the standartized form favouring interoperability with other data and analysis tools from WP3, WP5, WP6.

Key Outputs:

- service for integration of data and data analysis tools of Earth observations data into EU-MMO;
- EU-MMO online data analysis services for specific Earth observation datasets accessible through EOSC via EU-MMO frontend.
- Analysis Knowledge Base of Earth observations data analysis workflows metadata.

WP5 **Smart linking of analysis workflows to publications (Smart-Workflows)**. This workpackage will develop tools to link data analysis workflows to scientific publications. It will provide services for

- integration of analysis workflows from WP3, WP4 into different types of human and machine readable publications: from full scale refereed journal articles rapid short automatic nano-publications.
- identifying elements of data analysis workflows from WP3, WP4 in different types of human and machine readable publications: from full scale refereed journal articles rapid short automatic nano-publications.

Information on data analysis methods and workflows from publications will be used to identify scientific contexts of the usage of workflow elements known to EU−MMO. The scientific context information will be organised within dedicated ontology system. Metadata of each instance of publication-linked workflow will be memorized. The scientific context metadata will be added to the Knowledge Base, based on the dedicated scientific context ontology.

Key Outputs:

- EU−MMO service to integrate existing workflows into publications;
- EU−MMO service to extract information on analysis wokflows and their science context from publications.

WP6 **Workflow analytics for implementation of FAIR principles (**FAIR−Workflows). This workpackage will develop services for

- finding data analysis tools relevant to scientific context of research questions,
- composing data analysis workflows out of these tools and
- deploying them on EOSC-connected computing infrastructure developed by the WP2.

It will supplement all astronomy, cosmology and Earth observations data from WP3, WP4 and workflow-publications linking from WP5 with Cloud-based data analysis systems which will allow re-generation of the derived data from the raw data collected by telescopes and detectors. The raw and derived data and elements of the data analysis workflows will be classified based on

- workflow structure,
- usage statistics of workflow elements
- scientific context of the analysis.

The workflow metadata generated at each usage will be added to the Knowledge Base, structured based on dedicated workflow elements, usage and science context ontologies.

The Knowledge Base will be analyzed using machine learning technique to optimize searching and find interconnections and patterns in the Knowledge Base structure. The goal will be to develop in-built intelligence in the EU−MMO platform, which will help users to find and combine diverse types of data and compose, adjust and deploy data analysis workflows within specific scientific contexts of different applications in the domains of astronomy, cosmology and environmental sciences.

Key Outputs:

- intelligent Workflow Inference Engine assuring FAIR principles implementation for workflows deployable via EU−MMO;
- EU−MMO service for finding and composing workflows adjusted to science context of research questions.

WP7 **Operation of EU−MMO services thier integration into worldwide service networks, user support (**Operations**).** The task will be to deploy and maintain operational services powering EU−MMO platform, integrate the platform in the network of interconnected astronomical, cosmological, environmental science data platforms and services worldwide. This will require maintanance of service stability, version control and release cycle management, user support.

Key Outputs:

- Active EU−MMO platform services;
- Helpdesk for user support.

(b) Workpackage descriptions (Table 3.1b)

WP #	WP Title	Lead participant	Short name	Person-months	Start month month	End month month
1	Management				1	36
2	Platform				1	36
3	Astro-Data				12	36
4	Earth-Data		UNIGE/GRID		12	36
5	Smart-Workflows				12	36
6	FAIR-Workflows				12	36
7	Operations	University of Geneva	UNIGE		24	36
			Total PM:	??		

Table 3.1a: List of workpackages.

WP number:	1	**Lead beneficiary:**									
WP title:	Management										
Participant number:	1	2	3	4	5	6	7	8	9	10	11	12
Person-month per participant:												
Start month:	1	**End month:**				1						

Objectives.

- Establish the EU-MMO Management Team (MT) which will be in charge of financial, administrative and reporting elements of the project.
- Oversight of the scientific progress of WPs and their coordination.
- Control all deliverables and milestones, provide their scientific validation and dissemination.
- Coordinate project development with EOSC-hub.
- Provide networking and outreach to industry, SME and wide public.

Description of work.
WP1will assure efficient EU-MMO governance and implement the coordination and management of the project. It will establish the appropriate governance structure for the project. This will include nomination of the MT, planning of the project activities and coordination meetings, The MT will include Project Coordinator, Project Manager, Communication and outreach manager, the WP leaders and one representative of each project partner organisation.
The MT will be responsible for the project interfaces with

1. EOSC-hub with which EU-MMO will be integrated at medium integration level.
2. ESCAPE project which aims at connection of ESFRI infrastructures to EOSC. Some of those infrastructure are relevant to the multi-messenger astronomy domain and will be used by the EU-MMO platform.

Project governance tasks will include

- organise, prepare and minute meetings of the MT
- oversee the implementation of the decisions of the MT.
- Organize general meetings of the project.

Project coordination and management tasks will include

- supervision of the application of the EOSC general policy for service settings ??
- follow-up all EOSC regulations,
- promotion of EU-MMO to the global, European and national communities by presenting it at meetings and international conferences.
- distribution of the EC funding to the partners based on the Horizon 2020 rules and the Consortium Agreement.
- implementation of appropriate reporting structures within and between the various WPs. ??
- monitoring of the milestones and deliverables, chasing any outstanding actions/deadlines. ??
- production of project reports.
- establishment and maintainance of the central risk register, incl. entries for contingency and risk mitigation,

Project outreach activities will include: complete
Tasks:

T1.1 Governance, coordination and project management.

T1.2 Dissemination of project results and networking.

T3.1 Public outreach.

Deliverables.

WP number:	2	**Lead beneficiary:**										
WP title:	Platform											
Participant number:	1	2	3	4	5	6	7	8	9	10	11	12
Person-month per participant:												
Start month:	1	**End month:**				1						

Objectives.

Description of work.
Tasks:

Deliverables.

WP number:	3	**Lead beneficiary:**										
WP title:	Astro-Data											
Participant number:	1	2	3	4	5	6	7	8	9	10	11	12
Person-month per participant:												
Start month:	1	**End month:**				1						

Objectives.

Port data and and data analysis tools of multiple telescopes and astronomical messenger detectors and simulations

- radio, optical, X-ray, gamma-ray telescopes
- high-energy neutrino detectors
- gravitational wave detectors
- sky surveys analysis tools
- image, spectra and time series analysis tools
- numerical simulations for data and detector / telescope response modelling
- cosmological numerical simulations

into the EU−MMO framework. Generate and store metadata of the data and analysis workflows.

Description of work.

EU−MMO platform will provide access to on-the-fly analysis of telescopes and detectors operating across energy ranges and messenger types. Their integration into EU−MMO framework will be done in stages.
The initial focus will be on data and data analysis workflows relevant to the analysis of transient multi-messenger sources. We will integrate analysis workflows of X-ray and gamma-ray telescopes (INTEGRAL, SWIFT, Fermi), gravitational wave detectors (LIGO, VIRGO) and neutrino telescopes (ANTARES) to the common framework of EU−MMO. This collection of workflows will be used as a test case for a system which will be able to provide efficient access to data analysis deployable online and accessible via web interfaces.
In a parallel development, we will integrate in the system different types of data and analysis tools relevant to the domain of cosmology: large photometric and spectroscopic sky surveys, cosmological simulations.
The initial collection of diverse data analysis systems will allow to work out inteoperability standards which are generic enough to be further extended with additional instrument-specific data analysis systems. For the second stage the main focus will be not on the particular telescopes or detectors, but rather on tools allowing to port new data analysis systems into EU−MMO online data analysis framework. These tools will include annotation of the workflows in preparation for implementation of the FAIR principle (WP6). We will explicitly invite external experts to add further instrument / telescope / detector analysis workflows using these tools and also allow ”crowdsourcing” approach by enabling integration of relevant workflows from e.g. GitHub into EU−MMO environment.

Tasks:

T3.1 Integrate gamma-ray telescope data analysis into EU−MMO.

T3.2 Integrate X-ray telescope data analysis into EU−MMO.

T3.3 Integrate neutrino data analysis into EU−MMO.

T3.4 Integrate gravitational wave data analysis into EU−MMO.

T3.5 Integrate optical survey data analysis into EU−MMO.

T3.6 Integrate universal imaging, spectral and timing analysis tools into EU−MMO.

T3.7 Integrate analysis of cosmological simulations data into EU−MMO.

T3.8 Develop data and data analysis import tools for integration of telescope / detector data analysis systems into EU−MMO.

T3.9 Use the import tools to port further astronomical data analysis workflows into EU−MMO.

Deliverables.

D3.1 Service for integration of data and data analysis tools of telescopes and astronomical messenger detectors into EU−MMO;

D3.2 EU−MMO online data analysis services for specific telescopes (radio, optical, X-ray, gamma-ray) and simulations accessible through EOSC via EU−MMO frontend.

D3.3 Corpus of metadata of astronomical data and analysis workflows for telescopes, astronomical messenger detectors and simulations.

WP number:	4	**Lead beneficiary:**										
WP title:	Earth−Data											
Participant number:	1	2	3	4	5	6	7	8	9	10	11	12
Person-month per participant:												
Start month:	1	**End month:**				1						

Objectives.
Port data and data analysis workflows for

- satellite imagery of the Earth in different wavelength bands

into the EU−MMO framework.

Description of work.

The work will leverage on previous developements for the LiMES plaftorm. During the first stage of the project, the available sattelite imagery data and the analysis workflows of the LiMES platform will be integrated into the EU−MMO.
The user interface will allow users to define a study area the service will obtain the relevant satellite images, do the atmospheric correction, transform the signal into reflectance, and generate the requested products (e.g., pan-sharpened images, NDVI, and various other indexes). The output is a time series of images and statistics, which can be used to assess trends in status and level of protection of a selected site (such as given in Fig. 4).
For the second stage we will develop and use tools allowing to port new data analysis systems into EU−MMO online data analysis framework. These tools will include annotation of the workflows in preparation for implementation of the FAIR principle (WP7). We will explicitly invite external experts to add further Earth Observations data and analysis workflows using these tools and also allow crowdsourcing approach by enabling integration of relevant workflowsfrom e.g. GitHub into EU−MMO environment.
All the data will be made FAIR via generation of metadata fully describing their provenance, lineage and scientific context.
The Online Data Analysis services will enable people without high expertise in remote sensing to produce an assessment of the status of the 2,240 Ramsar Sites the 229 World Heritage Sites and sites of interest for UN-REDD. The results will be integrated into the existing websites of these organizations, such as the UNEP World Environment Situation Room, Ramsar Sites Information Service, IUCN World heritage websites. The results will also be shared through webservices so it can be widely accessible, e.g. via the Group on Earth Observations (GEO).
Tasks:

T4.1 Integrate LiMES platform workflows and data in EU−MMO.

T4.2 Develop data anad data analysis import tools for integration of Earth Observations data and data analysis systems into EU−MMO.

T4.3 Use the import tools to port further Earth Observations data and workflows into EU−MMO.

Deliverables.

D4.1 service for integration of data and data analysis tools of Earth observations data into EU−MMO;

D4.2 EU−MMO online data analysis services for specific Earth observation datasets accessible through EOSC via EU−MMO frontend.

D4.3 Analysis Knowledge Base of Earth observations data analysis workflows metadata.

WP number:	5	**Lead beneficiary:**									
WP title:	Smart−Workflows										
Participant number:	1	2	3	4	5	6	7	8	9	10	11	12
Person-month per participant:												
Start month:	1	**End month:**				1						

Objectives. Develop tools to link data analysis workflows to scientific publications.

Description of work.
The largest database of astronomical analysis descriptions is contained in Data selection and data analysis sections of scientific publications to add scientific context metadata to workflow descriptions. Those sections list relevant parameters of the workflows and provide descriptions detailed enough to allow verification of the results via re-running of the analysis chain.
At the first stage, we will use ScienceWISE keyphrase extraction service to mine information on details of data analysis and identify those details with the elements of the data analysis workflows ported into EU−MMO targeting scientific papers posted on arXiv.org preprint server. is there a commonly used repository for Earth observation papers?
At the second stage we will map descriptions of data analysis details extracted from publications to the workflow elements known to EU−MMO (from WP3-WP4). This will be done in two ways.
First, users of EU−MMO online analysis systems will be encouraged to link deployable workflows which lead to published results directly to the publications reporting the results, to foster reproducibility of the results. A dedicated service will be developed for this purpouse. In those cases extraction of analysis parameters from publications will be largely facilitated and precise unambiguous mapping of publication onto EU−MMO Knowledge Base representation of the workflows will be available.
Next, the set of explicitly linked publications and workflows will be used to train the Online Publication Service (OPA) which will be able to extract information on data analsysis details and parameters from papers which do not have workflows explicitly linked to them, but reporting data analysis of instruments known to EU−MMO (from WP3-WP4). The iWIE of WP6 will then be used to find one or several workflows most close to that descibed in the publication.
This will allow to automatically generate partial metadata on the provenance, lineage and scientific context of published results, to extend the Knowledge Base of EU−MMO and to scrutinize published scientific results. The main challenge will be to deal with the ambiguity of parameters of the workflows not explicitly described in the data analysis sections of research papers. In these cases the mapping between the paper and the workflows would also be ambiguous: single paper might correspond to several guesses data analysis workflows.
Tasks:

T5.1 integrate analysis workflows from WP 3, WP4 into different types of human and machine readable publications: from full scale refereed journal articles rapid short automatic nano-publications.

T5.2 Develop ScienceWISE data analysis ontology using the corpus of arXiv.org publications.

T5.3 Derive ontology of astronomical scientific concepts from arXiv.org publications.

T5.4 Setup Online Publication Analysis (OPA) engine which will extract information on analysis methods and parameters from publications. Train it using the corpus of explicitly linked workflows-publcations. .

T5.5 Use OPA to identify elements of data analysis workflows from WP 3-5 in different types of human and machine readable publications: from full scale refereed journal articles rapid short automatic nano-publications.

T5.6 Generate science context Knowledge Base using metadata of publication linked workflows extracted from the corpus of arXiv.org papers.

Milestones:

M5.1 Explicit integrateion of EU−MMO data analysis workflows into publications is enabled.

M5.2 OPA engine is able to infer workflow structure and parameters of explicitly linked publications – workflows.

M5.3 Extended ScienceWISE ontology of scientific concepts and analysis methods is created and embedded in OPA.

M5.4 OPA is trained and ready for analysis of arXiv.org publication corpus.

M5.5 EU−MMO Knowledge Base is extended with information on data analysis workflows extracted from arXiv.org publication corpus.

Deliverables.

D5.1 EU−MMO OPA service to integrate existing workflows into publications;

D5.2 EU−MMO OPA service to extract information on analysis wokflows and their science context from publications.

D5.3 Astronomical, cosmological and Earth Observations analysis elements and scientific context ontologies in Science WISE.

D5.4 EU−MMO Knowledge Base populated with information on data analysis details and methods extracted from publications.

WP number:	7	**Lead beneficiary:**									
WP title:	FAIR−Workflows										
Participant number:	1	2	3	4	5	6	7	8	9	10	11	12
Person-month per participant:												
Start month:	1	**End month:**				1						

Objectives. Develop services for

- finding data analysis tools relevant to scientific context of research questions,
- composing data analysis workflows out of these tools and
- deploying them on EOSC through EU−MMO platform.

Supplement all astronomy, cosmology and Earth observations data from WP3-WP5 and workflow-publications linking from WP6 with Cloud-based data analysis system which will allow re-generation of the derived data from the raw data collected by telescopes and detectors.
Classify raw and derived data and elements of the data analysis workflows based on

- workflow structure,
- usage statistics of workflow elements
- scientific context of the analysis.

to form the EU−MMO Knowledge Base.

Apply machine learning to the Knowledge Base power an intelligent Workflow Inference Engine (iWIE) which will be able to find, compose and deploy analysis chains based on scoring of workflow metadata within the Knowledge Base.
Develop and expose to users the in-built intelligence in the EU−MMO platform, which will help users to find and combine diverse types of data and compose, adjust and deploy data analysis workflows within specific scientific contexts of different applications in the domains of astronomy, cosmology and environmental sciences.

Description of work.
The Knowledge Base containing metadata on structure, usage and science contexts of deployment of data analysis workflows from WP3-6 will provide a basis for the in-built intelligence of the EU−MMO services.
We will map the data analysis metadata into a Knowledge Graph representation in which the elements will be linked and scored. The graph nodes and links will be annotated with their domain-specific meaning (e.g. input corresponding to time interval / energy bounds, nodes performing imaging, spectral extraction etc). The metadata representation will have dimensions which characterize analysis parameters (for example, the exposure time) along with dimensions which characterize physical properties of analyzed objects (for example, astronomical source flux listed in a catalog).
The links between analysis elements, their parameters, their science context will be given weights based on the usage statistics collected either from the Online Data Analysis services of EU−MMO or from the analysis of publications. Each real data analysis workflow instance will define a certain path through the Knowledge Graph. The set of commonly used paths will form patterns. Those patterns contain information needed for the expertise and intuition on the types of analysis suitable to particular science contexts. We will apply machine learning to score the workflow paths in the Knowledge Graph depending on the science contexts of specific astronomical, cosmological and Earth observations analyses.
We will add ”intuitive” element to the process of inference of workflows suitable to the science context by introducing a Knowledge Base representation based on the Word Embedding technique introduced by [**?**] and implemented in word2vec tool kit. In this representation the Knowledge Base content will be mapped onto vectors in a multi-dimensional Euclidean space, which is derived from the ”context” or ”environment” in which the Knowledge Base elements appear. Such a representation will allow to define distance between Knowledge Base elements. Syntactically/semantically close concepts will be mapped into close points in the Euclidean space. The dimensionality of the embedding Euclidean space is much smaller than that of the raw sparse representation of KNowledge Base. This might make application of classifying / feature finding techniques like e.g. support vector machines, more straightforward. The embedding in Euclidean space will enable explicit expressions for relations of closeness between concepts so that workflows similar to those mentioned in certain papers could be found via basic addition and subtraction of vectors.
The outcome of learning process will be scrutinized using available visualisation methods like e.g. Stochastic Neighbor Embedding, verifying the analysis workflow patterns corresponding to specific scientific contexts. After this assessment and interpretation of the results, we will be ready to use the Knowledge Base within an intelligent Workflow Inference Engine (iWIE) engine recommending data analysis workflows tuned to scientific contexts / research questions.
The iWIE will be integrated into EU−MMO Online Data Analysis services of different telescopes and sensors from WP3-WP5. The ODA fronends will be develop to assist users in the composition of workflows and choice of analysis parameters adjusted to the specific of the research quesitons, of the studied objects (astronomical source types, specific observation sites). Integration of iWIE in the ODA systems will provide explicit solution of the problem of sharing of the narrow instrument specific analysis expertise.

At a more advanced stage, the ODA frontends will also expose the Knowledge Graph infoboxes with the relevant information extracted from the Knowledge Base. This will expose the user with a variety of analysis workflows and science contexts in which the object of study was previously considered by EU−MMO users and discussed in research papers.
A specific use case for the iWIE will be within a transient astronomical source analysis service. We will setup a service which will extract information from astronomical nano-publications (ATEL, GCN, VO-Event). For each nano-publication discussing transient source result (e.g. a report on gamma-ray burst observation) the iWIE will be used to identify relevant data and data analysis workflows within EU−MMO platform to (a) scrutinize the results reported in the nano-publication and (b) complement them with additional analysis of relevant archival data. The service will alert the community by publishing the outcomes of the data analysis on EU−MMO platform or via nano-publication services.
Tasks:

T7.1 Use machine learning to classify and score information contained in the Knowledge base of analysis workflows and their science contexts.

T7.2 Interpret the results of learning. Compare the efficiency of learning algorithms.

T7.3 Set up the intelligent Workflow Inference Engine for suggestion / composition of analysis workflows adjusted to scientific context of research questions.

T7.4 Integrate iWIE into EU−MMO Online Data Analysis services of telescopes, sensors and detectors.

T7.5 Set up a service which will use WIE to identify analysis workflows relevant to the science context of astronomical nano-publications, deploy relevant workflows, inspect the results of analysis and generate new nano-publications reporting relevant results.

Milestones.

M7.1 Knowledge Graph representation of the Knowledge Base is available.

M7.2 Word Embedding representation of the Knowledge Base is available.

M7.3 Intelligent Workflow Inference Engine is functions.

M7.4 iWIE powers Online Data Analysis services of EU−MMO.

M7.5 iWIE powers transient astronomical analysis service which deploys data analysis in response to alerts from astronomical nano-publications.

Deliverables.

D7.1 intelligent Workflow Inference Engine assuring FAIR principles implementation for workflows deployable via EU−MMO;

D7.2 EU−MMO service for finding and composing workflows adjusted to science context of research questions.

D7.3 EU−MMO service to process nano-publications and generate multi-messenger results relevant to those publications.

WP number:	8	**Lead beneficiary:**	
WP title:	Operations		

Participant number:	1	2	3	4	5	6	7	8	9	10	11	12
Person-month per participant:												
Start month:	1	**End month:**				1						

Objectives.

Description of work. **Tasks:**

Deliverables.

3.2 Management structure, milestones and procedures

Describe the organisational structure and the decision-making (including a list of milestones (table 3.2a))

Explain why the organisational structure and decision-making mechanisms are appropriate to the complexity and scale of the project.

Describe, where relevant, how effective innovation management will be addressed in the management structure and work plan. Innovation management is a process which requires an understanding of both market and technical problems, with a goal of successfully implementing appropriate creative ideas. A new or improved product, service or process is its typical output. It also allows a consortium to respond to an external or internal opportunity.

Describe any critical risks, relating to project implementation, that the stated project's objectives may not be achieved. Detail any risk mitigation measures. Please provide a table with critical risks identified and mitigating actions (table 3.2b)

3.3 Consortium as a whole

The individual members of the consortium are described in a separate section 4. There is no need to repeat that information here

Describe the consortium. How will it match the projects objectives, and bring together the necessary expertise? How do the members complement one another (and cover the value chain, where appropriate),?

In what way does each of them contribute to the project? Show that each has a valid role, and adequate resources in the project to fulfil that role.

If applicable, describe the industrial/commercial involvement in the project to ensure exploitation of the results and explain why this is consistent with and will help to achieve the specific measures which are proposed for exploitation of the results of the project (see section 2.2).

3.4 Resources to be committed

Please make sure the information in this section matches the costs as stated in the budget table in section 3 of the administrative proposal forms, and the number of person months, shown in the detailed work package descriptions.

a table showing number of person months required (table 3.4a)

a table showing other direct costs (table 3.4b) for participants where those costs exceed 15% of the personnel costs (according to the budget table in section 3 of the administrative proposal forms)