dissemination of ASTERICS results (D2.1).

- Create brochures for informed communities and public-facing websites for ASTERICS activities (D2.2).
- Attend high-level events and meetings where ASTERICS results can be show-cased to external stakeholders, including industrial and commercial concerns.

Deliverables (brief description and month of delivery)

Nr	Description	Task	Month
D2.1, 2.2	ASTERICS project website live and brochure publication	2.1	3, 12
D2.3, 2.7	Educational resources for mass participation experiments	2.3	23, 35
D2.4, 2.8	Online mass participation experiments	2.5	23, 35
D2.5	Video resources	2.2	29
D2.6, 2.9	Open-access publications from mass participation experiments	2.5	32, 46

Table 3.1a.3: ASTERICS WP3: OBELICS (OBservatory E-environments LInked by common ChallengeS)

WP number	3					**Start Date**	M1
WP title	OBELICS (OBservatory E-environments LInked by common ChallengeS)						
Participant Nr	2				12	3	1
Short name	CNRS LAPP	CNRS IAP	CNRS APC	CNRS CPPM	CEA	INAF	ASTRON
Person months	152	36	24	24	36	108	96
Participant Nr	4	5	18	17	10	19	
Short name	UCAM	JIVE	UCM	IFAE	FAU	INFN	
Person/months	132	24	24	24	36	36	

Objectives

Enable interoperability and software re-use for the data generation, integration and analysis of the ASTERICS ESFRI and pathfinder facilities. An essential ingredient is the creation of an open innovation environment for establishing open standards and software libraries for multi-wavelength/multi-messenger data. Furthermore, development of common solutions for streaming data processing and extremely large databases is required. Study of advanced analysis algorithms and software frameworks for data processing and quality control is the third focus area. The

specific objectives are:

- Train researchers and data scientists in the ASTERICS ESFRI and pathfinder projects to apply state-of-the-art parallel software programming techniques, to adopt big-data software frameworks, to benefit from new processor architectures and e-science infrastructures. This will create a community of experts that can contribute across facilities and domains.
- Maximise software re-use and co-development of technology for the robust and flexible handling of the huge data streams generated by the ASTERICS ESFRI and pathfinder facilities. This involves the definition of open standards and design patterns, and the development of software libraries in an open innovation environment.
- Adapt and optimise extremely large database systems to fulfil the requirements of the ASTERICS ESFRI projects. This requires the development of use cases, prototypes and benchmarks to demonstrate scalability and deployment on distributed non-homogeneous resources. Cooperation with the ESFRI pathfinders, computing centres, e-infrastructure providers and industry will be organised and managed to fulfil this objective.
- Study and demonstrate data integration across ASTERICS ESFRI and pathfinder projects using data mining tools and statistical analysis techniques on Petascale data sets. This will require adaptable and evolving workflow management systems, to allow deployment on existing and future e-science infrastructures.

All tasks are built upon the state-of-the-art in ICT, in cooperation with major European e-infrastructures and are conceived to minimise fragmentation. Communications and links with other communities and e-science service providers are considered in order to contribute to the effectiveness of the proposed objectives.

Description of work

The partners have selected Dr. Giovanni Lamanna (CNRS-LAPP) as scientific coordinator of WP3. He will dedicate approximately 20% of his working time to the project. The leadership of tasks is shared between CNRS-LAPP, ASTRON, UCAM, INAF and UCM. All the ESFRI facilities (CTA, SKA, KM3NeT), and other major international projects or pathfinders, namely EUCLID, LSST, VIRGO/EGO, LOFAR, e-VLBI, HESS, MAGIC and ANTARES take part in all tasks and share responsibilities for task deliverables. The E-ELT (ESO) will be involved in particular in the development of open standards and common software libraries (task 3.2).

The work in WP3 is organised in four tasks. All tasks have well-defined connections with DADI (WP4), either through the definition of standards (task 3.2), making available of meta-data (task 3.3) or direct interfacing with VO libraries (task 3.4). The interface with WP4 is managed through a special subtask of task 4.3 led by INAF.

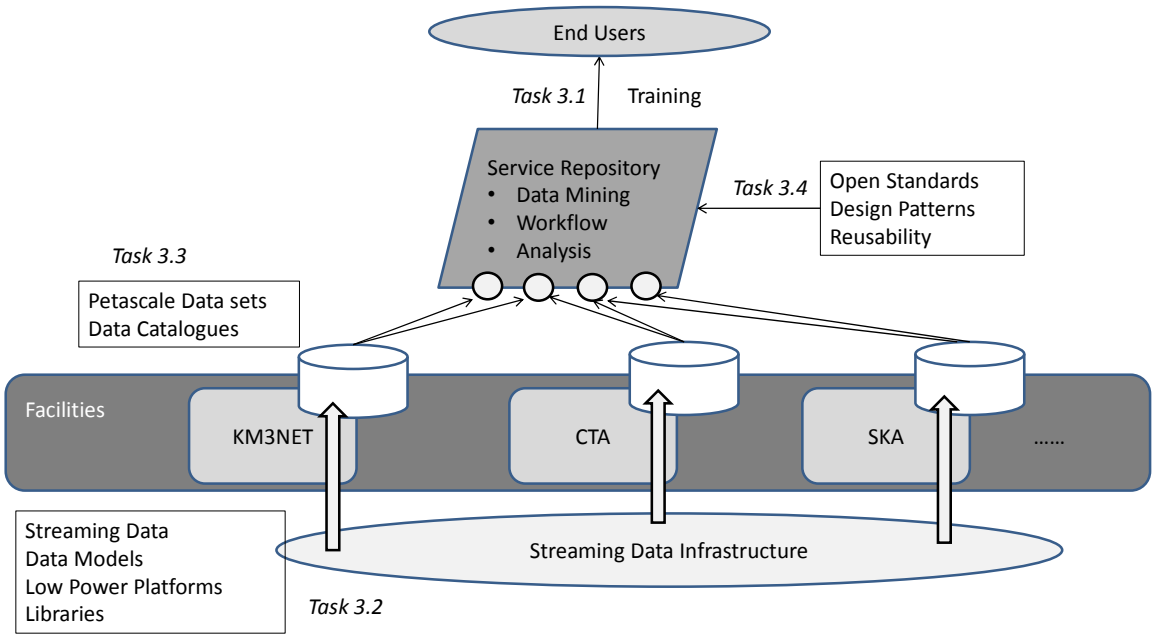

Figure 5: ASTERICS delivers a repository of services and Data Technology solutions for the ESFRI projects.

Task 3.1 MAUD: MAnagement, User engagement and data Dissemination

Partner	**LAPP**
Effort (PM)	44

As WP3 represents a major effort in its own right, this first task is devoted to the detailed day-to-day management of the work package, aiming to guarantee the federated character of all activities among ASTERICS ESFRI projects, and organising the three technical tasks in a coherent and well regulated manner. OBELICS is the cornerstone of the ASTERICS project and it also aims to guarantee the best interface with major e-infrastructures in Europe, as expected from the work programme of the INFRADEV4 call.

The task 3.1 addresses the following topics:

1. Defining and monitoring the planning, compiling progress reports and organising regular meetings.
   - 3 annual WP3 general workshops (M18, M30, M42).
   - 3 main thematic training events: (a) software parallelisation and big-data frameworks; (b) new computing technology; (c) archive and metadata-data-base systems. Where possible we invite communities and disciplines larger than those covered by ASTERICS.
2. Overall data dissemination, communication and exposure of the results to the target communities (in cooperation with WP1 & 2). Foster co-located periodical meetings and promote the ASTERICS-ESFRI users engagement in larger forums: e.g. new HEP-software foundation; e-infrastructures initiatives such as EGI; data-infrastructures initiatives such as EUDAT and RDA.
3. Managing and assuring coherent and cooperative interfaces with major computing/data centers (some already committing in a sustainable way to the support of the ASTERICS-ESFRI projects individually), providing test-bench platforms for the developments foreseen in all tasks (e.g. service agreements), as well as with transversal international e-initiatives, e.g. RDA and

EGI, and with international collaborations of data and computing resource providers, e.g. PRACE, EU-T0 and similar consortia.

4. ASTERICS aims to deliver a broad range of innovative software libraries and services to the facilities. The prioritisation of these software components and the specification of their requirements can only be made in the course of the project – arranging appropriate sub-contractors at this stage is premature. For a selection of the components, the generation is optimally executed in co-development with industry. Therefore, the consortium will organise industrial engagement for (precompetitive) co-development with industry via an internal call for sub-contracting technical works. WP3.1 manages this process in the same way as has been successfully organised for the FP7 CTA Preparatory Phase project. After the completion of this call, the consortium agreement and annex 1 will be adapted to incorporate any new subcontractors. A preliminary innovation plan of investments foresees some possible paths such as:

   a) "Professional software co-developments" with intellectual property of the delivered products transferred entirely to ASTERICS partners.

   b) Co-development of database software frameworks where private companies have already achieved a superior level of competence or are willing to follow-up the ESFRI use cases for cooperative work. An example is the case of applying web-technologies such Hadoop generally used on text files to astronomical data formats such as FITS and/or FITS-wrapped data packets.

   c) Consulting contracts to support technological survey (e.g. on workload management on distributed data centres; engineering of on-line data streaming processing/ground segments).

   d) Exploring a few co-funded PhD scholarships with major private companies (e.g. CUDA for GPU programming) around astronomical data analysis use cases.

   e) Sub-contracting to SMEs some computing benchmark prototypes, combining low-cost processors such as ARM and GPUs, integrating both on-line data streaming efficiency and data processing acceleration.

Task 3.2 D-GEX: Data GEneration and information eXtraction (UCM +INAF)

Partner	**INAF**	**UCM**	ASTRON	UCAM	CPPM	LAPP	IFAE	INFN
Effort (PM)	24	16	24	24	18	6	10	12

In this area of the data flow, there are common challenges to create more robust hardware and software solutions for the handling of ever increasing data streams, and to ensure interoperability between a variety of different data-sources. OBELICS will promote sustained cross-fertilisation via a three-step process: a) share studies and seek synergies, b) foster evaluation and adoption of innovative solutions, c) sharing common prototype frameworks and standards.

Following this approach, D-GEX will be concerned with:

1. Surveying the real-time streaming data architectures applied and envisaged for the ESFRI and pathfinder facilities, to establish best practices and agree on common software frameworks or common software modules, extending e.g. LOFAR, ASKAP, HESS, MAGIC, ANTARES and ALMA frameworks.

2. Developing new and common data models and high performance formats for data streaming, compatible with interoperability standards beyond the existing FITS, ROOT and HDF5 leading to common standards.

3. Developing prototype libraries that allow robust and optimised handling of secondary data streams and meta-data (environmental and engineering data, temporary local archive, device control software and observation scheduling), ensuring long-term & shared maintenance of the proposed products.

4. Benchmarking low-power computer platforms (including Multicore, MIC, Microservers, GPU, FPGA, ARM) and software technologies/methods for data-driven scalable parallel programming. This subtask will also follow a three-step approach, and will last the full ASTERICS project duration, since it will evolve by monitoring the continuous evolution of the technologies and could be also inspired by progress made in other scientific domains. The expected measurable value is the up-take of these new computing and information technologies by the ESFRI facilities and there platforms.

Task 3.3 D-INT: Data systems INTegration

Partner	**LAPP**	**ASTRON**	INAF	UCAM	UCM	IFAE	FAU	INFN
Effort (PM)	72	48	48	48	8	14	36	6

The major common challenge addressed in this task is scaling-up existing databases and storage architectures beyond the Peta-scale level, while allowing for more complex queries addressing both primary sensor data and secondary data-streams. This task will involve:

1. Collecting use cases from existing facilities (in particular LOFAR) that place extreme requirements on the databases and the e-infrastructures on which they are deployed, and develop these into benchmarks for future facilities and technologies (in particular the Extremely Large Data Base “XLDB” initiative) aiming for full interoperability. This will allow the ESFRI and related pathfinder facilities to optimally engage with providers of e-infrastructures.
2. Developing of prototype benchmarks for testing within a larger multidisciplinary context (in particular the XLDB initiative) aiming for interoperability. The current investigation and prototypes of some partners are around the Qserv solutions or new generation DBs such as Cassandra or MongoDB. These will be used as examples and compared to the use cases of other projects.
3. Developing a portfolio of open services for data integration, based on existing Data Management System services like FLUME, RUCIO and Hadoop but extending these with optimised software modules to support e.g. VO-integration (in collaboration with WP4), and data interoperability between primary and secondary data-streams. Some services are already well known and applied, therefore a minimal repository can be made available in the first year. The repository will continuously be filled with new results of common work.
4. Extending software frameworks for data catalogues and query solutions to maximise data integration. This requires the reduction of latency for high data rates and the integration of multi-parameter Instrument Response Functions. Benchmarking and verification will be done through real pathfinder data, as well as Monte Carlo simulations. The complete software will be made available through the service repository.

Task 3.4 D-ANA: Data ANAlysis/interpretation

Partner	**INAF**	**UCAM**	LAPP	IAP	APC	CPPM	CEA	ASTRON	JIVE	INFN
Effort (PM)	36	60	30	36	24	6	36	24	24	18

In this area of the data flow, there is a common challenge to assess the quality of Petascale data sets and execute automatic analysis to reduce their size. This task is therefore concerned with:

1. Developing a collection of statistically robust and domain independent open source software libraries for data analysis and data mining on Peta-scale datasets. This will enable a sustained

community-based effort towards excellent exploitation of all data generated by the ESFRI and pathfinder facilities. The initial set of libraries developed within this task are in particular:

- Statistically robust approaches (Bayesian and likelihood analyses) to advance cross-matching between catalogues and transients detected via different instruments
- Domain independent image analysis for simultaneous feature classification and extraction in multi-dimensional/multi-resolution data where the data are from multiple instruments.
- Effective likelihood reconstruction methods and new graphical processing approaches (mainly for event-based and signal-based projects but not exclusively) optimised for new computing technologies and maximum efficiency.

2. Establishing a common set of workflow architectures for the orchestration of compute intensive analysis of Peta-scale datasets on distributed computing infrastructures. This involves providing use cases and technical requirements, designing and testing of workflow engines on distributed compute-intensive systems, and improving existing authorisation, authentication and accounting protocols (e.g. eduGAIN services). In this respect, the cooperation with projects supported by the European Commission through the call EINFRA-7-2014, "Provision of core services across e-infrastructures" aimed to produce a global authorisation and authentication infrastructure is foreseen. Dedicated working meetings in the first stage of the project (M6) will be organized between the OBELICS partners and e-infrastructures providers, identifying the services to be adopted, explored or further developed through a shared approach.

All activities in this task will liaise and be complementary with WP4 (DADI), interfacing the respective activities will allow that the data, managed through the mechanisms identified and built by WP3, are archived, accessed, discovered and interoperated through the mechanisms defined by WP4.

Deliverables

Nr	Description	Task	Month
D3.1	Detailed WP3 Project plan	3.1	4
D3.2, 3.6, 3.10	Annual user engagement forum, workshops and training events	3.1	12, 24, 36
D3.3	Analysis Report on Standards and Libraries	3.2	12
D3.4, 3.17	Release of Software Libraries	3.4	12, 48
D3.5	Analysis Report on Resource Requirements	3.3	18
D3.7, 3.15	Processing Platform Technology Benchmark Report	3.2	24, 48
D3.8, 3.16	Database Technology Benchmark Report	3.3	24, 48
D3.9	Statistical Solvers Technology Benchmark Report	3.4	24
D3.11	Analysis Report on Frameworks and Architectures	3.2	36
D3.12	Repository of Services	3.3	36
D3.13	Repository of WMS Services	3.4	36
D3.14	Final Integral WP3 Report	3.1	48