Overview Task 3.2 D-GEX
OBELICS f2f Madrid
14 September 2016

J.L. Contreras

OBELICS and DGEX

| 3.2 DGEX | 3.3 DINT | 3.3 DANA |

1. **MAUD :** Management user engagementand Data disemination
2. **D-GEX:** Data Generation and information Extraction
3. **D-INT:** Data Systems Integration
4. **D-ANA:** Data ANAlysis / interpretation

Dedication to Tasks

MAUD	LAPP
PMs	**44**

D-GEX	ASTRON	CCPM	IFAE	INAF	INFN	LAPP	UCAM	UCM
140	24	18	10	24	12	6	24	16

D-INT	ASTRON	FAU	IFAE	INAF	INFN	LAPP	UCAM	UCM
280	48	36	14	48	6	72	48	8

D-ANA	APC	ASTRON	CEA	CCPM	IAP	INFN	JIVE	LAPP	UCAM
258	24	24	36	6	36	18	24	30	60

UNIVERSIDAD
COMPLUTENSE
MADRID

D-GEX: Work load (PM) and contact persons

INSTIT	PM	Persons	Collab
ASTRON	24	T.J. Dijkema	LOFAR, SKA
CCPM	18		ANTARES/KM3?
IFAE	10	T. Hassan, J. Rico	MAGIC/CTA
INAF	24	S.Lombardi, M . Molinaro	MAGIC/CTA
INFN	12	C.Bozza, D. Cesini	KM3
LAPP	6	J. Jaquemier	HESS/CTA
UCAM	24	B. Nikolic	SKA, LOFAR
UCM	16	J.L. Contreras, J. Rosado	MAGIC/CTA

D-GEX Goals I

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.

D-GEX Goals II

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.

D-GEX Goals and III

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.

Deliverables

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

Organization

ASTERICS Wiki: D-GEX

intra:wp3:task3.2

Task 3.2 D-GEX: Data GEneration and information eXtraction

Partner	INAF	UCM	ASTRON	UCAM	CPPM	LAPP	IFAE	INFN
Effort (PM)	24	16	24	24	18	6	10	12
Contact point	Marco Molinaro, Saverio Lombardi	José Luis Contreras	Tammo Jan Dijkema	Bojan Nikolic		Jean Jacquemier	Tarek Hassan	Cristiano Bozza

Edit

Activities

  - UCM

Each group can report its activities

Edit

From the proposal

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

The future

- On our side we see clear possibilities of cooperation *in the aspects which are common* (besides physics and VO):

  - Event based experiments: Common formats for event lists → see Jaime´s and Tarek´s presentation

  - Big Data experiments: Common file format (HDF5?) → FJ Franco @D-INT

  - Experiments in development: Observatory Use cases

Next steps

- We also see next steps in

  - Low-cost parallel computing (GPUs, etc).

  - Frameworks: data streaming

My reflections

- There is not a clear separation among tasks, same people and similar goals (eg UCM): makes difficult to present a task view, much easier at OBELICS level.

- Our main goal is to get ahead with our projects and sometimes is difficult to see the ASTERICS added value. ¿ How do we make it compatible ? Different to DADI.

- Timing is a problem: contracts for persons working on the same subject have to overlap !

- ¿ How do we publish our deliverables ?