OBELICS plans and actions

Giovanni Lamanna
LAPP-IN2P3-CNRS

1st OBELICS F2F
Rome, 26-27 January 2016

Commitments and deliverables

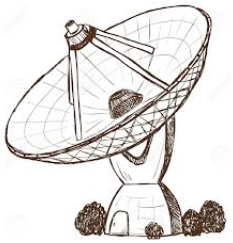
1) **D3.3** Analysis Report on Standards and Libraries
   Task 3.2
   M12: 1/5/2016.
   Leader Institutes: UCM, INAF…

2) **D3.4**, Release of Software Libraries
   Task 3.4
   M12: 1/5/2016 (next on M48).
   Leader Institutes: UCAM, …

3) **D3.2**, 3.6, 3.10 Annual user engagement forum, workshops and training events
   Task 3.1
   M12, M24, M36.
   Leader Institutes: LAPP, …

4) **D3.5** Analysis Report on Resource Requirements
   Task 3.3
   M18: 1/10/2016
   Leader Institutes: ASTRON, LAPP, …

OBELICS commons



1. Data generation

D-GEX: Data GEneration and information eXtraction:

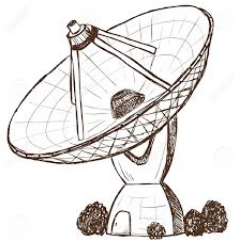
- ✓Surveying real-time or close-to-detector data streaming frameworks.
  (e.g. Hadoop, ACS and others; aiming at file and metadata management, fast algorithms integration, automatic remote acquisition, identification and ingestion..)
- ✓Standards on data model and data format.
  (e.g. Protocol buffer saving bandwidth; HDF5 simplifying big-data structure; evolution of scientific data FITS format; streaming protocols adopted in space projects...)
- ✓Prototype libraries handling secondary data streams.
  (environmental and engineering data, temporary local archive, device control software and observation scheduling)
- ✓Benchmarking low-power computer platforms. *(GPU + ARM, FPGA, Microservers, …)*

Commitments and deliverables

1) **D3.3** Analysis Report on Standards and Libraries
   Task 3.2
   M12: 1/5/2016.
   Leader Institutes: UCM, INAF, ASTRON, UCAM,…

   Proposal to get D3.3
   Report Editing coordinator: Marcos Lopez
   Contributors: event-based (CTA, KM3Net), Image-based (EUCLID, LSST), signal-based (LOFAR & SKA, Virgo?), …
   Milestones:
   m1- Authors assignments: 5/2/2015
   m2- Document outline and editing assignments: 15/2/2016
   m3 - First draft: 15/3/2016
   m4 - Second draft: 31/3/2016
   m5 – Final draft: 15/4/2016

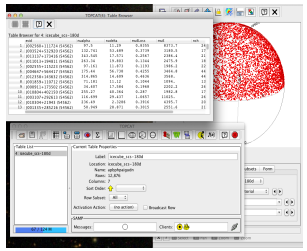
OBELICS commons



1. Data generation

2. Data systems integration

3. Data analysis



D-ANA: Data ANAlysis/interpretation

- ✓ Open source software for data analysis.
  - *a) Bayesian and likelihood analyses approaches for cross-matching between catalogues and transients detected via different instruments.*
  - *b) Simultaneous feature classification and extraction in multi-dimensional/multi-resolution data where the data are from multiple instruments.*
  - *c) Effective likelihood reconstruction methods and new graphical processing approaches.*

- ✓ Workflow architectures for Peta-scale datasets on distributed computing infrastructures.
  - *a) orchestration of compute intensive analysis of petascale datasets on distributed computing infrastructures (workflow engines on distributed systems, AAA protocols.)*

Commitments and deliverables

2) **D3.4**, Release of Software Libraries
Task 3.4
M12: 1/5/2016 (next on M48).
Leader Institutes: UCAM, …

Proposal to get D3.4
Report Editing coordinator: Bojan, Fabio
Contributors: Python lib. + Jupyter (CEA, JIVE, ASTRON); Gammas (LAPP); AAA (INAF); GPUs libraries ? (INAF, INFN, ASTRON); Bayes. Multi-lambda (UCAM), …
Milestones:
m1- Library contents assignment: (today or before) 5/2/2015
m2- Written detailed description and forum (on Redmine) and guidelines (for WEB): 18/2/2016
m3 – OBELICS WEB-PAGE preparation: 10/2/2016
m4 – First Web-repository prototype: 15/3/2016
m5 – Second Web-repository prototype: 15/4/2016
m6 – Release Web-repository : 1/5/2016

OBELICS commons

1. Data generation

2. Data systems integration

D-INT: Data systems INTegration

- ✓ Scaling-up existing databases and storage architectures beyond the Peta-scale level for complex queries.

  (Cooperative activities on "identification and archiving interesting data products")

  i) Developing prototype benchmarks of large size DB: Cassandra, MongoDB, Qserv.

  ii) Testing and adopting data-management- system services for data-sets integration: FLUME, RUCIO, …

  iii) Multi-parameter instrument response function integration.

Commitments and deliverables

4) **D3.5** Analysis Report on Resource Requirements
Task 3.3
M18: 1/10/2016
Leader Institutes: LAPP, …

Proposal to get D3.5
Report Editing coordinator: Tammo, Giovanni
Contributors: One Institute per project (ASTRON-SKA; ..)
Providing updated computing model and data management specifications.
Milestones:
m1- First collection of contributions: 15/2/2016
m2- First requirements extraction: 1/3/2016
m3 – Draft report on requirements (validated): 20/3/2016.
m4 – Analysis and definition of potential solution and e-infras: between15/4/2016 and 1/7/2016.
m5 – Draft report: 1/8/2016
m6 – Release report: 20/9/2016

Commitments and deliverables

3) **D3.2**, 3.6, 3.10 Annual user engagement forum, workshops and training events

Task 3.1
M12, M24, M36.
Leader Institutes: LAPP, …

To Be Discussed …

Project plan

First plan released at M4.
An update to come after first M12 deliverables: 1/5/15.

- Start-up actions

  (Time: from October 2015 to February 2016)

  i. Work plan and Work Breakdown Structure assessment.
  ii. First recruitments at work.
  iii. Task and sub-task assignments.
  iv. First D-GEX meeting and work plan approval.
  v. First D-INT meeting and work plan approval.
  vi. First D-ANA meeting and work plan approval.
  vii. Data dissemination plan (MAUD).
  viii. OBELICS general meeting.

Project plan

- First year events
  (Time: from January 2016 to May 2016)
  i. Pursue of co-located event or active participation: ~~EU-T0 & APPEC workshop on 24-25 January 2016;~~
  ii. Pursue active participation: RDA meeting.
  iii. Plan summer conferences participation.
  iv. Organization of D-INT workshop.
  v. Establishing e-infrastructure engagements for cooperation with OBELICS (MAUD).
  vi. Organize user-engagement meeting towards a *working open project: how to collaborate on coding projects of mutual interest* (MAUD).

- First year ASTERICS deliverables
  (Time: expected May 2016 + 2 ~~months maximum delay~~)
  i. D3.2 (MAUD)
  ii. D3.3 (D-GEX)
  iii. D3.4 (D-ANA)
  iv. Second year data dissemination plan (MAUD).

Project plan

- Second year activities.
  (Time: from May 2016 to May 2017)
  i. OBELICS general meeting.
  ii. After first 18 months WP3 internal review and project plan upgrading.
  iii. Training/Spring-session events dedicated to *Scientific Software Developers*, potential and optional subjects are *Parallel programming and new computing architectures* opening to other scientific domains.
  iv. To consider: OBELICS participation to EGI, PRACE, RDA, e-IRG and EUDAT forums.
  v. Workshop on *Data and computing infrastructures for open science,* opened to other scientific domains and commercial Cloud providers.
  vi. D-GEX meeting: status report and deliverables monitoring.
  vii. D-INT meeting: status report and deliverables monitoring.
  v. D-ANA meeting: status report and deliverables monitoring.

Events and dissemination

- One option: towards completion of D3.3 a workshop can be organized where:
  - based on computing model and archive requirements we convene scientists and e-infra. providers to learn more about:
  i) services for workflow management;
  ii) cloud computing and data-cloud (for open science?);
  iii) computing architectures (combining HTC, HPC and LPC);
  iv) defining training or briefing sessions for wm services.

3) **D3.2**, 3.6, 3.10 Annual user engagement forum, workshops and training events

Task 3.1
M12, M24, M36.
Leader Institutes: LAPP, and OBELICS tasks leaders

Action Points

- Actions list from each Task leader for the task plans and roadmap to 12 -18M deliverables.
- Training and workshops events organization
- Management: project plan, hiring plan, dissemination plan together with WP1.
- Industrial contracts ideas?