



H2020, APPEC & Computing

Giovanni Lamanna

*LAPP - Laboratoire d'Annecy-le-Vieux de Physique des Particules,
Université de Savoie, CNRS/IN2P3, Annecy-le-Vieux, France*

Zeuthen 5 November 2013





The discussions about “Astroparticle and Computing” took place in the last three years in three dedicated workshops organised in the context of ASPERA/APPEC :

Lyon 2010:

<https://indico.in2p3.fr/conferenceDisplay.py?confId=3845>

Barcelona 2011:

<https://indico.cern.ch/conferenceDisplay.py?ovw=True&confId=134280>

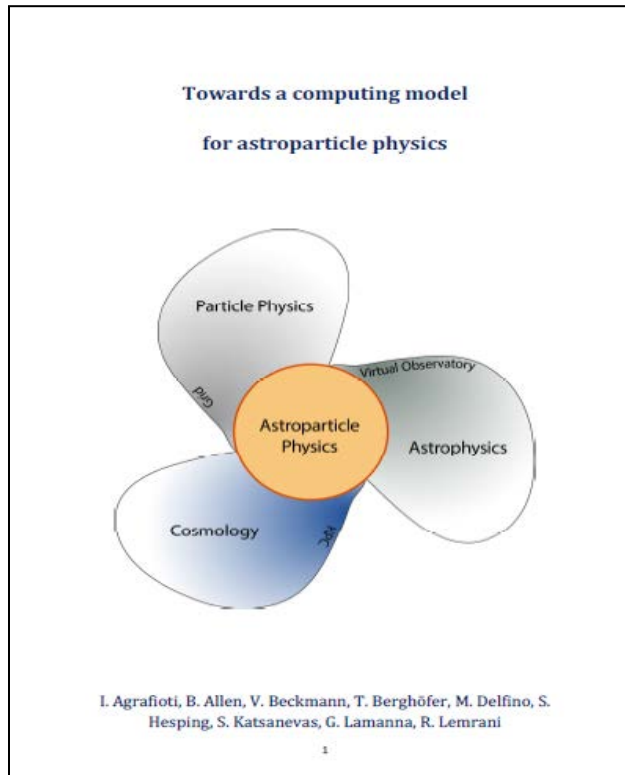
Hannover 2012:

<http://indico.cern.ch/conferenceDisplay.py?confId=159120>

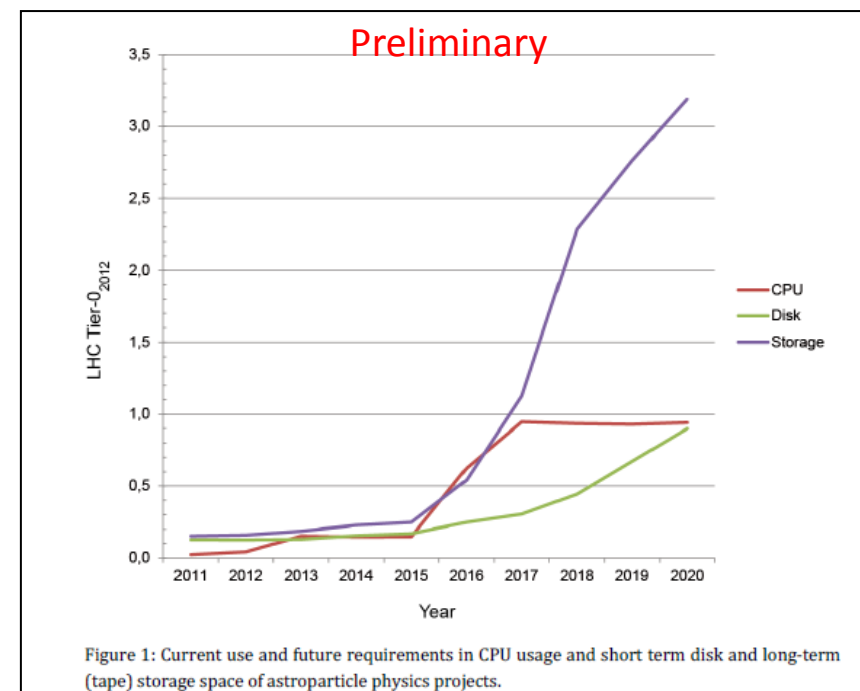
They demonstrate a high level of complexity of this topic and the need for further coordination before the White Paper for Computing and Astroparticle Physics can be released...



The discussions were kept going and contacts within the “Aspera/Appec core group” devoted to the computing issues as well. A summary document + dedicated H2020 preliminary meetings.



data with event-like structure	AUGER (total)
	HESS
	MAGIC
	CTA
	ANTARES
	KM3NeT
	IceCube
	FERMI
	AMS
	subtotal
data with signal-like structure	VIRGO
	advanced VIRGO
	LIGO
	advanced LIGO
	subtotal
data with image-like structure	PAU
	SNLS
	EUCLID
	LSST
	subtotal



ISSUES UNDER DEVELOPMENT AND INVESTIGATION IN AP PROJETS:

- Big Data projects: e.g. towards new DB technologies, distributed and federated archives, data mining, ...
- New computing models: e.g. distributed cloud infrastructures, new private-public paradigms, sustainable e-infrastructures, world-wide shared data pipelines, ...
- New software and middleware: e.g. parallel programming and MC simulations, new computing and storage technologies, ...
- Data management: e.g. open data access, scientific frameworks and user services, certification protocols, ..

CONSEQUENCES :

Potential sharing for common and/or close joined deliverables.

DIRECTIONS :

- Creation or participation to a Centre of excellence
- Participation to “e-infrastructure” initiatives
- Promote World-wide cooperation on computing
- Training network



EXCELLENT SCIENCE: B. Research Infrastructures

- Call 3 - e-Infrastructures
 - **EINFRA 1-2014 – Managing, preserving and computing with big research data**
 - EINFRA 2-2014 – e-Infrastructure for Open Access
 - EINFRA 3-2014 – Towards global data e-infrastructures – Research Data Alliance
 - **EINFRA 4-2014 – Pan-European HPC infrastructure and services**
 - **EINFRA 5-2015 – Centres of Excellence for computing applications (??)**
 - **EINFRA 6-2014 – Network of HPC Competence Centres for SMEs**
 - EINFRA 7-2014 – Provision of core services across e-infrastructures
 - EINFRA 8-2015 - Research and Education Networking – GÉANT
 - **EINFRA 9-2015 – e-Infrastructures for virtual research environments (VRE)**
- *Call 4- Support to Innovation, Human resources, Policy and International cooperation*
 - *INFRASUPP 4-2015 – New professions and skills for e-infrastructures*



Specific challenge: Development and deployment of integrated, secure, permanent, on-demand service-driven and sustainable e-infrastructures incorporating advanced computing resources and software are essential in order to increase the capacity to manage, store and analyse extremely large, heterogeneous and complex datasets, including text mining of large corpora. These e-infrastructures need to provide services cutting across a wide-range of scientific communities and addressing a diversity of computational requirements, legal constraints and requirements, system and service architectures, formats, types, vocabularies and legacy practices of scientific communities that generate, analyse and use the data.

Scope: Proposals should address at least one of the first five (5) activities, or activities 6, 7 or 8

- 1. Establishing a federated pan-European data e-infrastructure to provide cost-effective and interoperable solutions for data management and long-term preservation.** The needs for data access, replication, annotation, search, compute, analysis and reuse of information across disciplines should be accommodated in different research and education contexts. All these functions should expose standard interfaces for interoperation with other data sources to aggregate them or to be aggregated. Sustainability is of paramount importance, therefore robust business models should be proposed to encourage investment from all stakeholders. Foreseen challenges are technical, legal and organisational, including engaging e-infrastructure operators and other service providers;



-
2. **Services to ensure the quality and reliability of the data e-infrastructure**, including certification mechanisms for repositories and certification services to test and benchmark capabilities in terms of resilience and service continuity of e-infrastructures;
 3. **Federating institutional and, if possible, private data management and curation tools and services used across or at some point of the full data lifecycle**, including approaches for identification of open data sources and data collected with sensitive or restricted access features. Services and tools should be federated on the basis of an open architecture and should offer or coordinate support to the development of Data Management Plans;
 4. **Large scale virtualisation of data/compute centre resources** to achieve on-demand compute capacities, improve flexibility for data analysis and avoid unnecessary costly large data transfers.
 5. **Development and adoption of a standards-based computing platform (with open software stack) that can be deployed on different hardware and e-infrastructures** (such as clouds providing infrastructure-as-a-service (IaaS), HPC, grid infrastructures...) to abstract application development and execution from available (possibly remote) computing systems. This platform should be capable of federating multiple commercial and/or public cloud resources or services and deliver Platform-as-a-Service (PaaS) adapted to the scientific community with a short learning curve. Adequate coordination and interoperability with existing e-infrastructures



-
- 6. Support to the evolution of EGI (European Grid Infrastructure) towards a flexible compute/data infrastructure capable of federating and enabling the sharing of resources of any kind** (public or private, grid or cloud, etc.) in order to offer computing and storage services to the whole European scientific community. The proposal will address operations for supplying services (IaaS, PaaS, SaaS) at European level, engagement of and tailoring of services to new user communities and dissemination activities.
 - 7. Proof of concept and prototypes of data infrastructure-enabling software (e.g. for databases and data mining) for extremely large or highly heterogeneous data sets scaling to zetabytes and trillion of objects.** Clean slate approaches to data management targeting 2020+ 'data factory' requirements of research communities and large scale facilities (e.g. ESFRI projects) are encouraged.
 - 8. Enable the creation of a platform and infrastructure for mining text aggregated from different sources/publishers that responds to the needs of users (researchers).** This includes the definition of technical requirements (e.g. on interoperability, metadata standards and aggregation of new services) as well as addressing legal and contractual issues to serve the needs of text mining communities. The project should also provide consulting and counselling services to solve problems related with the legal framework and permissions to text mine collections, and to advise researchers on the benefits and practice of text mining. The development of the proposed platform and services should be informed by the studies on policy and licencing issues associated with Text and Data Mining that will be funded from the Call on "Science with and for Society"

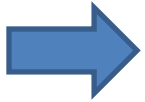
Timeline: publication: **11 December 2013 // deadline: 14 April 2014**

Indicative budget: 2014: 33 M, 2015: 17 M

Instrument: Research & Innovation Action

Note: The 2015 budget will be dedicated to projects selected from a reserve list drafted in 2014 but not funded. Maximum 8 M€ for subtopic (6 - EGI).





General considerations:

- This call looks tuned to an articulated ecosystem made of a series of FP7 and new projects which are aimed to be transversal.
 - The role to play for us is to propose “use cases” with our project and show intention to implement common solutions and extendable.
- e.g. Topic: 1) Service providers and private sectors new models (mostly driven by agencies); 2) &3) certification mechanics for “observatory-mode” data access paradigm (some commonalities with some AP future projects); 4) virtualization of “analysis framework and archive access” (idem); 5) federating heterogeneous cloud resources: the worldwide type of some of AP projects could represent stakeholders.
- However in any of the five we should target cooperative efforts with other partners (CERN, ESO, ESA and other existing e-infrastructures)
 - One more comprehensive approach is the “Users Forum” proposed by CERN to prepare the way to an interdisciplinary global new “e-infrastructure”.
 - (EGI)6 topic could more appropriate (and secure) than topic 5) for the cloud development for our community. We should maybe verify our role within it.
 - 7) is really ambitious if we can merge some main “Big Data” projects (CTA, EUCLID, LSST et al.) towards a common development work around DBs.



Specific challenge: Establishing a limited number of Centres of Excellence (CoE) is necessary to ensure EU competitiveness in the application of HPC for addressing scientific, industrial or societal challenges. CoEs will develop a culture of excellence, both scientific and industrial, placing computational science and the harnessing of 'big data' at the centre of scientific discovery and industrial competitiveness. CoEs may be 'thematic', addressing specific application domains such as medicine, life science or energy; 'transversal' on computational science (e.g. algorithms, analytics, numerical methods etc.); or 'challenge-driven', addressing societal or industrial challenges (e.g. ageing, climate change, clean transport etc.); or a combination of these types. This topic will be carried out in the frame of the Public-Private Partnership (PPP) in HPC, contributing to the implementation of the EU strategy on High Performance Computing (HPC), in particular to achieving excellence in HPC application delivery and use.

Scope: The CoE's are expected to be:

- integrated: encompassing not only HPC software but also relevant aspects of hardware, data management and storage, connectivity, security, etc.;
- multidisciplinary: with domain expertise co-located alongside HPC system, software and algorithm expertise;
- user-driven, with the application users and owners playing a decisive role in governance; and
- distributed with a possible central hub, federating capabilities around Europe, exploiting available competences, and ensuring synergies with national/local programmes;



Proposals for CoEs will address:

- Provision of services such as: developing, optimising (including if needed re-design) and scaling HPC application codes towards peta and exascale computing; testing, validating and maintaining codes and managing the associated data; quality assurance; co-design of hardware, software and codes; consultancy to industry and SMEs; research in HPC applications; and addressing the skills gap in computational science.
- Working in synergy with the European Technology Platform for HPC and with the pan-European HPC infrastructure, including by identifying suitable applications for co-design activities relevant to the development of HPC technologies towards exa-scale.
- Sustainability embracing a wide range of service models and funding from a mixture of sources, including through sponsorship by industry or hybrid public-private models. Clear business plans need to be presented in the proposal.
- Creating communities around specific codes that impact the target sectors, involving ISVs (independent software vendors) where appropriate, and exchange of best practices in particular for SMEs.
- A governance structure driven by the needs of the users. Commercial management expertise will be needed along with technical expertise to manage industry clients and supply chains.

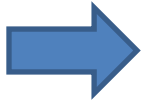
Timeline: publication: **11 December 2013 // deadline: 14 April 2014, 12 March 2015**

Indicative budget: 2014: 22 M, 2015: 18 M

Instrument: Research & Innovation Action

Note: 8-10 CoEs are expected to be funded in this Topic in order to test the concept. **A follow up Call is expected in the future that will build on the results and lessons learnt from the present Call. International co-operation is encouraged. The 2015 budget will be dedicated to projects selected from a reserve list drafted in 2014 but not funded.**

- Call 3 - e-Infrastructures
 - EINFRA 4-2014 – Pan-European HPC infrastructure and services
 - EINFRA 5-2015 – Centres of Excellence for computing applications
 - EINFRA 6-2014 – Network of HPC Competence Centres for SMEs



General considerations:

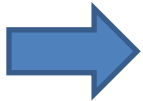
- These calls (5-2015 in particular) appear as tuned also to the federated role of our major Computing Centers that with AP projects would be HTC (High Throughput Computing) Tier0s and potential Centre of Excellences. Unfortunately these calls are really HPC dedicated. However meeting with CERN has provided important perspectives for the “second round” calls.
- The AP community should be in the User Forum and the AP agencies bring the AP use cases. We should follow the evolution of the CERN proposal.

Specific challenge: GÉANT is recognised as the European communications commons that supports the rise of compute- and data-intensive collaborative research and education through innovative services, operational excellence and global reach. There is a clear need to further develop and maintain GÉANT in this role.

Timeline: publication: 11 December 2013 // deadline: 15 September 2014

Indicative budget: 2015: 25 M with Specific Grant Agreement established under the GEANT FPA // Maximum 100% re-imbursement rate for all costs.

Instrument: Specific Grant Agreement established under the GÉANT FPA and implemented as a Research & Innovation Action



General considerations:

- We use to neglect it but actually it deserves some exploration: although the proposal are expected to be submitted solely by legal entities operating the NRENs we should investigate the cooperation for: connectivity of remote AP sensors (e.g. CTA is a use case) ; pan-European (and not only) dedicated network for distributed computing model and big data archives of future AP projects.

Specific challenge: There is yet considerable potential and room for development in the use of virtual research environments. The objective is to address this challenge by supporting capacity building in interdisciplinary research communities to empower researchers through development and deployment of service-driven digital research environments, services and tools tailored to their specific needs. These virtual research environments (VRE) should integrate resources across all layers of the e-infrastructure (networking, computing, data, software, user interfaces), should foster cross-disciplinary data interoperability and should provide functions allowing data citation and promoting data sharing and trust.

Scope: Each VRE should abstract from the underlying e-infrastructures using standardised building blocks and workflows, well documented interfaces, in particular regarding APIs, and interoperable components. Over time VREs will be composed of generic services delivered by e-infrastructures and domain specific services co-developed and co-operated by researchers, technology and e-infrastructure providers, and possibly commercial vendors. The VRE projects should clearly identify and build on requirements from real use cases, e.g. for integration of heterogeneous data from multiple sources and value-added services for computing, simulation, and data exploration, mining and visualisation.



They should re-use tools and services from existing infrastructures and projects at national and/or European level as appropriate. Where data are concerned, projects will define the semantics, ontologies, the 'what' metadata, as well as the best computing models and levels of abstraction (e.g. by means of open web services) to process the rich semantics at machine level (the so called 'how' metadata). They may also support proof of concept, prototyping and deployment of advanced data services and environments, providing a toolset and desktop with easy to use functionalities and access to top-of-the-range connectivity and computing.

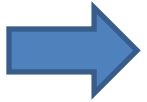
VREs may target any area of science and technology, especially interdisciplinary ones, including ICT, mathematics, web science and social sciences and humanities. Focusing on the ICT infrastructures needed for addressing the Societal Challenges is especially encouraged. Proposals should indicate the number of researchers they target as potential users.

This topic is complementary with topic 1.3.1.4, as VREs integrate data, network and computing resources for interdisciplinary research communities, whereas 1.3.1.4 addresses interoperability of services and common solutions for cluster of ESFRI and other research infrastructure initiatives in thematic areas.

Timeline: publication: 11 December 2013 // deadline: 14 January 2015

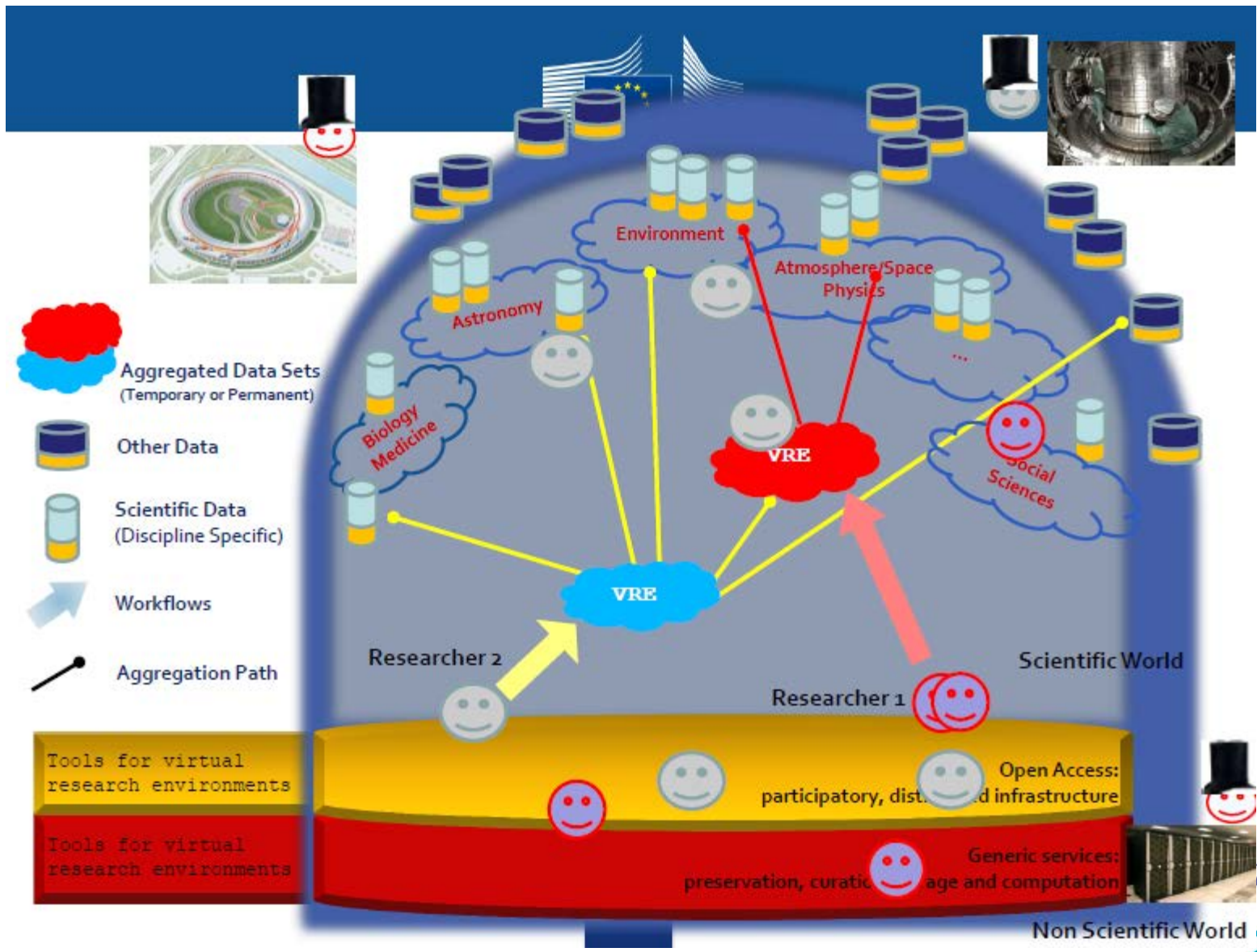
Indicative budget: 2015: 42 M

Instrument: Research & Innovation Action



General considerations:

- Some potential key issues: Big data new generation data base, new analysis software framework, Parallel programming/GPUs analysis platforms, API for new CR simulators and Grid/Cloud production on demand, Scientific Gateways / Virtual Machines. To succeed it should involve as well some of the Astro. Community for the Virtual Observatory issues or anyhow some high-level data and software products to be “used” as a service from a common user-interface.
- The correct implication of Web- and ICT scientists will be required.
- We should contact other (ESFRI?) projects in other domains to explore the potential cooperation in a same VRE.
- An important commitment towards “development and deployment of generic e-infrastructure services”



Call 1 – Developing new world-class research Infrastructures

1. Design Study [...]
- 2 .Preparatory Phase [...]
- 3 .Support to the individual implementation and operation of ESFRI projects [...]
- 4 .Support to the implementation of clusters of relevant Research Infrastructures in a given thematic area**

Common challenges on a large spectrum of issues, **many related to the data aspects but not limited to them**, exist between ESFRI projects, other world class Research Infrastructures, e-Infrastructures and Integrating Activity projects.

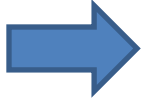
The objective of this action is to **exploit synergies**, to **optimise technological implementation**, and to ensure **coordination, harmonisation and interoperability of services, data and applications in a given research thematic area**.

To this extent, EU financial support will be provided to the coordination of common activities, the development of **common technologies** and **services** and the implementation of common and **efficient solutions on issues ranging from architecture of distributed infrastructures to distributed and virtual access management, from development of critical components to new/revised data acquisition, access, deposit, sharing and re-use policies and services**.

Among the fields one call is expected from

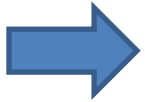
Astronomy

A project under this heading should aim at synergies in the development of data policy and solutions, including processing, storing and curating the huge data flows generated by the next generation astronomy research infrastructures such as the SKA and the E-ELT, as well as of key critical components such as the next generation correlator for the SKA and the necessary software development;



General considerations:

- Therefore there should not be room for other AstroP. Projects in Call 1.4.
- VRE could be the solution for APPEC but to add value should be built on the AP future project requirements



Maybe not one single project?

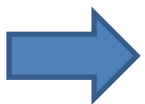
Maybe a Transversal generic work plan + a collection of work packages distributed per AP project and federated by the common and shared deliverables.

ONE MORE OPTION: CALL 1 -3 SUPPORT TO INDIVIDUAL ESFRI PROJECT: E.G. CTA

The research infrastructures identified in the ESFRI roadmap have benefitted from EU support for their preparatory phase. Some of them already moved to the implementation phase and have started their operation. In this initial phase, they may need support to become fully operational.

Scope:

Support will be provided for central coordination, operation, access provision, enlargement of the membership, training and innovation activities. Activities can include setting up and initial running of the central coordination office, enhancement of the technical architecture, detailed R&D and engineering work, development of innovative components, users' access, data management (including possible open access to data), inter-operability, standardisation, outreach, training and international cooperation. Specific attention will be given to the role of industries, in particular to facilitate where relevant the access of SMEs as a user and as a partner of the research infrastructure for technological developments e.g. through technology transfer activities as well as the development of services to industry. The activity can also support the development of Regional Partner Facilities.



G.L. general considerations:

- It could be that the gamma-ray/CTA community would apply for that and the “data management” issues would be within. HOWEVER CTA SITE DEV. CAN BE PRIVILEGED...





RESEARCH INFRASTRUCTURE Work Programme 2014-2015 (DRAFT)

CALL 3 → E-INFRASTRUCTURES

**MANAGING,
PRESERVING AND
COMPUTING WITH BIG
RESEARCH DATA**

**E-INFRASTRUCTURES
FOR OPEN ACCESS**

**TOWARDS GLOBAL DATA
E-INFRASTRUCTURES:
RESEARCH DATA
ALLIANCE**

**PAN-EUROPEAN
HIGH PERFORMANCE
COMPUTING INFRASTRUCTURE
AND SERVICES**

**CENTRES
OF EXCELLENCE
FOR COMPUTING
APPLICATIONS**

**NETWORK OF
HPC COMPETENCE
CENTRES FOR SMES**

**PROVISION OF
CORE SERVICES
ACROSS
E-INFRASTRUCTURES**

**RESEARCH AND
EDUCATION
NETWORKING –
GEANT**

**E-INFRASTRUCTURES
FOR VIRTUAL RESEARCH
ENVIRONMENTS (VRE)**

CALL 4 → SUPPORT TO INNOVATION, HUMAN RESOURCES, POLICY AND INTERNATIONAL COOPERATION FOR RESEARCH INFRASTRUCTURES

**In blue:
Appec interest**

**E-INFRASTRUCTURE
POLICY DEVELOPMENT
AND INTERNATIONAL
COOPERATION**

**NEW PROFESSIONS
AND SKILLS
FOR E-INFRASTRUCTURES**

Conclusions (and decisions?)

- The “APPEC-Computing team/contact person” inform about cooperative and transversal participations to the various project under
EINFRA 1-2014 – Managing, preserving and computing with big research data
(8 sub-topics where transversal activities can host also APPEC community)
- For the call: **EINFRA 9-2015 – e-Infrastructures for virtual research environments (VRE)**: we need one contact “data management” person per main AP projects in order to start planning the needs, the current expertise, the common development for shared deliverables... Some ideas are well already established...
- Coherence with any “data and e-Science” activity proposed within the Call 1 ESFRI project (e.g. CTA, SKA,...)
- Take part to the “User Forum” initiatives (which might be supported with EINFRA 1? Or maybe “Coordinated” centrally with CALL 4-7 to monitor e-infra. interoperability)
- “Centre of Excellences” later on (but with direct agencies commitments).
- Marie Curie training network and support to “data scientists” to be considered...

What next?

Verify the real wish to be part of a common project !

Start official survey among AP collaborations and then meet concerned people