

The Role & 2018 expectations of the EOSC HLEG

Silvana Muscella, CEO Trust-IT Services; Chair of the High Level Expert Group on European Open Science Cloud

2nd ASTERICS-OBELICS Workshop

16-19 October 2017, Barcelona, Spain.



H2020-Astronomy ESFRI and Research Infrastructure Cluster
(Grant Agreement number: 653477).

Contents

- What is the EOSC HLEG expected to do ?
- Starting with some solid Building Blocks / «Assets»
- An EOSC in Practice
- Some insights from EOSC European Commission
- A Procurement Model
- How are the EOSC Science Demonstrators (SDs) doing
- Mapping these SDs onto a common service model going forward
- What's in it for the Astro Physics community & what role can they play?

EOSC HLEG: Mission & Context

- The EOSC HLEG has been mandated to support the European Commission in the set-up data-driven infrastructure that builds on:
 - what exists,
 - caters for the whole scientific community and
 - brings together the building blocks developed through different categories of stakeholders to provide actionable options for the design of the EOSC governance and services.
- The HLEG EOSC will provide its contribution as part of the implementation of the Horizon 2020 Work Programme 2016-17 'Europe in a changing world: inclusive, innovative and reflective societies', specifically under "25. Support to Open Science“

User-Oriented Open Science – Pragmatic Approach

- An «EOSC in practice»
- The Group wishes to provide a further elaboration and insights on the basis of outputs from funded related projects, national projects or other community fora (ie: EOSC Pilot & EOSC hub, RDA, FAIR, OSSP, etc.)
- Structure thinking about EOSC into **technical, organizational and financial** aspects
- Introduce and analyse **Incentive Mechanisms** to whet the appetite for all stakeholders
- Services should be offered by **different, existing infrastructures** not by one single infrastructure
- Have a **functioning methodology** that is practical that can be followed
- Functional Thinking vs Principle thinking
- User oriented vs Service thinking
 - Users using commercial solutions & find out why & take into account
 - Need to know what we are building to design the Governance Framework – think in terms of options
 - What are the architecture options, define all of them to be able to make choices

Introduce a Business Thinking oriented approach. The EOSC should be an environment for sharing and making data available to all. Usability and usefulness of the EOSC are central

Other relevant past Best Practices – Being Aware of our Assets

Ensuring long-term sustainability of research data across all disciplines, with strong & flexible governance model

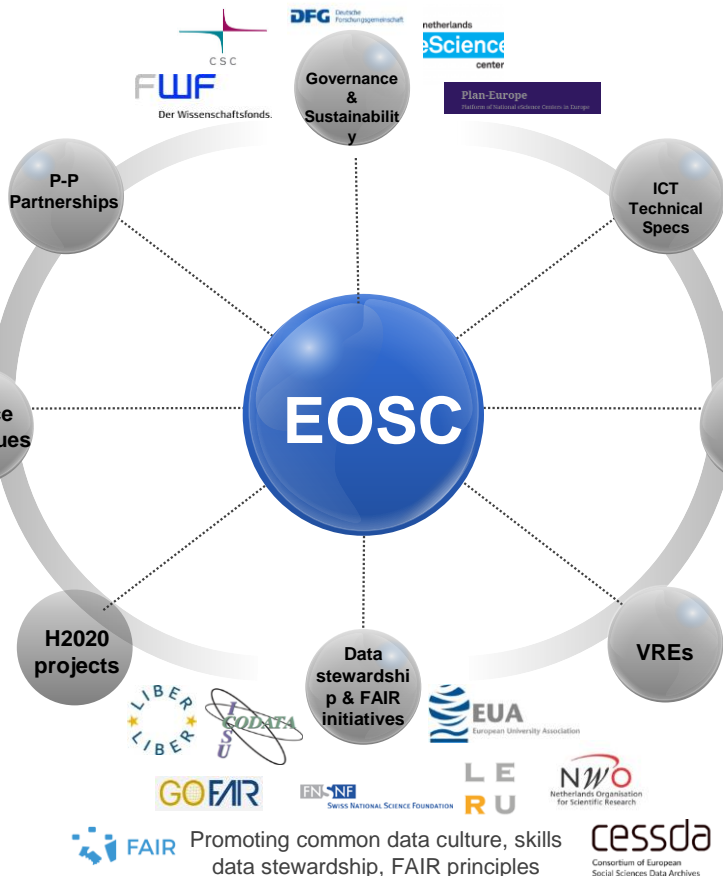
Boosting the value of EU Big Data research as part of the EC strategy to promote a data economy



Harmonising stakeholder's offerings under a common service catalogue framework



Leveraging on existing best practices & real use cases



Leveraging RDA as a global neutral discussion forum as well as the technical & social infrastructure solutions & ICT technical specifications developed by RDA WG & IG that enable data sharing, exchange, & interoperability

Building on the implementation & governance best practices, & multi-disciplinary approach to community building of ESFRIs & ERICs

EOSC data catalogue capitalise on recommendation from major users like ELIXIR, EOS; ENES & GEOSS

Facilitating communication between scientists & computer infrastructures

Leverage on existing “Building Blocks” & relevant Working Service Models

EOSC – the way forward

EOSC Declaration (Sept 2017)

- key input for the Roadmap
- for agreement and poss. specific commitments by stakeholders
- basis for follow-up discussion with MS (ERAC SWG 5+)

EOSC Stakeholders Forum (Nov 2017)

- open to ALL categories represented at the Summit & endorsing EOSC Declaration
- by end August, EC to define application procedure & working modalities
- initially based on EOSCpilot project (Stakeholder Engagement Event 28-29 Nov)

EOSC Roadmap (Dec 2017)

- **Governance structure** , incl. mandate & selection procedure for Executive Board

- Broad (federated) architecture



EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR RESEARCH & INNOVATION
The Director-General

Brussels, 10 July 2017

EOSC Declaration

RECOGNISING the challenges of data driven research in pursuing excellent science;

GRANTING that the vision of European Open Science is that of a research data commons, widely inclusive of all disciplines and Member States, sustainable in the long-term,

CONFIRMING that the implementation of the EOSC is a process, not a project, by its nature iterative and based on constant learning and mutual alignment;

UPHOLDING that the EOSC Summit marked the beginning and not the end of this process, one based on continuous engagement with scientific stakeholders, the European Commission,

PROPOSES that all EOSC stakeholders consider sharing the following intents and will actively support their implementation in the respective capacities:

Data culture and FAIR data

- **[Data culture]** European science must be grounded in a common culture of data stewardship, so that research data is recognised as a significant output of research and is appropriately curated throughout and after the period conducting the research. Only a considerable cultural change will enable long-term reuse for science and for innovation of data created by research activities: no disciplines, institutions or countries must be left behind.
- **[Open access by-default]** All researchers in Europe must enjoy access to an open-by-default, efficient and cross-disciplinary research data environment supported by FAIR data principles. Open access must be the default setting for all results of publicly funded research in Europe, allowing for proportionate limitations only in duly justified cases of personal data protection, confidentiality, IPR concerns, national security or similar (e.g. 'as open as possible and as closed as necessary').
- **[Skills]** The necessary skills and education in research data management, data stewardship and data science should be provided throughout the EU as part of higher education, the training system and on-the-job best practice in the industry. University associations, research organisations, research libraries and other educational brokers play an important role but they need substantial support from the European Commission and the Member States.
- **[Data stewardship]** Researchers need the support of adequately trained data stewards. The European Commission and Member States should invest in the education of data stewards via career programmes delivered by universities, research institutions and other trans-European agents.
- **[Rewards and incentives]** Rewarding research data sharing is essential. Researchers who make research data open and FAIR for reuse and/or reuse and reproduce data should be rewarded, both

EC gathering endorsement and commitments on the EOSC declaration

- EOSC Declaration won't be modified
- Action list/commitment its evolutive part
- Roadmap will follow (and will have future updates)

By endorsing the principles of the EOSC Declaration, stakeholders signal their intention to be involved in the making of the EOSC (eg. by taking specific action, by joining the Executive Board, by providing inputs via the annual stakeholder forum, or again by joining consortia , to implement the EOSC via Horizon 2020).

Objectives:

- Identifying 'doers' out of wide range of stakeholders
- Promote accountability



EOSC Declaration / Services & Architecture

[EOSC architecture]

The EOSC will be developed as a data infrastructure commons serving the needs of scientists...EOSC will federate existing resources...service provision will be based on local-to-central subsidiarity...Users should contribute...continuous dialogue to build trust and agreements...

[Implementation]

Resources, components and initiatives of pan-European relevance will be federated on the basis of objective criteria...to deliver EOSC main functionalities...

[Legacy]

The EOSC should incentivise the re-use of existing building blocks, state-of-the-art services and solutions...It should facilitate learning from the past...

[User needs]

Users should see the EOSC as a one-stop-shop...Services and functionalities shall be user driven and determined by clear use cases...

[Service provision]

Research Data Infrastructures, e-infrastructures and commercial operators will develop and provide services based on user needs...Services will be offered at highest Technology Readiness Levels (TRLs) and kept future-proof...avoid lock-in by individual service providers...

[Service deployment]

The EOSC shall support different deployment models...Software sustainability should be treated on an equal footing as data stewardship.

Source : Carmela Aseró European Commission, DG Research & innovation e-IRG Workshop, Tallinn, 3 October 2017

EOSC Roadmap

Governance – (3-layered structure, mandates, working procedures)

- Member States and EC (strategic)
- Executive Board (operational)
- Stakeholders Forum (advisory)

Architecture – launching EOSC stage 1 by 2020

- Supporting EOSC components (geographical & thematic)
- Core data & services through an EOSC Hub
- Advanced services by certified actors
- Catalogues of data & Catalogues of services
- Rules of participation
- EOSC Portal as a universal entry point
- FAIR data / FAIR Action Plan

Financing – stage 1 / stage 2

- H2020 – RI, in particular through WP2018-20
- Developing an EOSC Business Model for long-term sustainability

EOSC: An example of commons credit model ... Others to be investigated

The Commons Credits Model Pilot is designed to provide investigators with access to cloud based computing resources as a means to seed the **American National Institutes of Health (NIH)** Commons (<https://datascience.nih.gov/commons/>) with useful digital artifacts of biomedical research.

- 2 year (originally 3 year) pilot to test this business model to facilitate researcher use of cloud resources (enhance data sharing and potentially reduce costs).
- Contract with the CMS Alliance to Modernize Healthcare (CAMH) Federally Funded Research and Development Center (FFRDC) managed by the MITRE corporation

Mapping this to EOSC:

- The NIH already has a full **set of standards** in place, ranging from interoperability to business relationships to access profiles, vital for EOSC
- **Finding the right cloud(s) provider(s)** – For the EOSC, choosing a legally-compliant cloud provider will be paramount. With a view to protecting personal data, the GDPR poses specific obligations onto the cloud provider, from data security to subcontracting conditions;
- **Creation** of a **marketplace** for "conformant" cloud providers – The selection of all interested cloud providers abiding by NIH standards is thoughtfully assigned directly to researchers (pre-selected via grants),
- They are provided with a # of "**coins**" they spend to obtain cloud services from provider of their choice.
- Underlying **plan** goes visibly beyond earmarking funds, rather it aims **to create a pool of top** (= most reliable, legally compliant) **cloud providers** that would be governed by market rules and whose services would become cheaper overtime, due to critical mass achievement (concentration of researchers' preferences).

Trusting only compliant cloud providers offering adequate guarantees with respect to data storage and processing services, foster and rapidly spread good practices Other Examples ?

Who can provide cloud resources?

Commons Credits Model Pilot Provider Conformance Requirements

01 MAR 2016

Definitions:

1. Digital Object: An electronic artifact, including, but not limited to data, software, metadata, and/or workflows that can be stored or manipulated in an electronic information system.
2. Digital Object Steward: The individual or organization that created and/or controls a digital object and that has formal responsibility for its security, integrity and/or availability.
3. Investigator: A user who interacts with the Commons.
4. Cloud Credits Coordinating Center: The organization (including subcontractors, where relevant) that distributes computing resources (Credits) to stewards and prospective users of digital objects for use with providers.
5. Provider: An organization that makes a conformant cloud infrastructure available to users of the Commons and accepts NIH Commons Credits
6. Reseller: an entity which provides capabilities as a result of reselling or providing access to another provider's capabilities.
7. FISMA: Federal Information Security Management Act (44 USC § 3541 *et seq*) enacted as Title III of the E-Government Act of 2002, defines federal agency responsibilities for Information Assurance.
8. NIST: National Institute of Standards and Technology
9. IaaS: Infrastructure as a Service, based on NIST definitions¹
10. PaaS: Platform as a Service, based on NIST definitions
11. SaaS: Software as a Service, based on NIST definitions
12. REST: Representational State Transfer; an implementation independent protocol for exchanging information over networks.
13. SLA: Service level agreement
14. CPU: Central Processing Units
15. VM: Virtual Machines
16. FTP: File Transfer Protocol
17. SFTP: Secure (SSH) Files Transfer Protocol.

General Requirements:

1. Providers must offer one or more of the following cloud services: IaaS, PaaS or SaaS. When included in a provider's offering to reduce the effort needed for developing or running computational or visualization tools, PaaS or SaaS-only offerors must also include an available data access API (or equivalent), which can be used by recipients of credits and the general public,

- “Conformant Providers”
- Can be IaaS, PaaS, SaaS
- Meet standards promulgated by NIH for:
 - Capacity
 - Accessibility
 - Interfaces
 - Identifiers and Metadata
 - Networking
 - Authentication/Authorization
 - Information Assurance



What we can learn about user needs through EOSCpilot's Science Demonstrators (SDs)

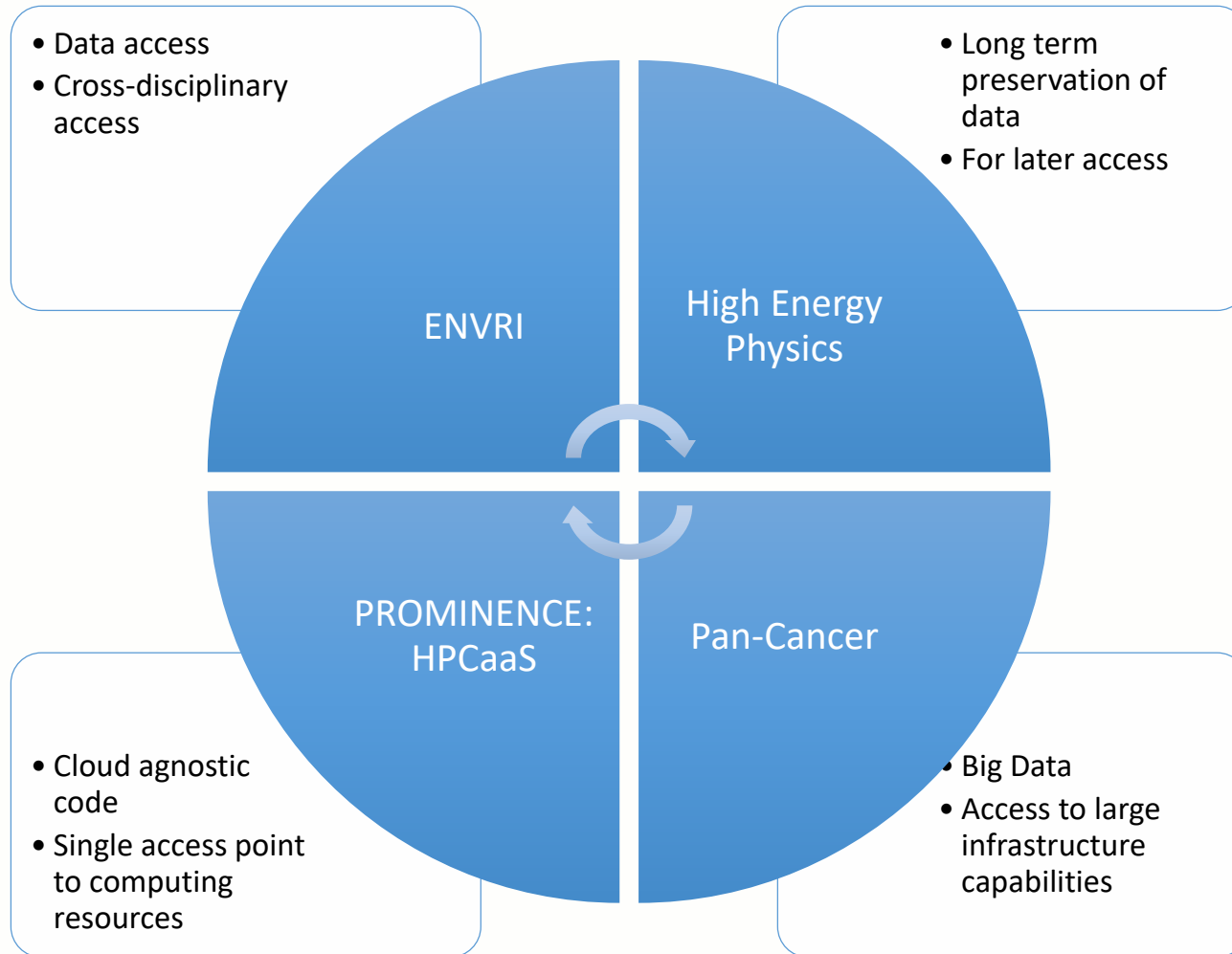


- ❧ **Advanced Virtualisation:** Transparent access to Computational Resources Progressing with a framework which will help conceptualise the range of stakeholders and interoperability objectives

- ❧ **A Market Place** for modular services to be combined in traceable **Workflows:** Build, Run and Share Workflows
 - ❧ Data workflows for extraction, analysis, curation, publication, preservation, etc.
 - ❧ The workflows should run transparently on EOSC Computational Resources



Transparent Access to Computational and Data Resources





Modular and Traceable Workflows

EPOS/VERCE

- Abstract **workflows** for distributed data-intensive applications
- Support for composition
- Executable in numerous parallel environments

Genome SD

- Framework for computational workflows
- Write complex parallel **workflows**
- Transparent deployment on multiple platforms

CryoEM

- Scipion: an image processing framework to “glue” software for **workflow** combinations
- Traceable and Reproducible **workflows**

LOFAR

- Data **workflow** project to facilitate data access
- Both to power and non power user

Common Keywords: Workflows, Container technologies (Docker), User-friendliness, Cloud agnostic



- EOSC as a broker for **Cloud Resources**
- EOSC as a **market place of atomic micro-services** that can be pipelined and combined in traceable **workflows**.
- Necessity to distinguish the power user from the non power user
 - The power user wants flexibility and large parametrisation possibilities
 - The non power-user can work in a more constrained environment, but it needs to be more user-friendly
 - -> One idea would be to start with the power user flexible services, and then offer packaged workflows to the non-power user.

The EOSCpilot WPs and Science Demonstrators are a mine of ideas and examples on how to set these two services up

What's in it for the Astrophysics Community?

- Contribute to shape the design, while still in this phase
- Support the Workflow technologies and pipelines that allows users to change settings, input datasets, etc
- Consider the LOFAR data model (Physical Sciences / Astronomy)
 - Easy Access to LOFAR data & knowledge extraction through the Open Science Cloud
- Advanced virtualisation with transparent allocation of computational resources of all kinds of sizes.
- Other

Looking forward to receiving your comments on the proposed ideas

Acknowledgement

- H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).