



**EOSC**pilot  
The European Open Science  
Cloud for Research Pilot Project  
[www.eoscpilot.eu](http://www.eoscpilot.eu)



**Science & Technology  
Facilities Council**

# The European Open Science Cloud Pilot

Brian Matthews

Science and Technology Facilities Council



**EOSC**pilot  
The European Open Science  
Cloud for Research Pilot Project  
[www.eoscpilot.eu](http://www.eoscpilot.eu)



# European Open Science Cloud - its in the air!

Discussed for a couple of years

- A systemic change in the modus operandi of science and research
- Affecting the whole research cycle and its stakeholders



Brussels, 19.4.2016  
COM(2016) 178 final

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN  
PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL  
COMMITTEE AND THE COMMITTEE OF THE REGIONS

European Cloud Initiative - Building a competitive data and knowledge economy in  
Europe

{SWD(2016) 106 final}  
{SWD(2016) 107 final}

Final draft

Monday 20 June 2016

## A Cloud on the 2020 Horizon

Commission High Level Expert Group on the European Open Science Cloud

Realising the European Open Science Cloud: first report and recommendations

Preface by Barend Mons, Chair



*This report aims to lay out a high level, living roadmap for the realisation of the European Open Science Cloud (EOSC). The High Level Expert Group, with ten members from European countries, Japan and Australia, discussed extensively in several meetings, conferences, policy events and met with key stakeholders (30 November 2015) and research funders (15 March 2016). Based on these consultations, on many 'white papers' and on a range of presentations and feed-back at international meetings, we are confident that our recommendations count on a high-level of consensus amongst all stakeholders. This was a solid basis to embark on this challenging journey with the Commission, the Member States and International partners in concert.*

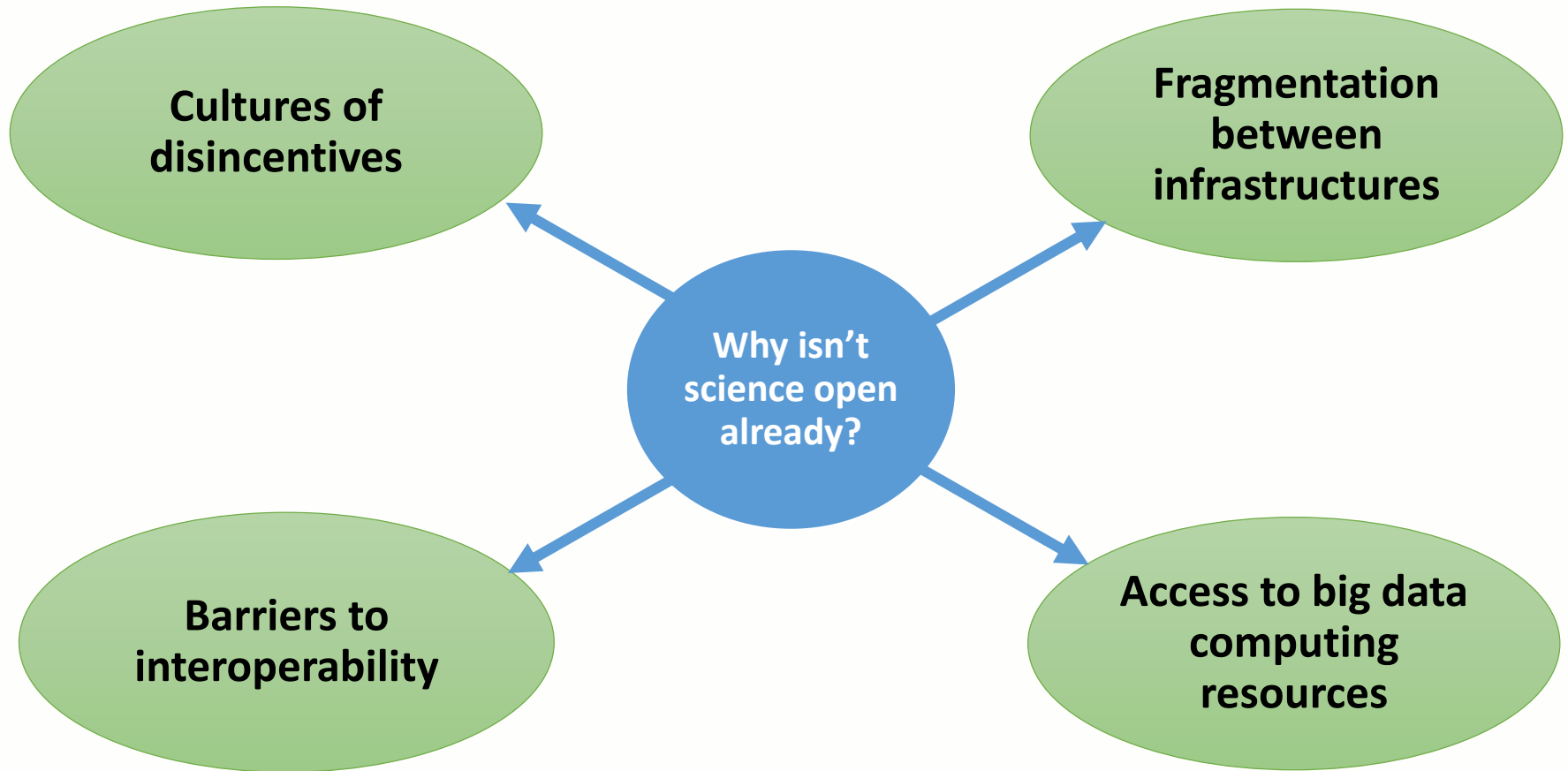
*The title of this first report may have a slightly threatening ring to it and indeed, if we do not act, there might be a looming crisis on the Horizon. The vast majority of all data in the world (in fact up to 90%) has been generated in the last two years. Computers have long surpassed individuals in their ability to perform pattern recognition over large data sets. Scientific data is in dire need of openness, better handling, careful management, machine actionability and sheer re-use. One of the sobering conclusions of our consultations was that research infrastructure and communication appear to be stuck in the 20th century paradigm of data scarcity. We should see this step-change in science as an enormous opportunity*

Commissioner  
Carlos Moedas  
Open Science  
Presidency  
Conference  
Amsterdam, 4  
April 2016



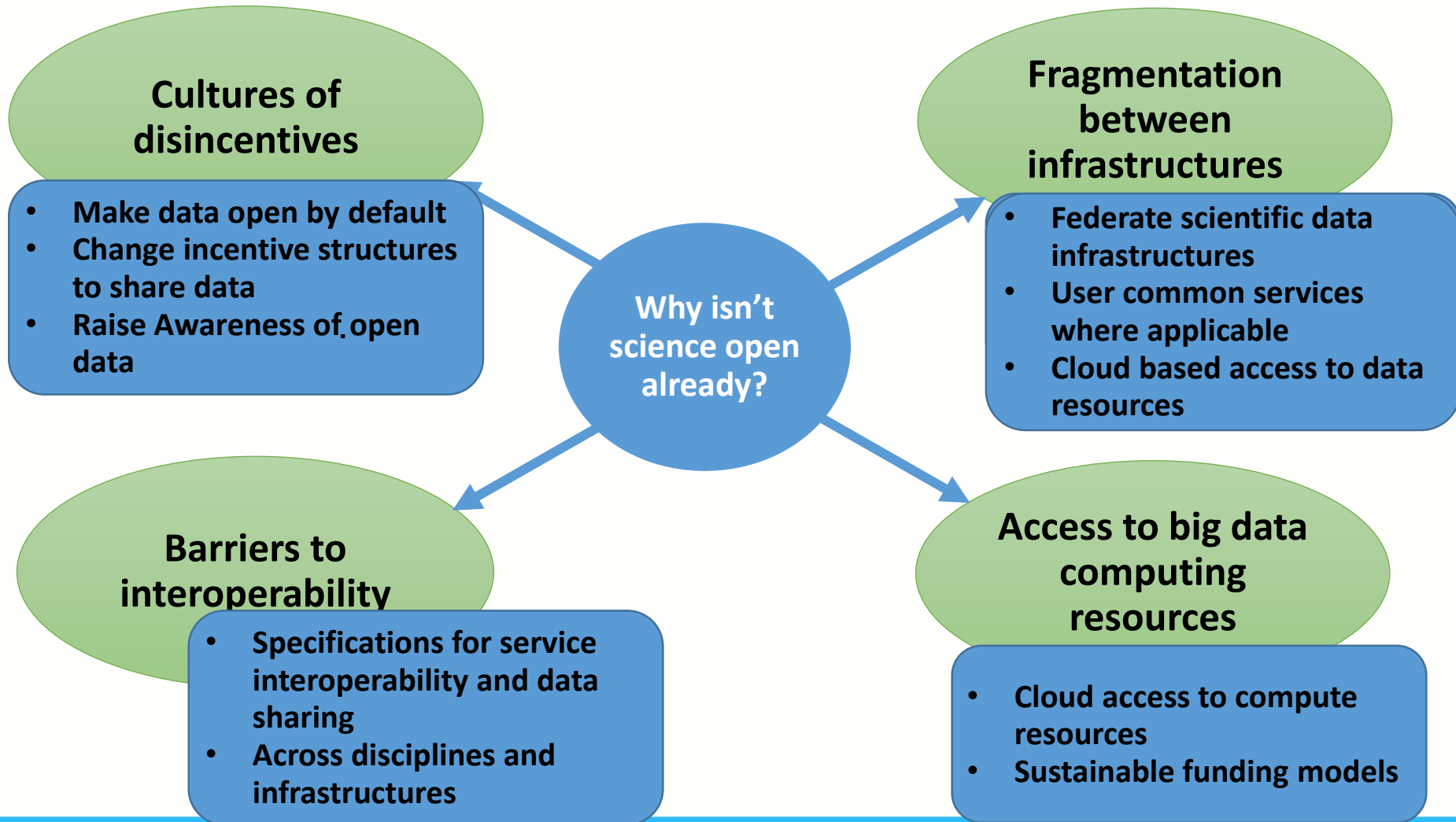


# *Why is Europe not fully tapping into the potential of data ?*



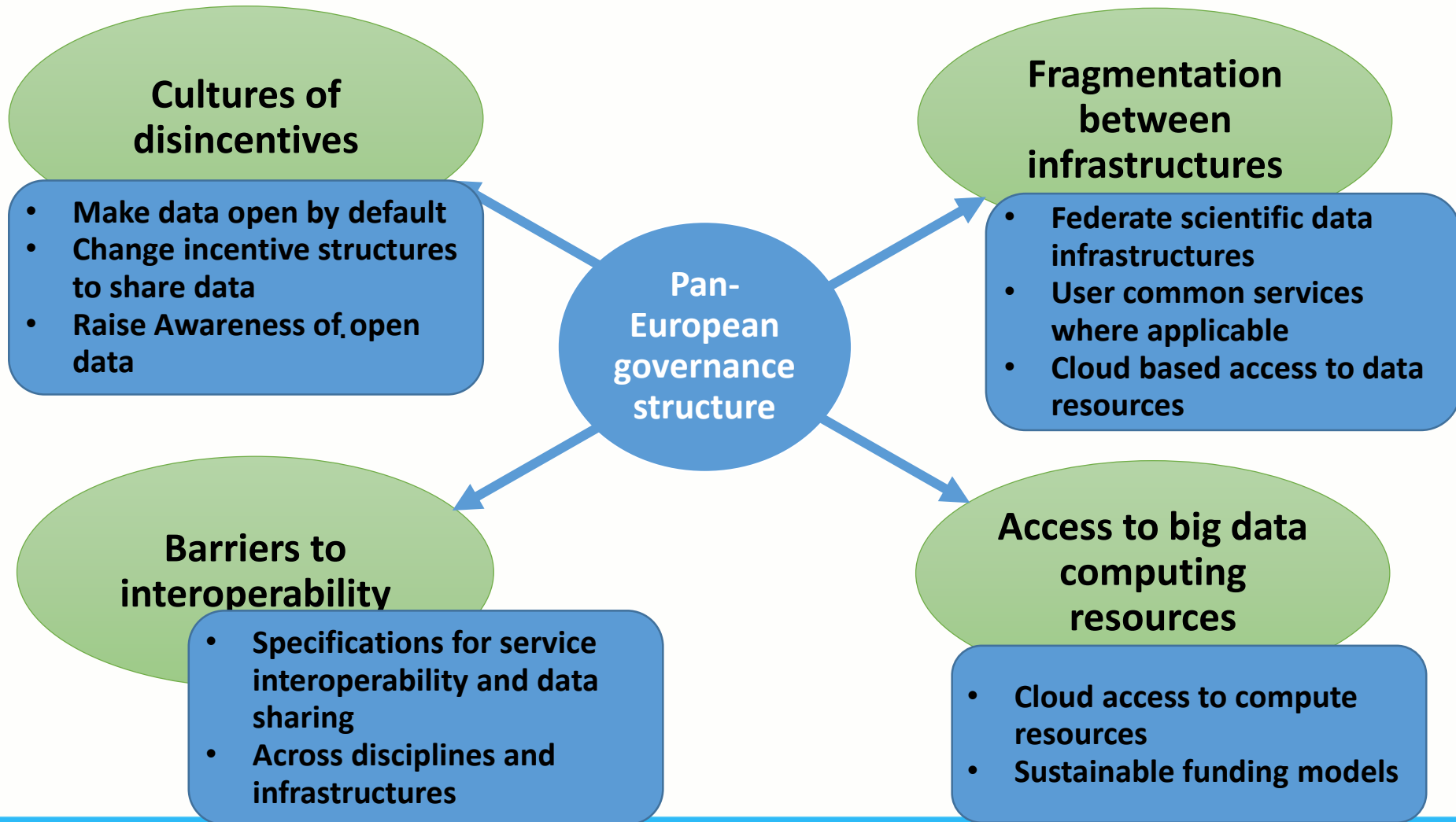


# A coordinated effort to change how research is done



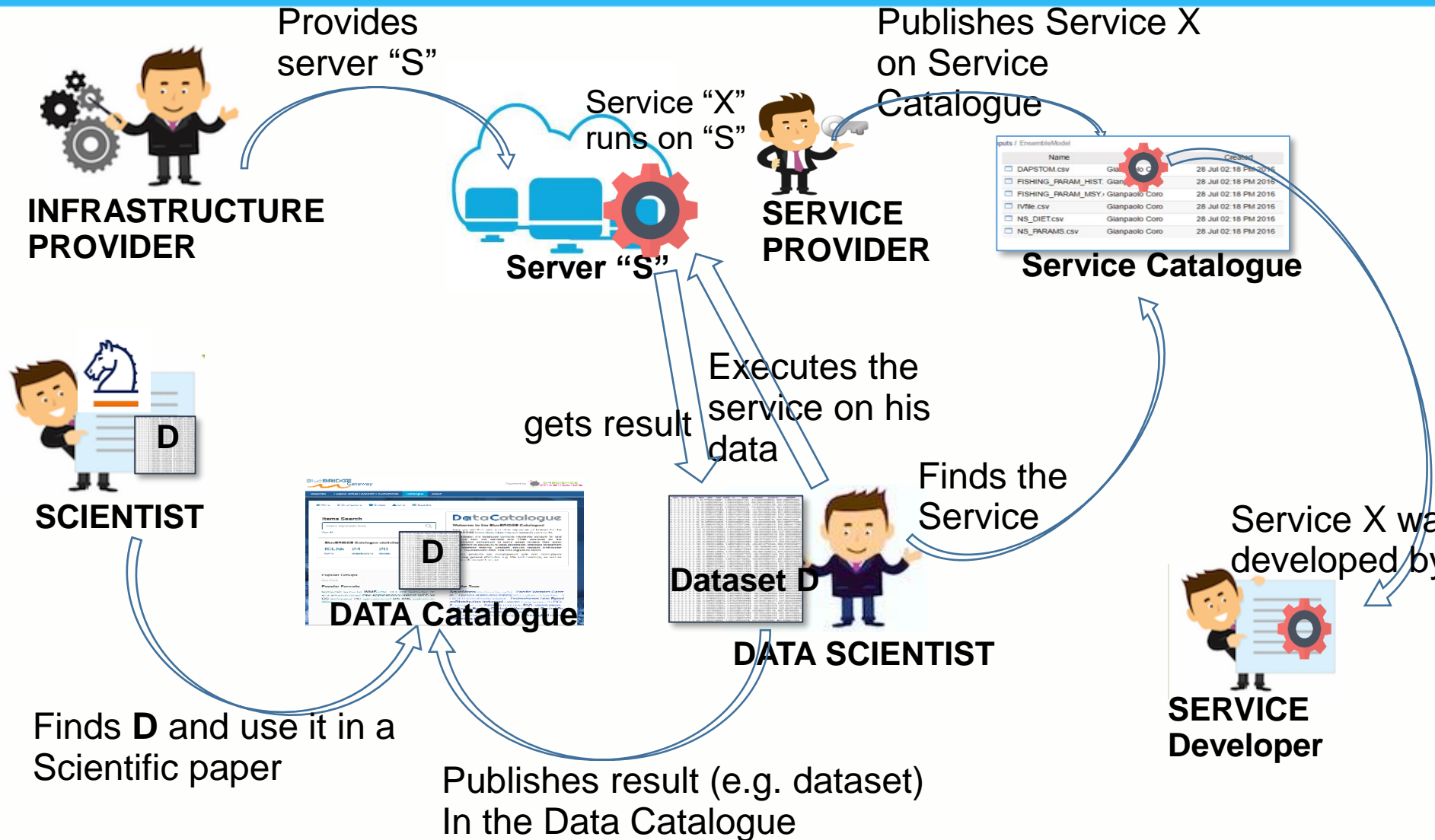


# *Propose a European Open Science Cloud*





# Typical scenario





What do we need to do?



# EOSCpilot Challenges

Three types of challenges addressed by the EOSCpilot:



Scientific Challenges are really *Opportunities*

**Scientific Challenges:** deploying the EOSC to deliver Open Science

Technical Challenges are *Barriers to overcome*



**Technical Challenges:** developing technical solutions that meet the scientific needs

Cultural Challenges are also *Barriers*



**Cultural Challenges:** adopting new, more open ways of working





# Actions to bring about an EOSC

- Bring the current Research Infrastructures together
  - We do not want to replace their work
- Bring the e-Infrastructure projects together
  - GEANT , PRACE
  - EGI, EUDat, OpenAire
- Interoperate between their services
  - Catalogue of services
  - Allow people to select services to build new infrastructures
  - Set up appropriate rules of engagement
  - Allow access to data and compute
- Interoperate between their data
  - FAIR data catalogues – accessible outside their discipline.
  - Interoperable standards and metadata
- Allow new resources to be added
  - Cloud providers, HPC providers, data providers, service providers
- Within the common governance and resourcing processes
- And a skills and competencies framework
- Need some set of core services and processes to hold the EOSC together



## Setting the EOSC in the right direction

### First of the EOSC projects

### 10M€ over 2 years

- Jan 2017 – Dec 2018

### 33 Partners + 15 3<sup>rd</sup> parties

- Led by STFC
- A range of e-Infrastructure providers, research institutes, research consortia, across disciplines.
- EGI, EUDat, OpenAire, PRACE, GEANT
- ELIXIR, ICOS, ECRIN, BBMRI, DESY, CERN, XFEL, CEA
- STFC, CNR, DANS, DCC, BSC, MPG, CNRS

### Try to answer some basic questions

- What is the EOSC going to provide?
- How is the EOSC going to operate ?
- How is the EOSC going to change how science is done ?







# *EOSC*pilot: High Level Aims

The *EOSC*pilot project will support the first phase in the development of the EOSC.



## **Propose a governance framework** for the EOSC

-  and contribute to the development of European open science policy and best practice;

## **Develop a number of demonstrators**

-  functioning as high-profile pilots
-  integrate services and infrastructures
-  to show interoperability and
-  Demonstrate the potential benefits in a number of scientific domains;

## **Engage with a broad range of stakeholders,**

-  crossing borders and communities, to build
-  And to define the skills required for adoption of an open approach to scientific research.



# Workpackages

## 1. Governance

- Propose a governance framework

## 2. Policy

- Devise a policy environment

## 3. Demonstrators

- Use real demonstrators to drive the requirements for the EOSC

## 4. Services

- Specify service architecture, catalogue and pilot services

## 5. Interoperability

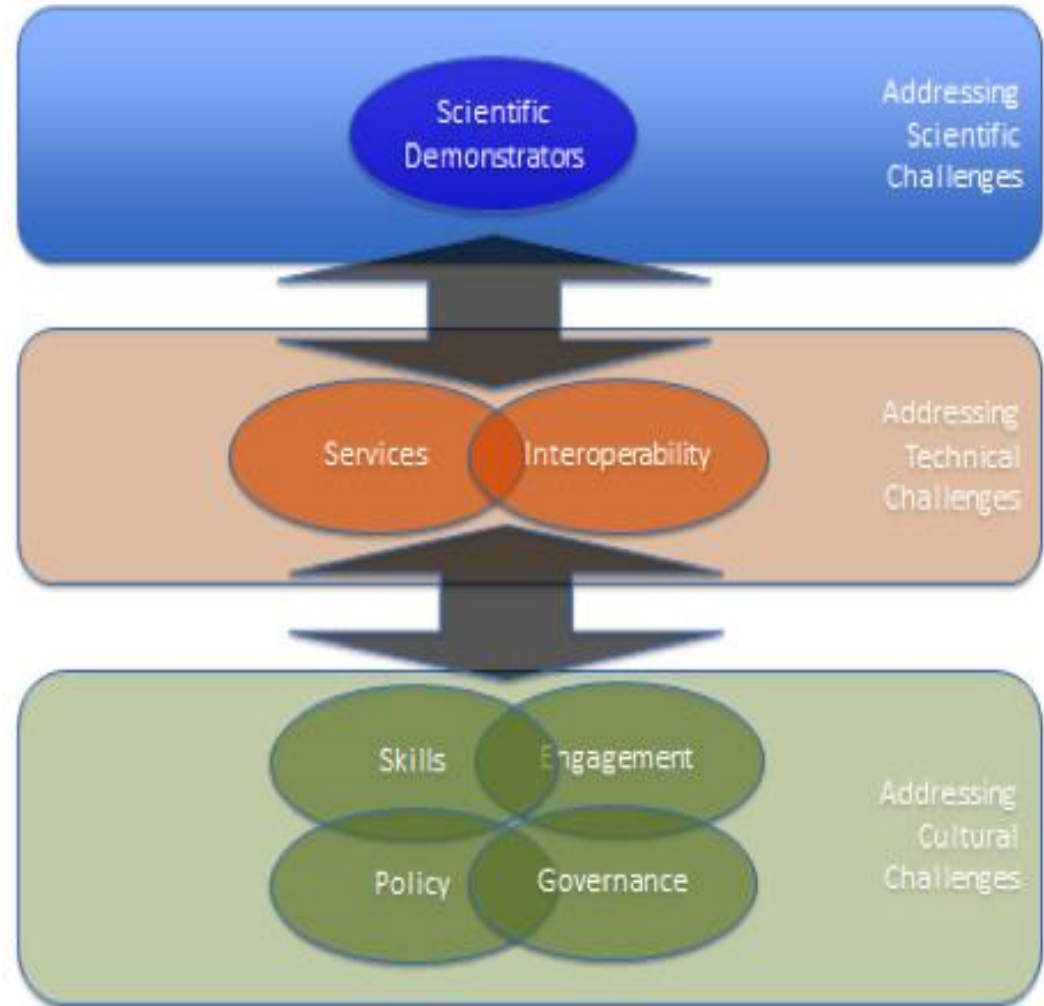
- Identify interfaces and standards to drive interoperability

## 6. Skills

- Specify a skills and competencies framework for the EOSC

## 7. Engagement

- involve as many stakeholders as possible.



## First 5 Demonstrators

- **Environmental & Earth Sciences** - ENVRI  
Radiative Forcing Integration to enable harmonised data access and integration across multiple research communities
- **High Energy Physics** - WLCG: large-scale, long-term preservation and re-use of HEP data in the EOSC open to other researchers
- **Humanities** – TEXTCROWD: Collaborative semantic enrichment of text-based datasets by make new software available on the EOSC.
- **Life Sciences** - Pan-Cancer Analyses & Cloud Computing within the EOSC to accelerate genomic analysis on the EOSC
- **Physics** - The photon-neutron community to improve the community's computing facilities by creating a virtual platform for all users

## Second 5 Demonstrators

- **HPaaS for Fusion** - Culham Science Centre, UK
  - **Life Science Leveraging EOSC** to offload updating and standardizing life sciences datasets and to improve studies reproducibility, reusability and interoperability-CRG, Spain
  - **Seismology**: EPOS Virtual Earthquake and Computational Earth Science e-science environment in Europe- University of Liverpool, UK
  - **CryoEM** Linking distributed data and data analysis resources as workflows in Structural Biology with cryo-Electron Microscopy: Interoperability and reuse CSIC, Spain
  - **Astronomy Open Science Cloud** access to LOFAR data - ASTRON, NL
- 5 more demonstrators to be selected in the autumn.



# Architecture Framework



**Use functions and tools** to perform their domain specific **research activities** and to **collaborate**

Scientists

**Develop new analytical models**, new processes and tools to **analyse data and derive knowledge**

Data scientists

**Develop services for scientists** by understanding their **requirements** and the Open Science vision

Service developer

Service provider

**Operate and provide access** to their own **portfolio of services** according to declared SLAs

Managers

**Facilitate the operation, assistance and quality assurance** of the EOSC system and the **coordination** among its different stakeholders

Infras provider

**Provide access to** computational, storage and network **resources**, according to declared **SLAs**



# HOW: EOSC as a System-of-Systems



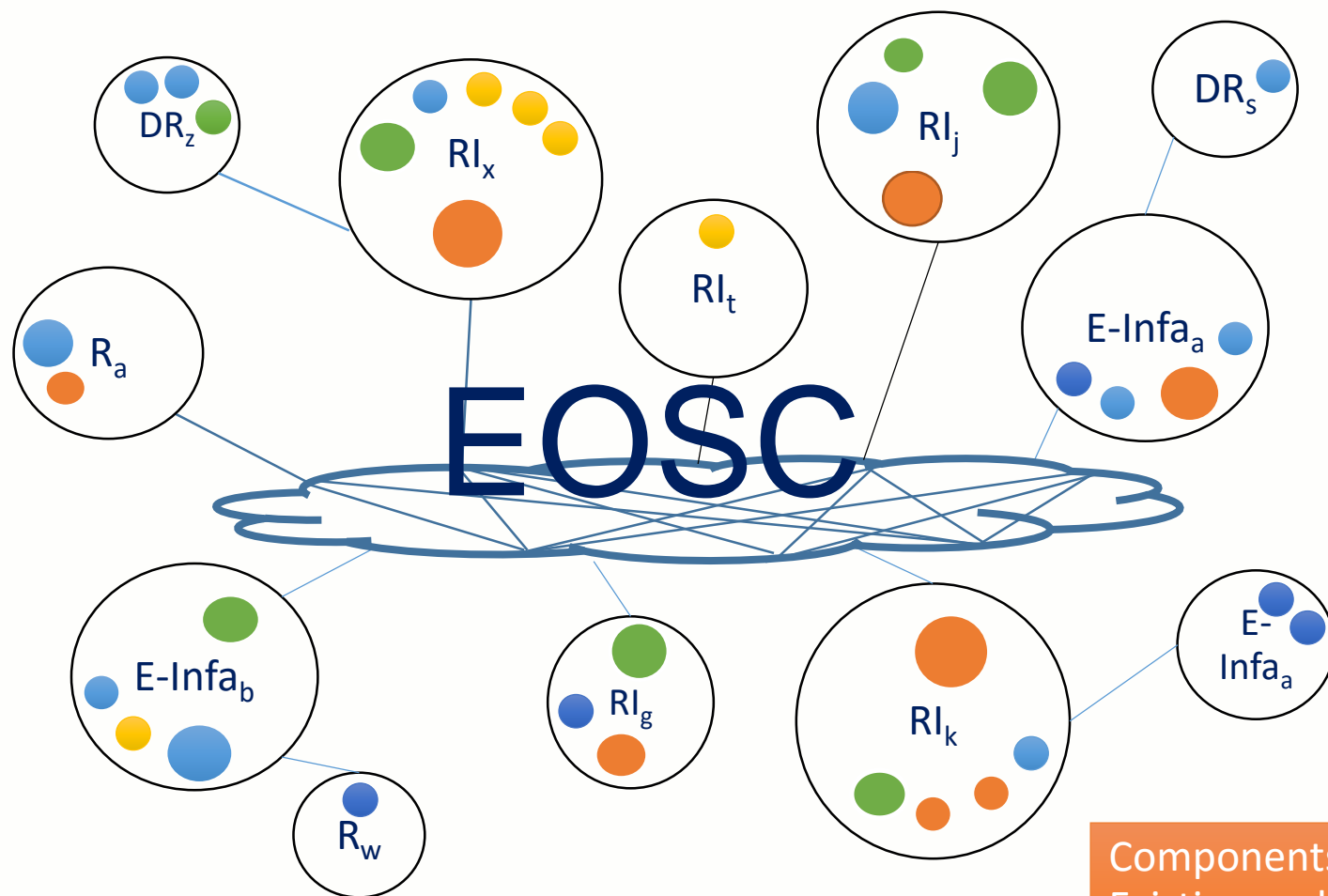
**EOSC**

Components:  
Existing and emerging RIs,  
e-Infras, data repositories,  
registries,...





# HOW: EOSC as a System-of-Systems



Components:  
Existing and emerging RIs,  
e-Infras, data repositories,  
registries,...



# System-of-systems (SoS): characterising properties

## Operational and managerial independence

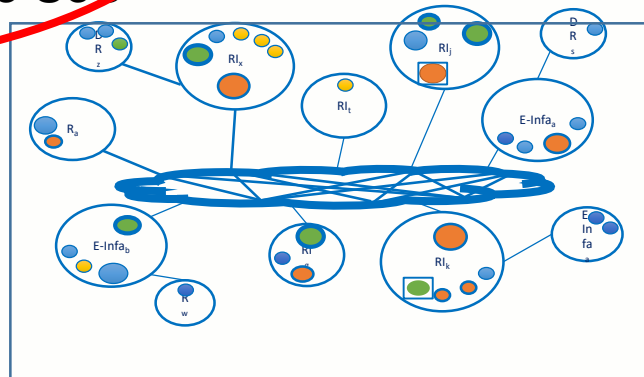
**Each system is independent** and it achieves its purposes by itself and for its own objective rather than for the purposes of the SoS

## Evolutionary development

A SoS **evolves** with time and experience

## Geographical distribution

A SoS is **distributed** over a large geographic extent



## Added-Value

A SoS has capabilities and properties that **do not reside in the component systems**

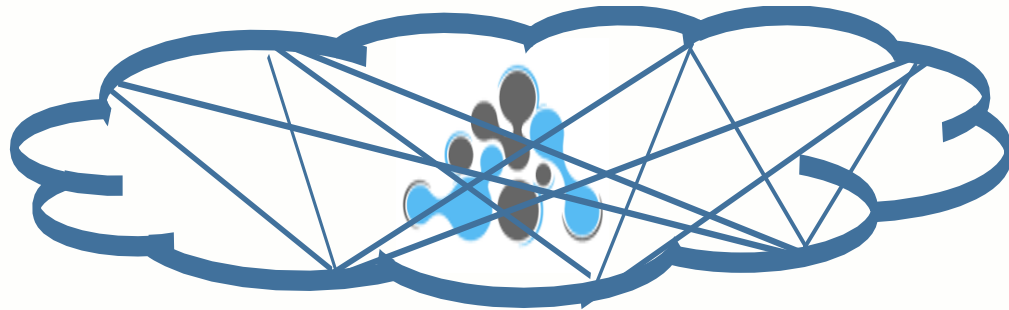
## Heterogeneity of constituent systems

A SoS consists of multiple, heterogeneous, operating systems **embedded in networks at multiple levels**



## EOSC enabling services “categories”

- Validation and monitoring of EOSC principles of engagement
- Actors & Services coordination and monitoring
- Bridging across components
- Allowing ***Interoperability across services, data, infrastructures***





# EOSC enabling services examples:

## Validation and monitoring of EOSC principles of engagements

- Scientists must have a known identity, artifacts deposited and findable
- Data Scientists produced Developed software services in EOSC must have an explicit licence
- Infrastructure providers must ensure the established SLA

## Actors & Services coordination and monitoring

- Licence compatibility checker
- ...

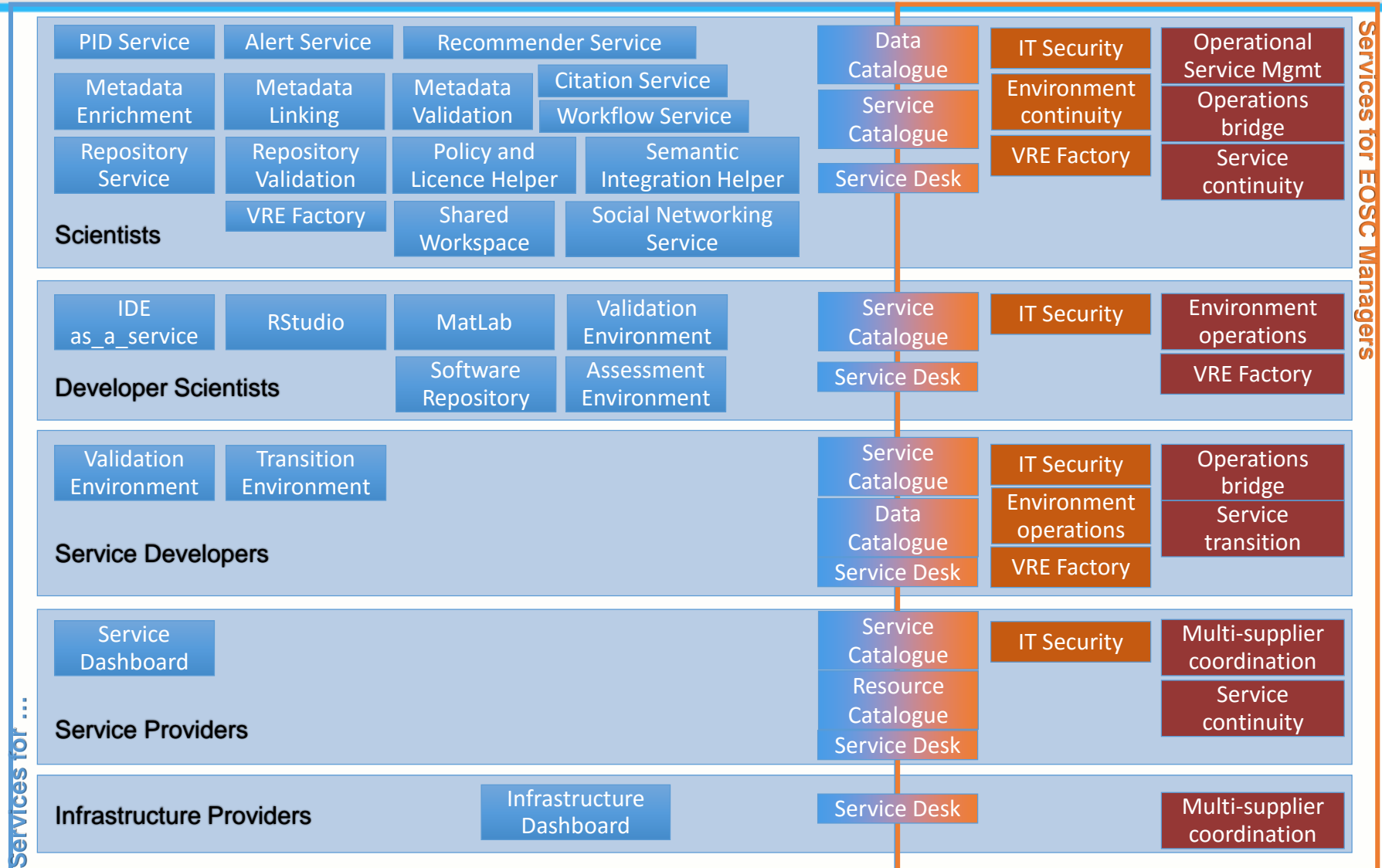
## Bridging across components

- EOSC Service & Data catalogue
- Metadata format transformer
- ...



# Proposed EOSC class of Services

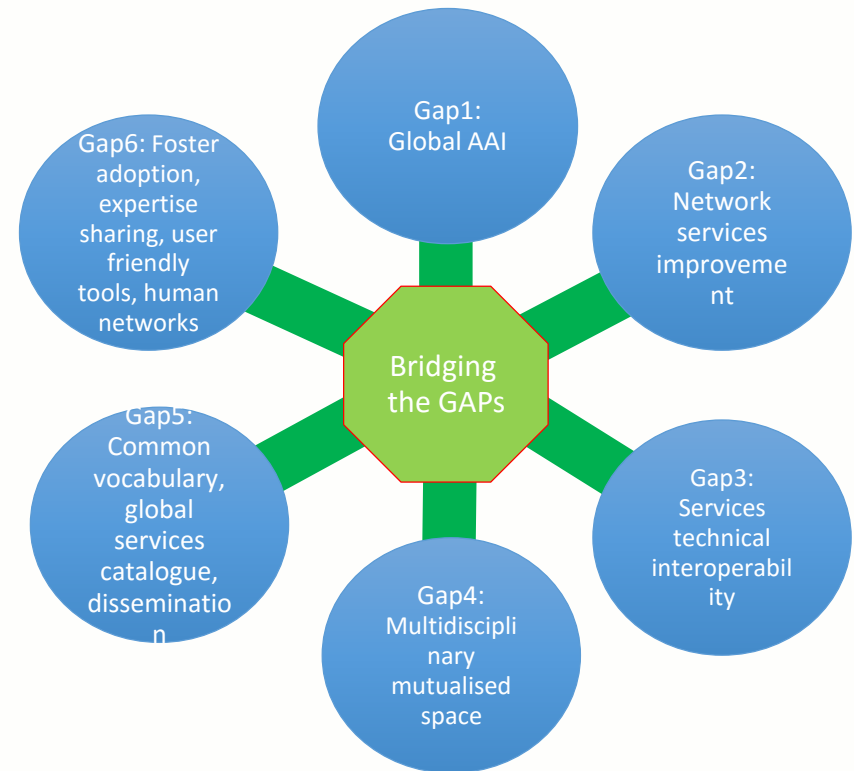
## - Pisa, 9/15 '17





# Interoperability

Propose an architecture, validated technical solutions and best practices for enabling interoperability across multiple federated e-infrastructures, overcoming current gaps expressed by user communities and resource providers.





- EOSC Pilot a work in progress
- under progress for this autumn
  - Architecture framework and service portfolio
  - Data interoperability framework
  - Governance and policy model
- Stakeholder Event: Brussels 28-29 November

Thank you!

**Brian.Matthews@stfc.ac.uk**



**With Slides from:**

Donatella Castelli, **Massimiliano Assante**, **CNR-ISTI**, Italy  
Volker Beckmann, CNRS, France



**Science & Technology**  
Facilities Council