

Open Science 2019

19th-20th March 2019, Berlin

BUILDING AN EOSC IN PRACTICE,
a summary of the 2nd HLEG work



CSIC

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Dr. Isabel Campos Plasencia
CSIC - Spanish National Research Council

Outline

- » **Background** on EOSC
 - The shortest possible version of the history
- » Explaining the **rational behind the recommendations** of the EOSC HLEG:
- » **Minimum Viable Ecosystem**
 - Implications for **Researchers**,
 - ... for **Infrastructure Managers**,
 - ... for **Software developpers**
 - ... for **Policy makers**
- » Some **reflexions** looking towards Implementation

European Commission

Digital Single Market (DSM)



April 2016: European Cloud Initiative COM(2016) 178
as part of the 'Digitizing Industry' package

- **EuroHPC**: European Strategy on High Performance Computing
- Widening access and building trust: **Open Science**
- European Open Science Cloud: **EOSC** as the instrument to support **Open Science**

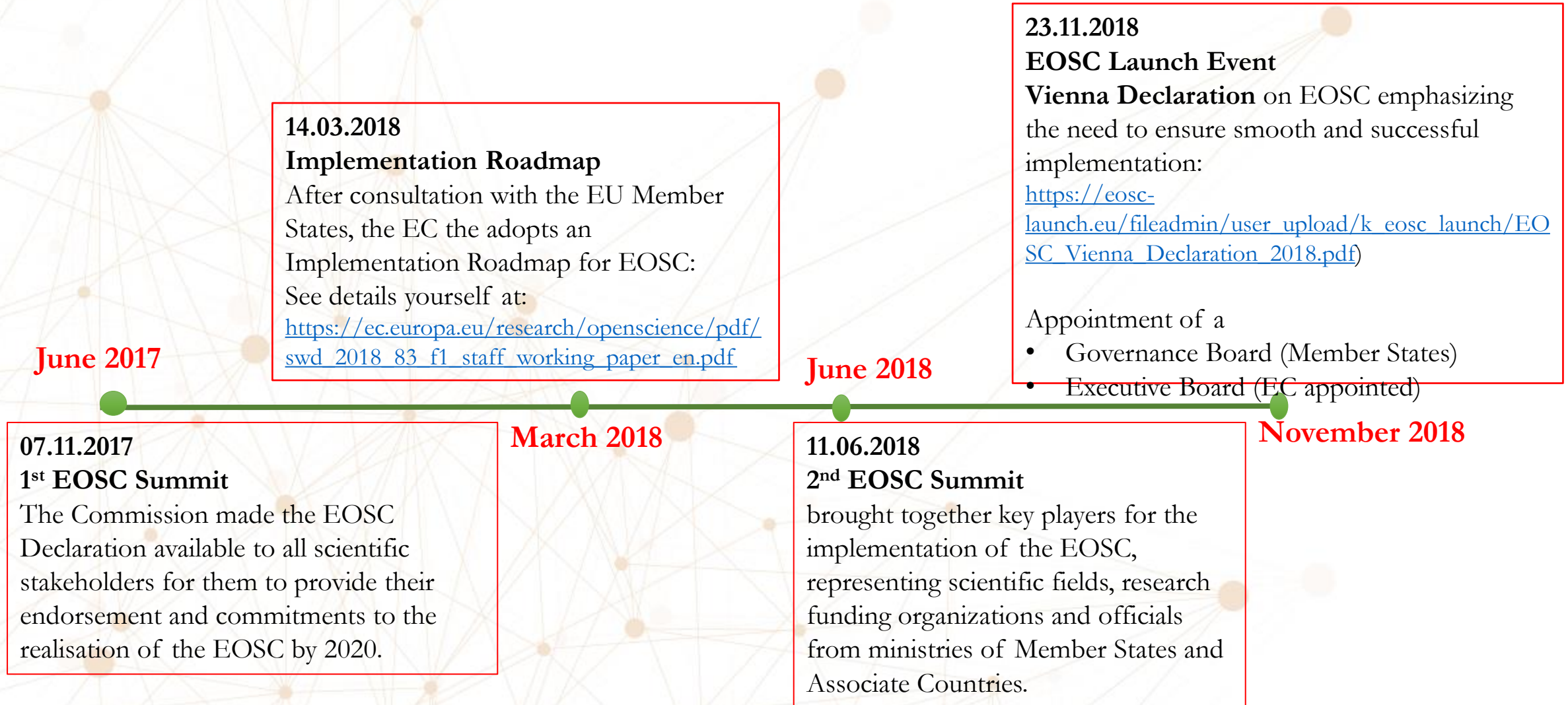
Idea and Realization of an EOSC

The history (The shortest possible version)

April 2016: European Cloud Initiative COM(2016) 178 as part of the 'Digitizing Industry' package

The EOSC aims to give Europe **a global lead in scientific data** infrastructures.... It will offer 1.7 million European researchers and 70 million professional in science and technology a virtual environment with free at the point of use, open and seamless services for storage, management, analysis, and **reuse of research data, across borders and scientific disciplines.**

Source : COM(2016) 178final, p.6



BUILDING AN EOSC IN PRACTICE

EC High Level Expert Groups: 1st and 2nd

Details	Additional Information	Meetings	Subgroups	Statistics	Members
Name: Commission High Level Expert Group European Open Science Cloud (E03353) Closed					
Abbreviation: HLEG EOSC					
Policy Area: Research and Innovation					
Lead DG: RTD - DG Research and Innovation					
Type: Informal, Temporary					
Scope: Limited					
Mission: The Expert Group will advise the Commission regarding the shape of a 'European Open Science Cloud' initiative. Overall, the group should provide strategic advice to the Commission on the strategy for the European Open Science Cloud initiative as part of the Digital Single Market.					
Task: Assist the Commission in the preparation of legislative proposals and policy initiatives					
Contact: RTD-OPEN-SCIENCE@ec.europa.eu					
Publication in RegExp: 16 Sep 2015					
Link to Website: http://ec.europa.eu/programmes/horizon2020/en/experts					
Last updated: 30 Jun 2017					

1st

... to provide strategic advice to the Commission on the strategy for the European Open Science Cloud initiative as part of the Digital Single Market.

June 2016

Details	Additional Information	Meetings	Subgroups	Statistics	Members
Name: Commission High Level Expert Group 'European Open Science Cloud' (E03513) Active					
Abbreviation: HLEG EOSC					
Policy Area: Research and Innovation					
Lead DG: RTD - DG Research and Innovation					
Type: Informal, Temporary					
Scope: Limited					
Mission: The Expert Group will advise the Commission regarding how to achieve the implementation of 'European Open Science Cloud' initiative. Overall, the group should provide methodological advice to the Commission on the fulfilment of the objectives set by European Open Science Cloud initiative as part of the Digital Single Market. Care should be taken that the 'European Open Science Cloud' does not duplicate existing initiatives and that it adds value by integrating existing Commission and Member States' initiatives as well as the outcome of past, ongoing and forthcoming projects.					
Task: Assist the Commission in the preparation of legislative proposals and policy initiatives					
Contact: rtd-open-science@ec.europa.eu					

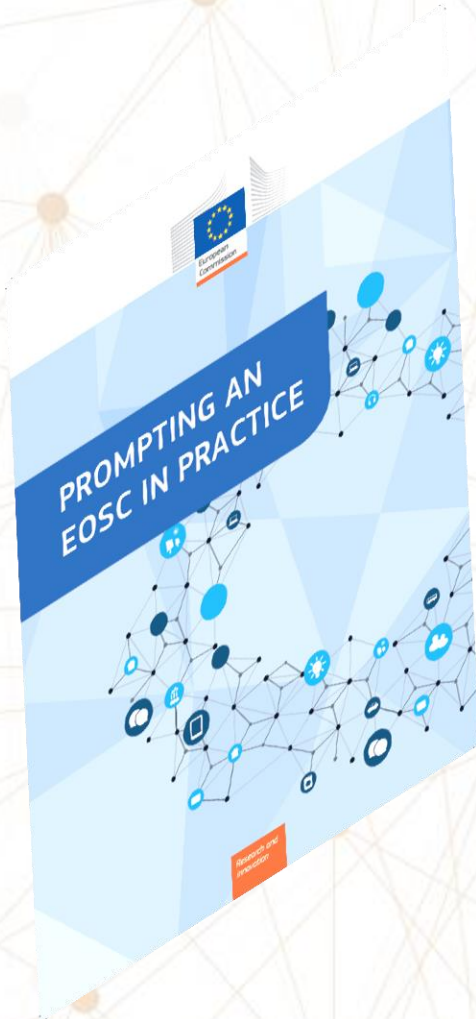
2nd

- Provide methodological advice to the Commission on the fulfilment of the EOSC objectives
- Care should be taken that EOSC **does not duplicate existing initiatives** and, **adds value** by integrating existing initiatives as well as the outcome of past, ongoing and forthcoming projects.

June 2017

December 2018

Key elements in the report



- ✓ **Vision** – The EOSC as interlinking people, data, services & trainings, publications, projects, and organisations across borders and scientific disciplines
- ✓ **MVE** – Discussion on how to make EOSC a Minimum Viable Ecosystem
- ✓ **Business models** – 3 valid alternatives for funding the EOSC have been outlined
- ✓ The **set of practical recommendations**

Have a look yourself:

<https://publications.europa.eu/en/publication-detail/-/publication/5253a1af-ee10-11e8-b690-01aa75ed71a1/language-en/format-PDF/source-80622260>

... **EOSC** can only be **gradually implemented**.

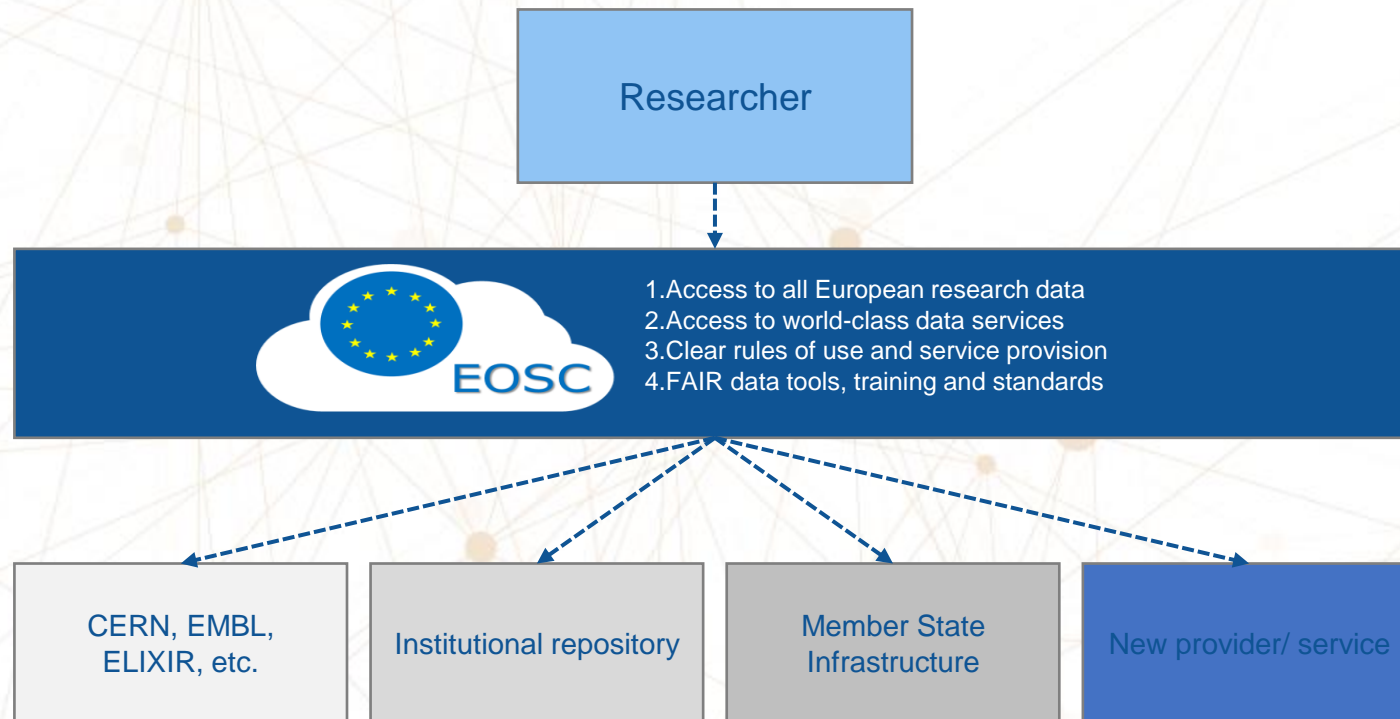
... **EOSC will NOT** provide science with functionalities that are not already provided by large Cloud providers for generic purpose.

However, it has the potential to deploy added value services for science:

- » ... for **providing seamless access** to instruments and the borderless **exchange of data and knowledge** (FAIR) and to produce results beyond the sum of the individual findings.
- » ... for **simplifying funding models** and moving away from the project based funding nightmare.
- » ...to foster **the synergy between scientific domain knowledge and IT domain knowledge**

BUILDING AN EOSC IN PRACTICE

A well-known EC-RTD slide on the EOSC vision



- **Easy access** through a universal access point for ALL European researchers
- Cross-disciplinary access to data and services unleashes potential of **interdisciplinary research**
- Services and data are interoperable (**FAIR data**)
- Data funded with public money is in principle open (as **open as possible, as closed as necessary**)
- EOSC will help **increase recognition data intensive research and data science**

Seamless environment enabling interdisciplinary research

The EOSC will allow for universal access to data and a new level playing field for EU researchers

Let us turn the argument around: “data sharing” being used as an argument/vehicle to solve a more fundamental problem:

- E-science based knowledge remains confined in disciplinary silos.
- By promoting “data sharing”, the hope is expanding E-science knowledge across disciplines that will share data.

Seamless environment enabling interdisciplinary research

Building the EOSC in practice

Starting with a Minimum Value Ecosystem

EOSC as a Minimum Viable Ecosystem

- EOSC aims to put Europe at a global lead regarding scientific data infrastructures.
- The provision of infrastructures, technology development, and human resources to support it will take place in a **very heterogeneous landscape**.
- Addressing this challenge requires the definition of a smallest common denominator: the EOSC **Minimum Viable Ecosystem (MVE)**

The way to do it is through MVP strategy

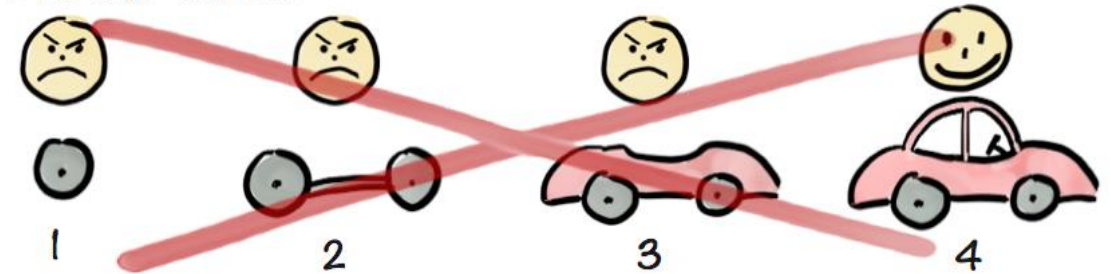
» Minimum Viable Product is

- “a version of a new product which allows a team to collect the maximum amount of validated learnings about customers with the least effort.”

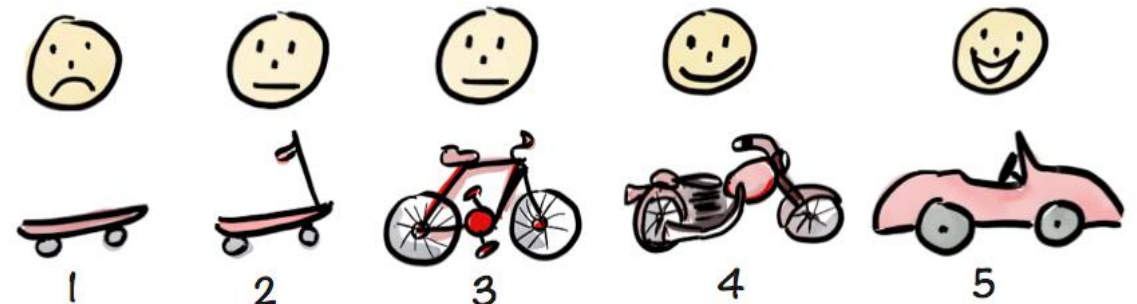
» MVP includes

- “the feature or features required to solve a **core problem** for a set of users and be released to market”.

Not like this....



Like this!



by Henrik Kniberg

- » **Minimum Viable Ecosystem** would enable EOSC to emerge as
 - a collaborative effort,
 - in an iterative way: EOSC is not a project, is an iterative process
- » Identification and Understanding of the needs
- » Finding the opportunities which includes
 - identification of researchers current needs,
 - their actions and desired results
- » Deciding what features to build
 - What are the current blocking factors
 - What will researchers win by removing each blocking factor
 - Establish a roadmap with priorities



Recognise *that the implementation of the European Open Science Cloud is a process, not a project, by its nature iterative and based on constant learning and mutual alignment (EOSC Vienna Declaration, 23.11.2018)*

There are three+one key user groups

» Researchers

- they are the customer and customer is king

» Software developers

- they are the makers of the whole thing
- they make the ecosystem rich or poor

» Infrastructure providers

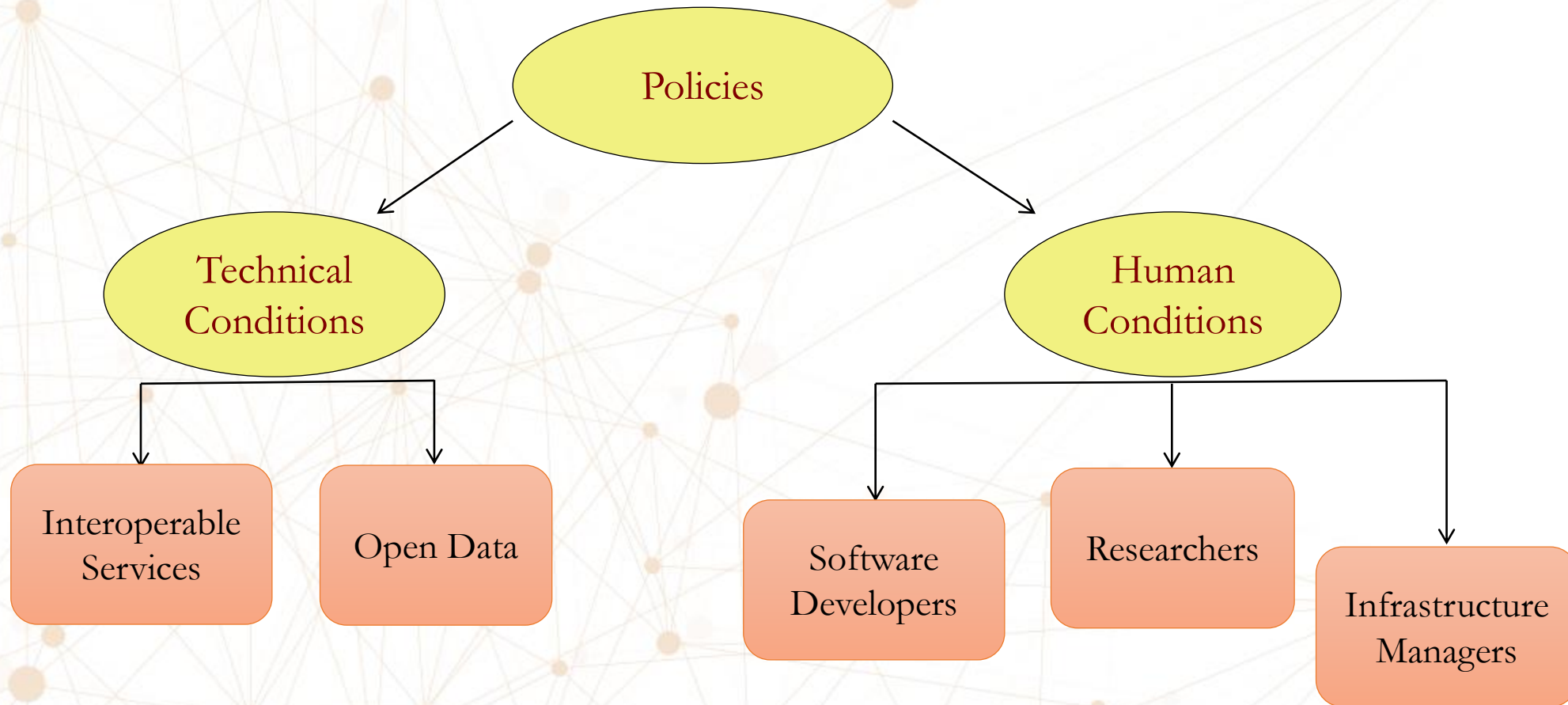
- they get an opportunity for new businesses on their platforms

» Research funding organizations

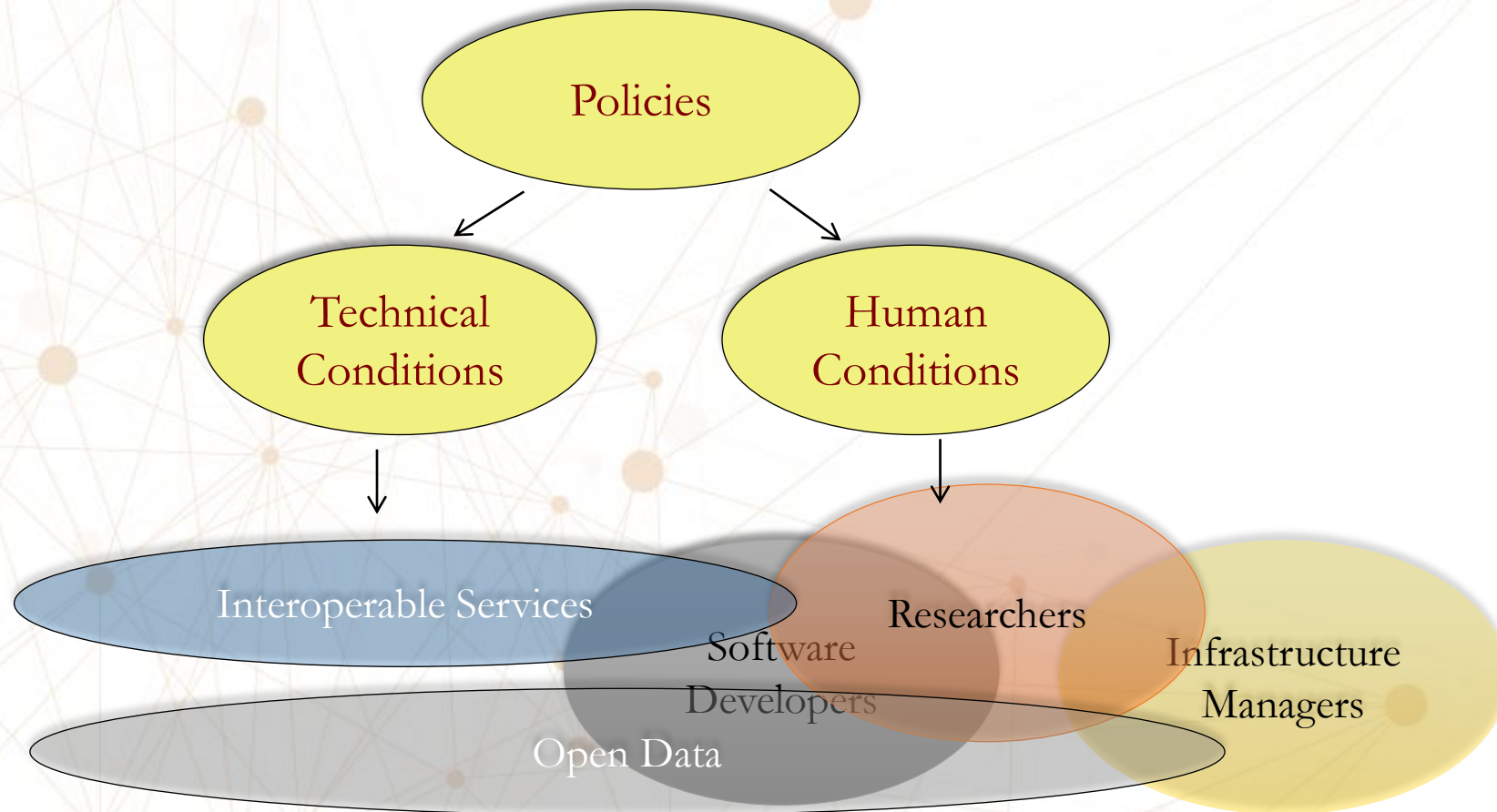
- they foot the bill
- they care for efficiencies of the system



An MVE will emerge if political, technical and human/sociological conditions are met



Reality is more complex/ richer: inter-dependencies



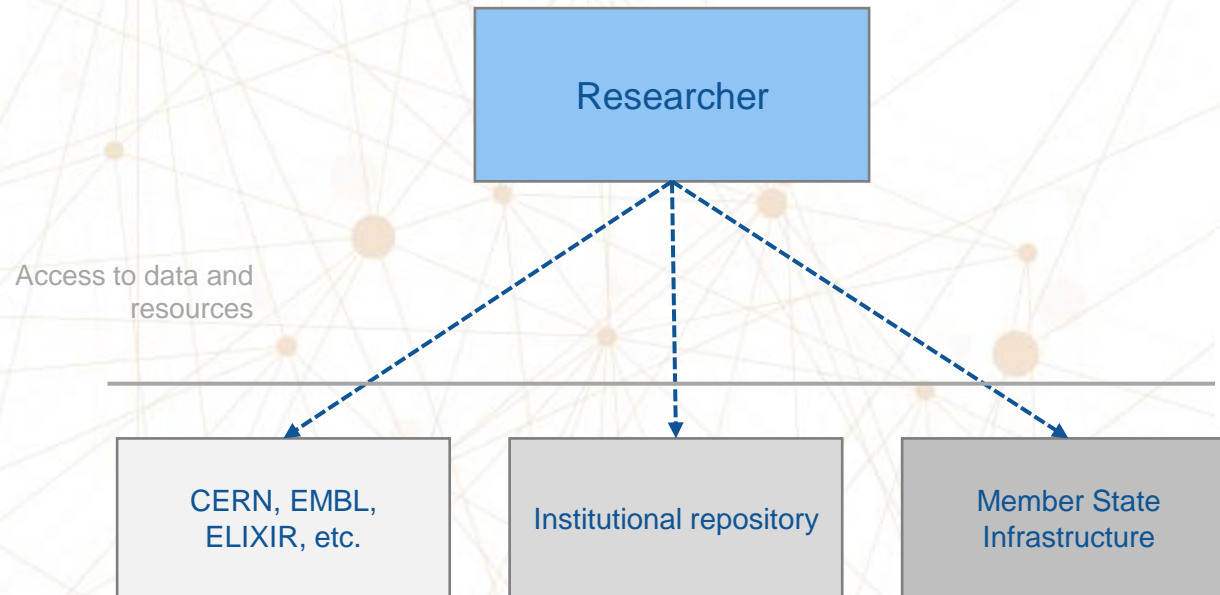
Building the ecosystem for Researchers

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

Call for the European Open Science Cloud *to provide all researchers in Europe with seamless access to an open-by-default, efficient and cross-disciplinary environment for storing, accessing, reusing and processing research data supported by FAIR data principles (***EOSC Vienna Declaration, 23.11.2018***)*

Scientific infrastructures produce large amounts data (PBs...)

European researchers face data fragmentation and unequal access to quality information sets



- Fragmented access (across scientific domains, countries and governance models; varying access policies)
- Limited cross-disciplinary access to data sets (i.e. interdisciplinary research)
- Non-interoperable services and data
- Closed data

Limited and limiting access for an ordinary European researcher

Scientific infrastructures produce large amounts data (PBs...); **Factual**

Let us answer out of experience

- *Fragmented access (across scientific domains, countries and governance models; varying access policies)?*
 - Large Collaborations have a strategy to solve this problem for themselves: dedicated funding lines
 - For the average research group, there are nightmare scenarios.
- *Limited cross-disciplinary access to data sets (i.e. interdisciplinary research): Factual*
 - Is it needed? Sometimes it is.
 - Difficult to overcome, as today's scientific challenges and career evaluation policies impose a huge specialization on researchers.
- *Non-interoperable services and data: Factual*
- *Closed data: Plenty, and will continue so (embargo periods, lack of actual time, lack of demand,...)*

Limited and limiting access for an ordinary European researcher

Bringing the key players (i.e. Researchers) on board:

Fundamentally wrong ansatz

~~There seems to be a lot of politics in the
European Science funding~~



~~Politicians and Committees
define what Science should deliver~~

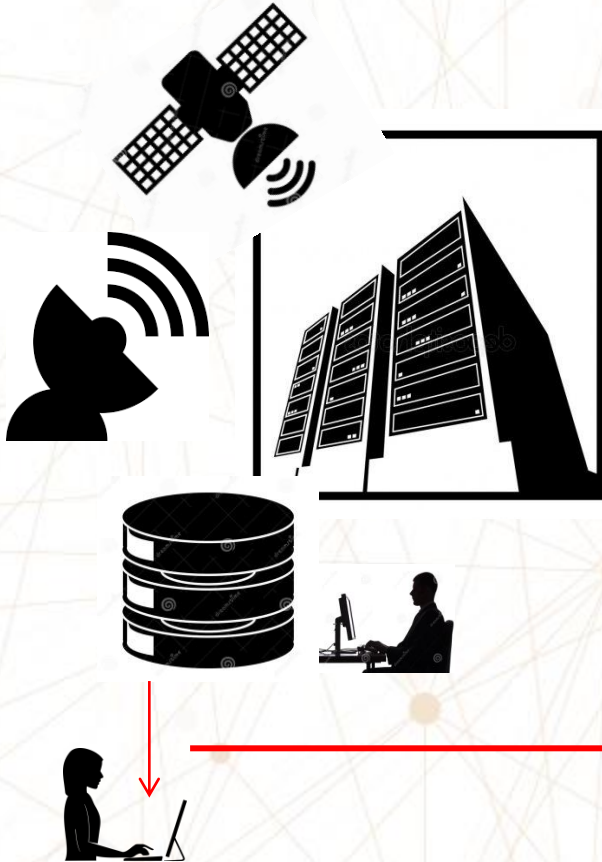
Key Players



Good Science & Technology is still done
by individuals & small teams, who are
enthusiastic about a particular subject,
technology or scientific idea.

BUILDING AN EOSC IN PRACTICE

Insufficient access to computing, processing & storage resources



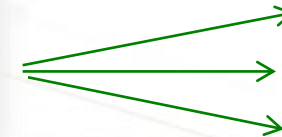
✓ **Researchers can barely cope with the post-processing** of the data generated during regular work in HPCs or Observation facilities on daily work

✓ **Problems to store** it permanently locally for later analysis, let alone, **Data Preservation**

✓ Analysis and re-analysis requires **flexible access to processing** and **data sharing** facilities for daily work



Your disk is almost full
Save space by optimizing storage.



Sharing?



No real access to Fast Networks

Despite **GEANT** theoretical capacity, most (90%) researchers in Europe have no access to Gigabit Speed Networks,

→ **Last mile problems** are plaguing Universities and research performing institutions

→ **Lack of availability of fast protocols**

TCP/IP keeps being the limiting factor

UDP: commercial solutions (\$\$\$)

`scp -- > 10Mb/s`

GridFTP → 150 Mb/s (TCP/IP enhanced with some UDP)

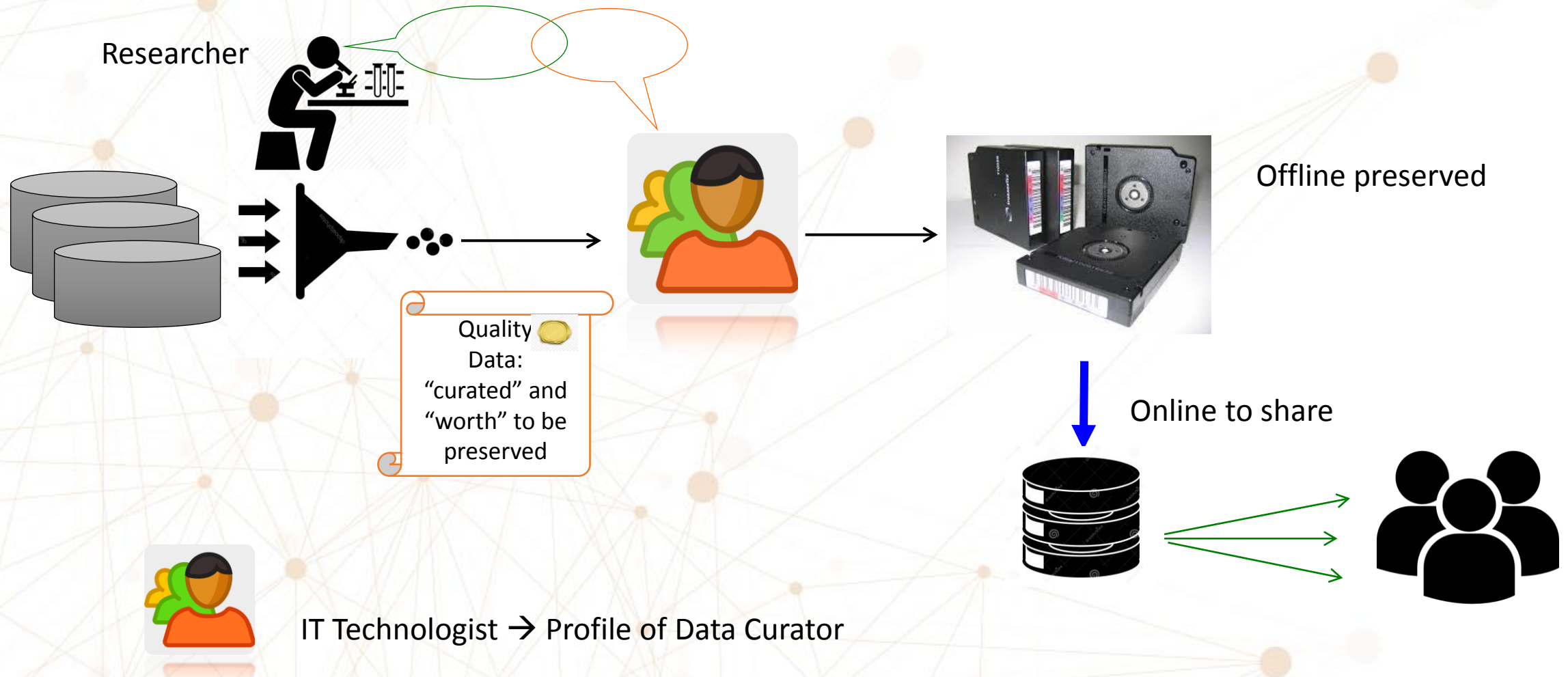
But very limited availability (**essentially HEP-related centers**).

Example of knowledge silo



BUILDING AN EOSC IN PRACTICE

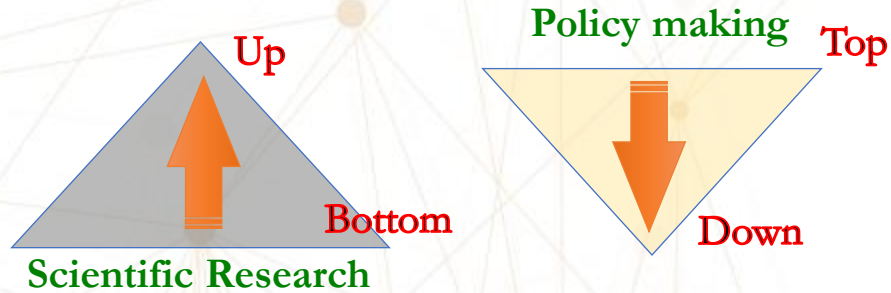
Research Data worth to be preserved and/or shared



IT Technologist → Profile of Data Curator

Agreement that this is a helpful career profile in Research Performing organizations

BUILDING AN EOSC IN PRACTICE



**There is a gap to close.
Recommendations HLEG:**

1- Improved Fast Network access to EOSC data through policies for better harmonisation of research networks with particular attention to “last mile” – researchers have to be able to get to the data

2- Ensure that researchers who cooperate to lodge curated data in trusted repositories are recognised in funding and promotion schemes

3- Ensure policy initiatives focused on strengthening research infrastructure provider participation in EOSC – both at EU level and member state level

BUILDING AN EOSC IN PRACTICE

- Ensure that researchers who cooperate to lodge curated data in trusted repositories are recognised in funding and promotion schemes



- ✓ Recognition is necessary, but **all the required IT work cannot rest on the shoulders of researchers.**
 - ✓ **Researchers** are responsible **for assessing the value of data**, taking into account the specificities
- ✓ Research projects cannot take care of sharing and preserving research data (*): **researchers have no time**
 - ✓ Contributing to the solution of real problems today in Science is an enormous endeavour.
 - ✓ Once they have explored a path, they need to move on quickly to the next step.
 - ✓ Specially so for young researchers, due to the very high competition for research jobs: dedicating time to these matters would ruin their career.
- ✓ **Keeping research data available to be shared and curated requires a layer of technical support:**
 - ✓ Profile: people not necessarily understanding scientific details, but only the global meaning of the data.
 - ✓ Interesting career path to educate highly skilled professionals: **Data Management Experts**,

(*) With the **exception** of very Large Collaborations (LHC, ESO,...) that have already that support onboard

BUILDING AN EOSC IN PRACTICE

3- Ensure policy initiatives focused on strengthening research infrastructure provider participation in EOSC – both at EU level and member state level



Your disk is almost full
Save space by optimizing storage.

A very good “**side effect**” of **EOSC** would be a global substantial(*) **increase of the resources available** to research groups.

(*) Substantial = at least 2 orders of magnitude with respect to today’s situation in the upcoming 2 years

✓ **Investments need to be made at the level of national organizations** to support Science

- We could not do efficient data distribution and analysis without the Grid (**EGI**),
- **PRACE** has allowed tackling problems using HPC, that we could not have even address before

Successful

For many leading research groups, the resource gap is at the level of supporting daily work on exploratory projects.

Infrastructure Managers

Recall that the Council - in its conclusions of 29 May 2018 - welcomed the implementation roadmap and the federated model *for the European Open Science Cloud*.

Resolve *to harness the many ongoing and planned activities at EU and Member States level to cooperate in establishing an inclusive partnership with a view to developing the European Open Science Cloud as a federated infrastructure that can enhance value-based, open, trusted, user-centric digital services across borders within the Digital Single Market (DSM) (**EOSC Vienna Declaration, 23.11.2018**)*

BUILDING AN EOSC IN PRACTICE

From a technical standpoint:
We know how to federate infrastructures

Infrastructure building Capacity: Community funded and/or National Infrastructures



Building Capabilities via User Community Engagement



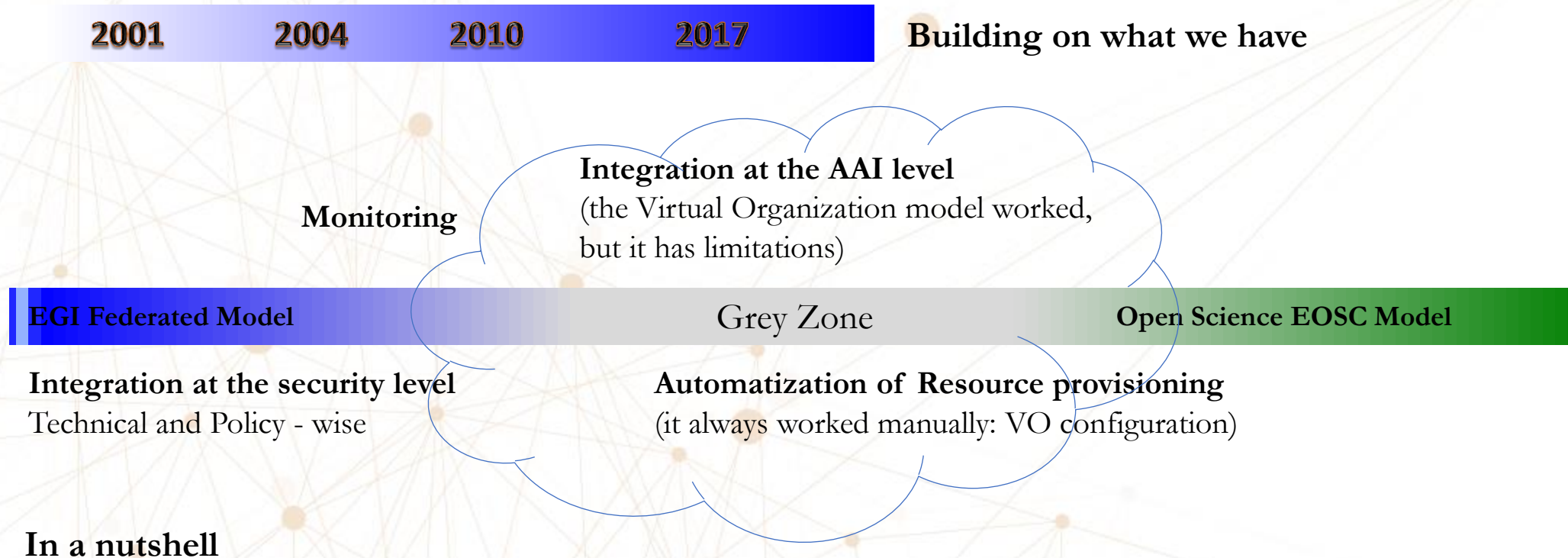
Technical Standpoint in Federating Infrastructures
and it works (see accounting.egi.eu) :



- **2018 Figures**
 - **4.4 Billion CPU core wall time** delivered in **2018** = worth **220 M€** in Amazon
 - 970,600 computing cores
 - 356 PB disk & 380 PB tape storage
 - **1170 Open Access publications**
 - +41 new international projects
 - 31 large scale ESFRI projects/landmarks supported

BUILDING AN EOSC IN PRACTICE

Technical Standpoint in Federating Infrastructures: **Making the leap to EOSC**



- Policy Level: Actions to be supported by technical developments, notably in the front of Security (data privacy policies) and access policies
- Technical Integration Level: work to be done/extended, building on the previous experience.

BUILDING AN EOSC IN PRACTICE

Incentives for Research Infrastructure Managers *“Generating Economic Growth via added value”*

- » Integration resources in the EOSC is a way to achieve a higher a more efficient usage.
- » Supporting International Research Collaborations
- » Prototyping and piloting Innovative Services for cutting-edge Research

Expectations of gaining certain
Economy of Scale

Expanding the user base in
National infrastructures

Early Career Professionals Education

Independent IT Research at the State
of the art

High Quality economic growth

Policy Recommendations to Incentivize Infrastructure Managers

- Discuss the practical elements of the business model with Member States

- Incentivize Infrastructure managers through funding streams & maintenance costs for shared facilities and data repositories. This implies technicians to support data management.

Discuss the practical elements of the business model with Member States

Excellence Driven Acces

Exclusively dependent on the scientific excellence, evaluated through peer review conducted by experts

Market Driven Acces

Agreement between the User and the e-Infrastructure that will lead to a “payment” for the Access

Private providers will envision a fee-model for the Access

Wide Acces

Guarantees the broadest possible Access to scientific data and digital services to Users wherever they are based

Discuss the practical elements of the business model with Member States

Excellence Driven Acces

Exclusively dependent on the scientific excellence, evaluated through peer review conducted by experts

In practical terms

A model based on **Wide Access mode, modulated by a negotiated agreeable Access restriction**, is the pragmatic way to start moving with EOSC

Wide Acces

Guarantees the broadest possible Access to scientific data and digital services to Users wherever they are based

Some of the Financing Instruments analyzed

Direct funding

Provide resources to develop innovative services when they are not economically profitable in commercial providers

Capacity mismatchings and access problems for outside stakeholder

Cloud vouchers ("coins")

Easy way for Commercial providers to access the Marketplace

Poor model to maintain data sets in view of the money-flow gaps in Research funding, and also data curation/preservation lifetimes.

Hybrid approaches

Supports access by both commercial and non-commercial organizations into market, allowing both groups to work in areas where they have specialized abilities.

EOSC as ecosystem for Software Developers

Highlighting that Europe is well placed to take a global leadership position in the development and application of cloud services for Science.

Reaffirm the potential of the European Open Science Cloud to enable first-class data-driven science and to stimulate new business models benefiting our society and the economy.

Recognise that such services will create opportunities for both public and private sectors, notably by intensifying reuse of public sector information while preserving data integrity, and ensuring access, transparency within and across borders. (EOSC Vienna Declaration, 23.11.2018)

Incentives for Software Developers: “*Engaging Human Talent*”

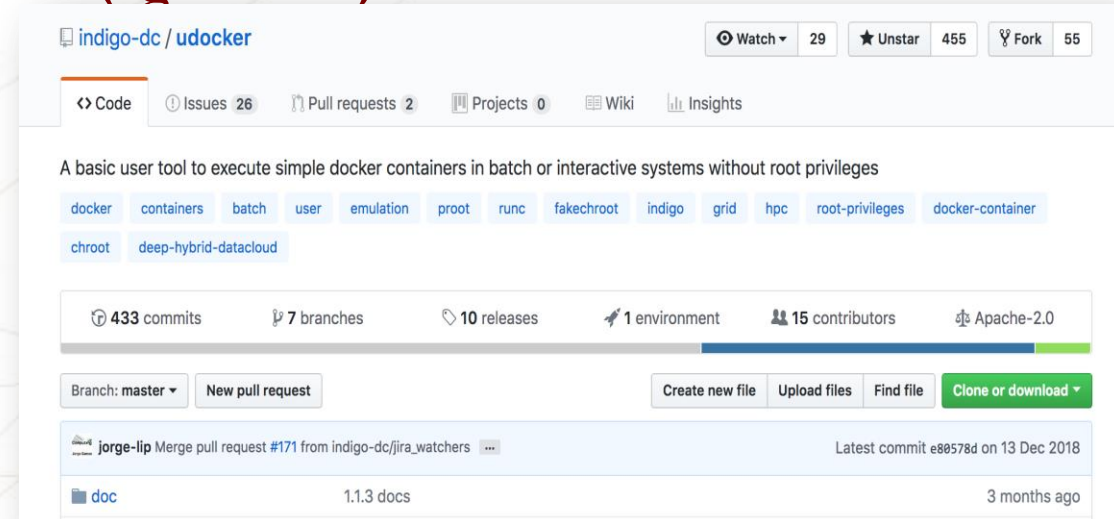
Breakthrough ideas leading to innovation need
to be awarded with the proper Recognition

» Indicators of recognition in popular software repositories (eg. Github)

- # of forks
- # of stars awarded by peers and users
- # of downloads

» Additional Indicators through EOSC

- # of services built using a software component
- # of deployments on Resource Centers
- # of users/scientific communities served



Incentives for Software Developers: “*Engaging Human Talent*”

» “***EOSC-Ready***” as a **Branding** for software products, will harness the potential of the European developers, both in academia and industry.

» *What does it mean “EOSC-Ready”? Trustable*

- **Open Source** → contributed upstream (potentially sustainable)
- Based on **Open Standards**
- Adheres to code style quality standards → **Software Quality**
- It **respects the EC directives** on cloud security (**NIS**), personal data protection (**GDPR**) and Free Flow of non-personal Data (**FFD**).



Incentives for Software Developers: “*Engaging Human Talent*”

“EOSC-Ready” as a branding for software products, will harness the potential of the European developers, both in academia and industry.

The Devil might be in the details: **how & who would be awarding such label?**

Nothing is black or white, **there is a grey scale:**

- » Is Closed Source software not good to be used to deploy services in EOSC?
 - Certainly not for AAI.
 - Nor for core services that support the federation layer (IMHO)
- » Open Standards: difficult compromises always among which standard...
- » Software Quality procedures: need to be substantiated in an agreeable way (building on community recognized best practices)
- » EU Directives: Is the EC Market too regulated? How will this affect EOSC? (interesting debate)

Recommendations to Incentivize Software Developers

- **Stronger Open policies for trusted software and associated data services**

Trusted services rely on the existence of a software ecosystem adhering to the principles of software openness, which builds on open standards.

The recommendation is the promotion of an open software ecosystem, supported by a consistent, community-recognized, software certification*.

(*) i.e. NOT an ISO-like certification

- » **EOSC** is understood by policy makers as **the instrument to support** the deployment of services towards a more open environment in science and technology.
- » **EOSC should become a framework to enable cooperation**
- » Now **is time for implementation** at the two levels:
 - **National implementations** → **should translate in more resource availability**
 - **Harmonization at the EU level:** both at the **technical** and **policy** levels.
- » The **EC investment** to “harmonize” is being very generous(*) **order(400M€)**
 - Ongoing EOSC-supporting projects is currently order of (100M€).
 - Recently awarded 157M€ more, and order (100M€) still to come.

(*) *IMHO: specially when compared with other research programs:*

250M€ → 1350 Marie Curie Individual Fellowships to the most talented postdocs in Europe for 2 years.

BUILDING AN EOSC IN PRACTICE

Call INFRAEOSC-04-2018-2019 **Four Multi-ESFRI clusters funded**

Connecting ESFRI infrastructures through Cluster projects (95M€)

Call INFRAEOSC-01-2018-2019 **OCRE**

Access to Commercial Services through the EOSC-hub 12M€

The political & financial support is there,
now should be “EOSC Implementation Time”

Call INFRAEOSC-05-2018-2019

(a) *Setup of an EOSC coordination structure (10M€)* **EOSC-Secretariat** (EOSC-Pilot was **another 10M€**)

(c) *FAIR data uptake (10M€)* **FAIRsFAIR**

(b) *Coordination of EOSC-relevant National Initiatives and support to prospective EOSC service providers (30M€)*

- **EOSC-synergy,**
- **EOSC-Pillar,**
- **NI4OS-Europe,**
- **EOSC-Nordic,**

+ Thematic on Photon Science **Expands**



Isabel Campos Open Science 2019 / Berlin

A background network diagram consisting of numerous orange circular nodes of varying sizes connected by thin, light gray lines, creating a complex web-like structure across the slide.

Thanks for listening!
Questions?

IBERGRID 2019

23rd – 27th, September 2019



**In parallel: KoM of the Iberian-led
EOSC National implementation project**

