

EOSC emerges from the Clouds

how a misty metaphor delayed practical action

(but also how the sky's bright again today)

PREAMBLE:

Caveat emptor. Read the disclaimer

These are *purely personal* observations,
made over the past ~10 years.

*

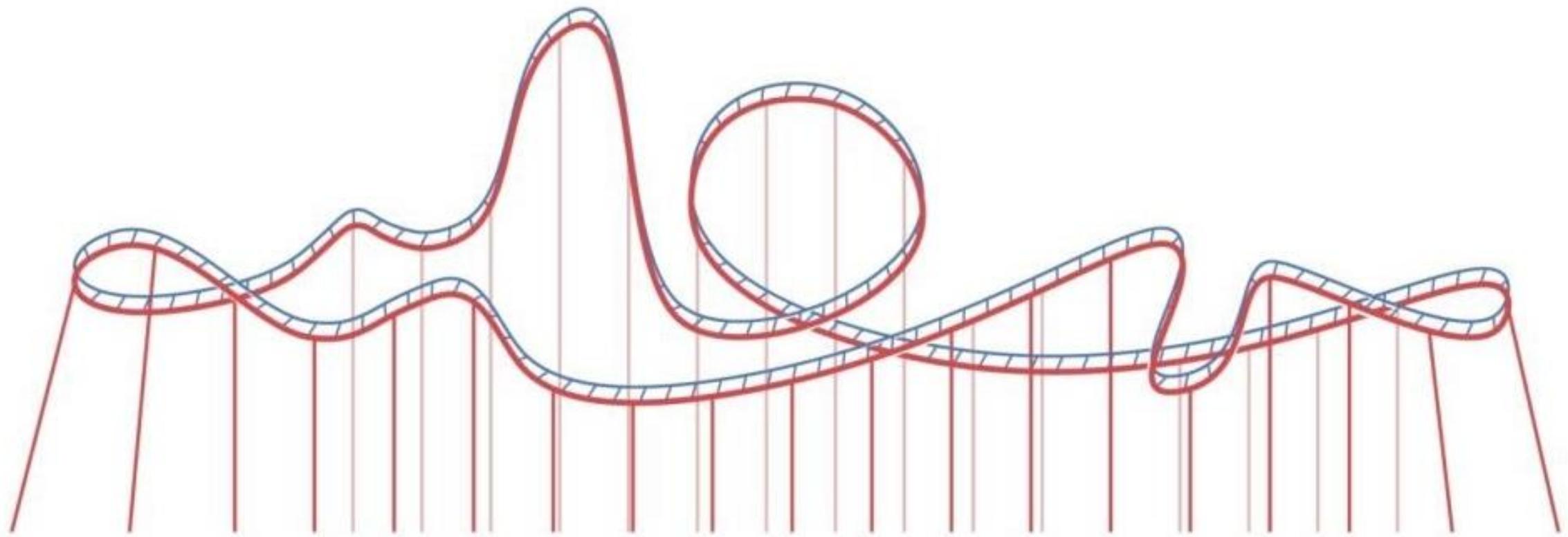
I won't claim I'm definitely right
– I *can* say I've been around for a while though.

Different people will see this differently.
People more important than me are probably more correct ;)

* About 20 years. Most of that time spent at NRENs and in science outreach / eInfra development. Doesn't stop me from getting it wrong.

And I know my own
perspective gets a
little *cloudy* from
time to time ;)





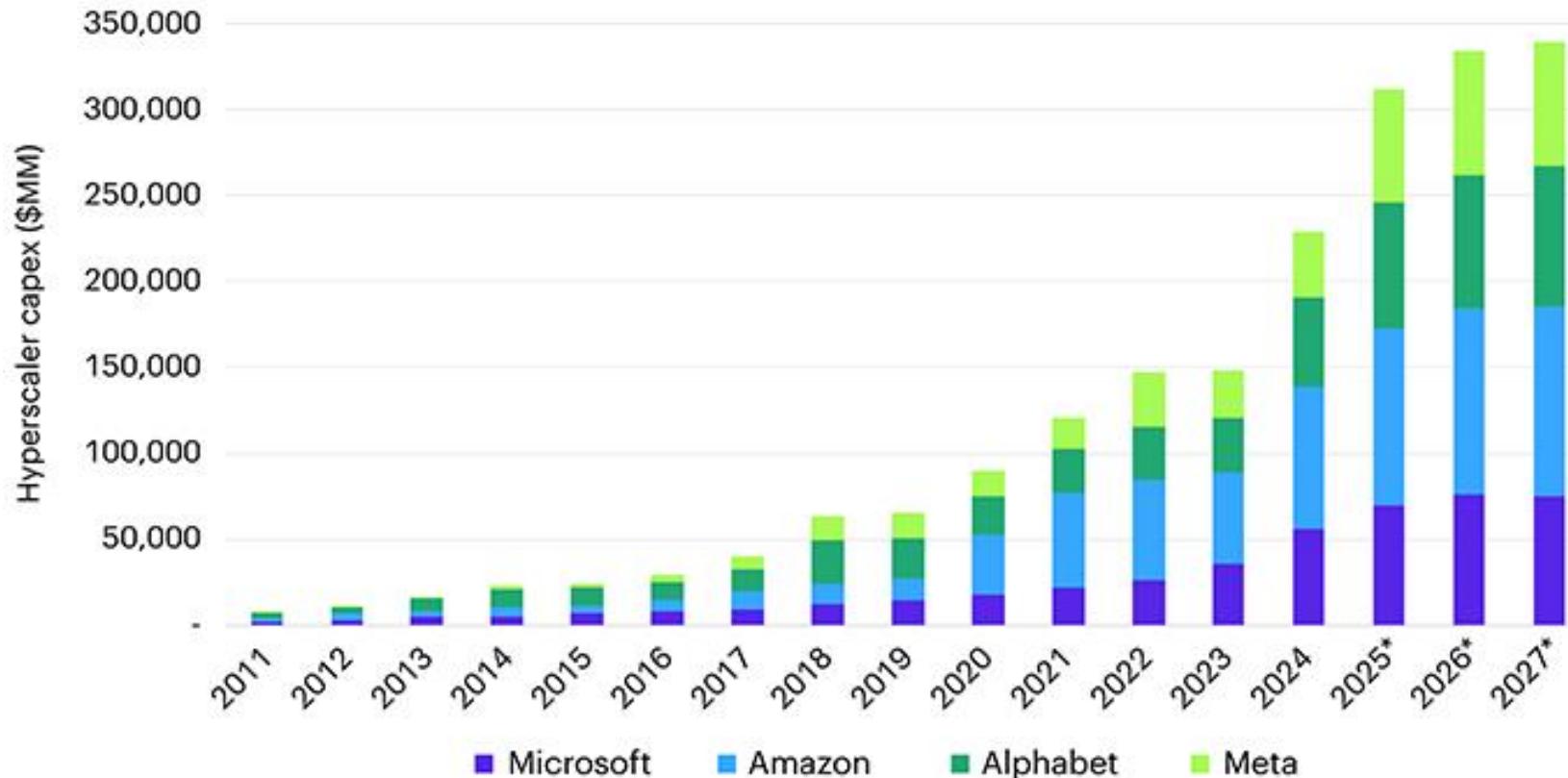
A SHORT HISTORY OF EOSC

How did we get here and what can you learn from our journey?

FIRST CHAPTER.

EOSC, 2015-2020

HYPERSCALERS' CAPEX CLIMBED IN 2024



a) Boom! Hyperscalers happen. The EC realises US hyperscalers are moving into the R&E data ecosystem. Circa 2015

EOSC, from 2015-2020

b) Idea generation starts. How should EC counter this?

b1) perhaps build a central European R&E answer to the hyperscalers?

---> but that's a huge task.

and dangerous, too (lawsuits). And out of the comfort zone (operations... hardware...)

b2) so perhaps... *don't* begin at the bottom? and *don't* call it infrastructure?

Much less scary to begin at the top (of the OSI stack) and call it a **cloud**.

Thus begins the EOSC of high level concepts:

“free movement of data” and “FAIR”. Circa 2017

INTERLUDE. What *is* a cloud?

“somebody else’s computer”

It’s the things all your stakeholders *need* done...
...but nobody *wants* to do.

So it becomes a bucket of hot hopes –
“if only I could get other people to do my work”...

Which doesn’t encourage ownership or a sense of investment in assets;
it’s closer to selling airline tickets. Nobody wants to *maintain* the plane.

By way of
comparison

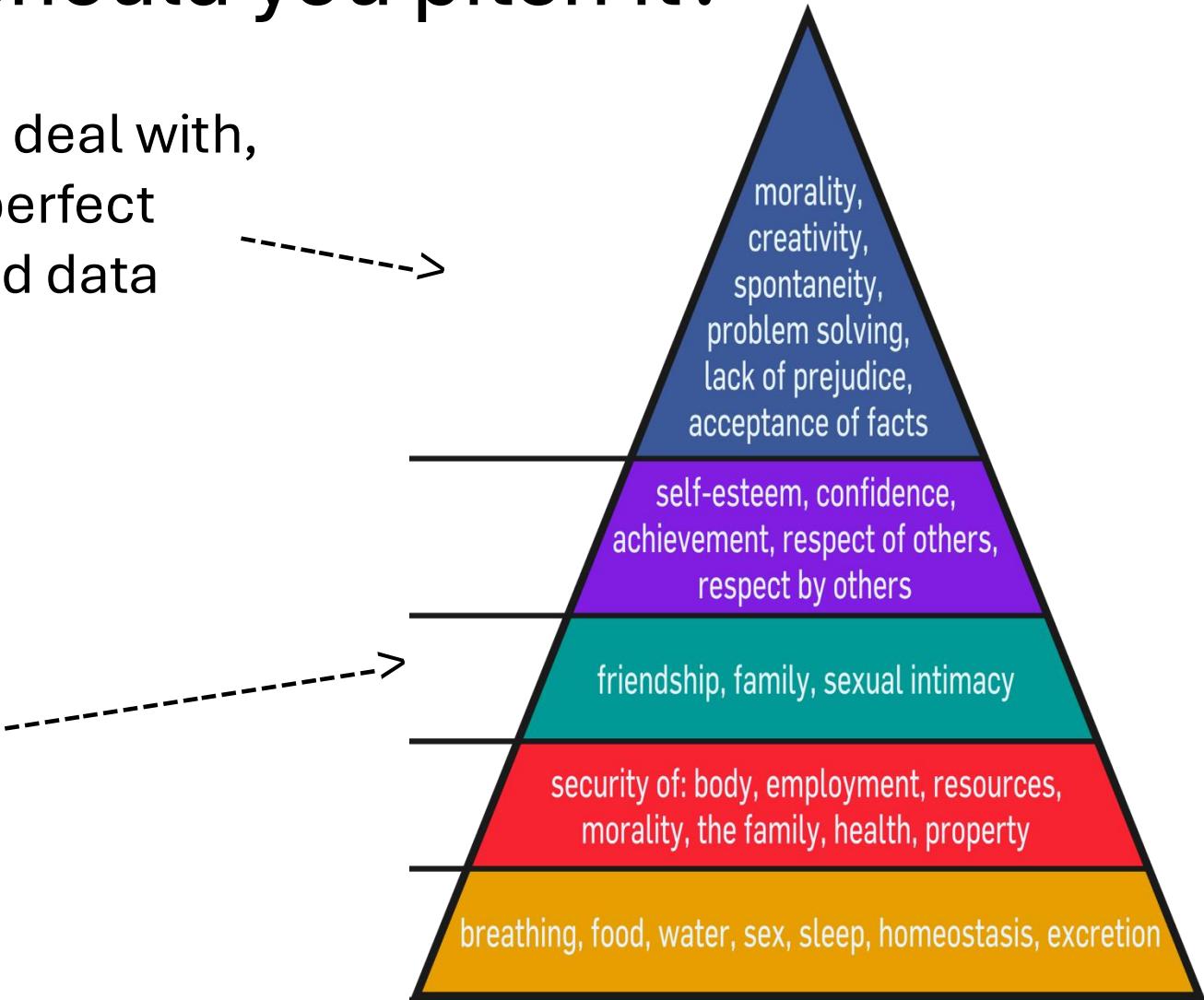
Aviation companies
don't build “flights”.
They *market* flights.

They *build* airline
infrastructure.

Cloud, elnfra, whatever it is... where should you pitch it?

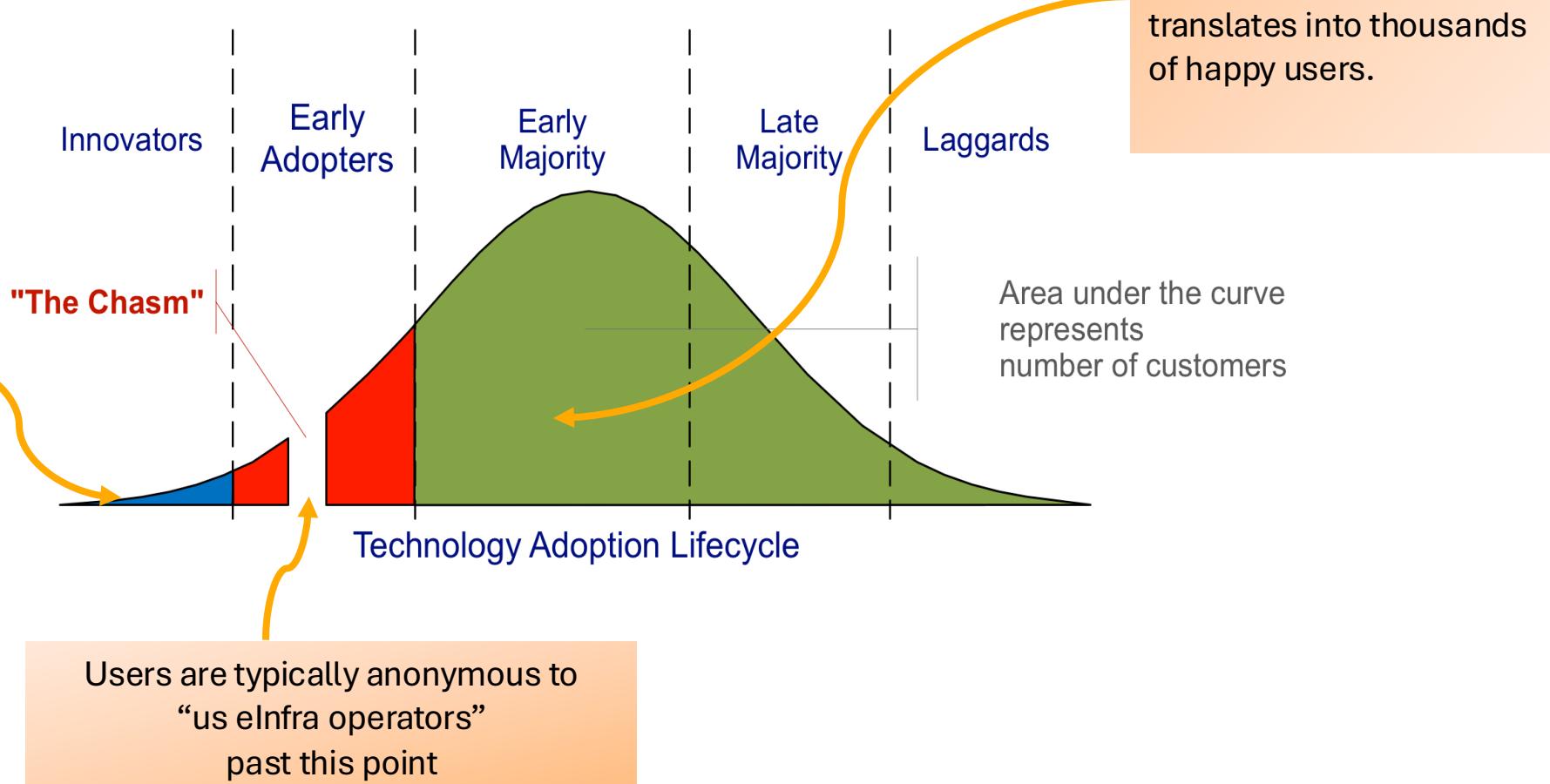
We, and the science counterparts we deal with, think (or pretend) all users are here (perfect understanding of sizes and tools, good data hygiene, identified collaborators, a global namespace, metadata etc)

But inevitably, some (imho the majority...) are here (ad hoc data bottlenecks, ad-hoc collaborations, low training levels, need a quick fix). High barriers and daunting interfaces increase mistakes, increase FUD, lower training retention.



science clouds are like any business. Critical mass matters.

Vocal power users are here – big, but few.
Keywords: FAIR, ontologies, grids, HPC, knowledge graphs.
This area is *relatively* well funded



Okay. Moving on,

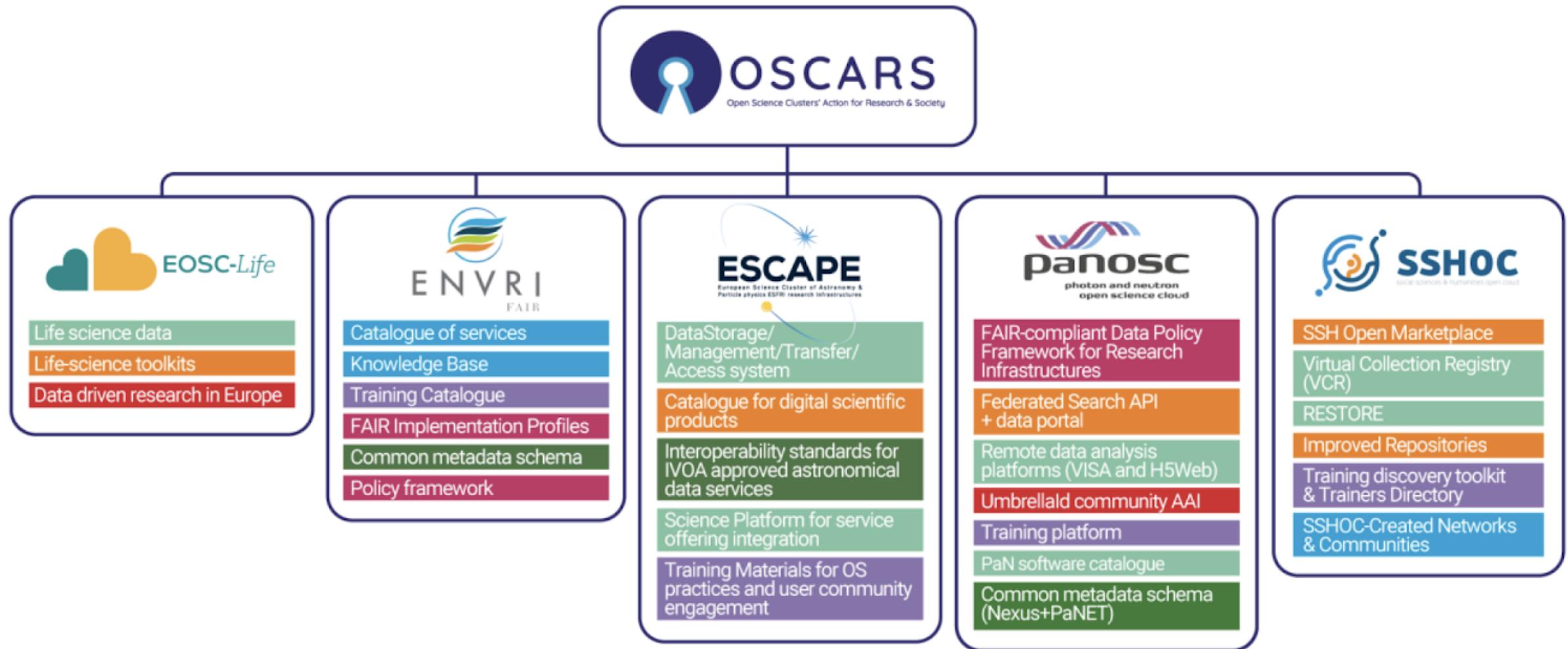
EOSC, from 2015-2020

- c) Leave the building of EOSC to grants
 - c1) Do *not* establish a tech architecture; that requires making choices which we've just conveniently pushed away into those grants.

This is the EOSC of circa 2017-2021: fund already-ongoing projects in what's broadly the “research data space” and call them EOSC; don't touch any technical decisions and leave it to business-as-usual in the existing Science infrastructures to set the direction, or rather *directions* of development

Example of a funded high-level EOSC project:

this is from 2024! – meaning, the effects of architectural laissez-faire will continue for a while



RESULTS CATEGORIES

- Technical Harmonisation
- Policy Harmonisation
- Discovery/Access Platform
- Virtual Research Environment (VRE)
- Training Resource
- Knowledge Centre
- Authentication and Authorization Infrastructure (AAI)

EOSC, from 2015-2020

- d) Having outsourced tech decision making to the R&E ecosystem, you can now focus on doing what you know how to do: Governance. Financing. Legal structure.

Remarkably large amounts of effort have been poured into this activity, while arguably it was relatively unclear what was being governed.

“ IF THE ONLY TOOL YOU HAVE IS A HAMMER,
IT IS TEMPTING TO TREAT EVERYTHING
AS IF IT WERE A NAIL ”

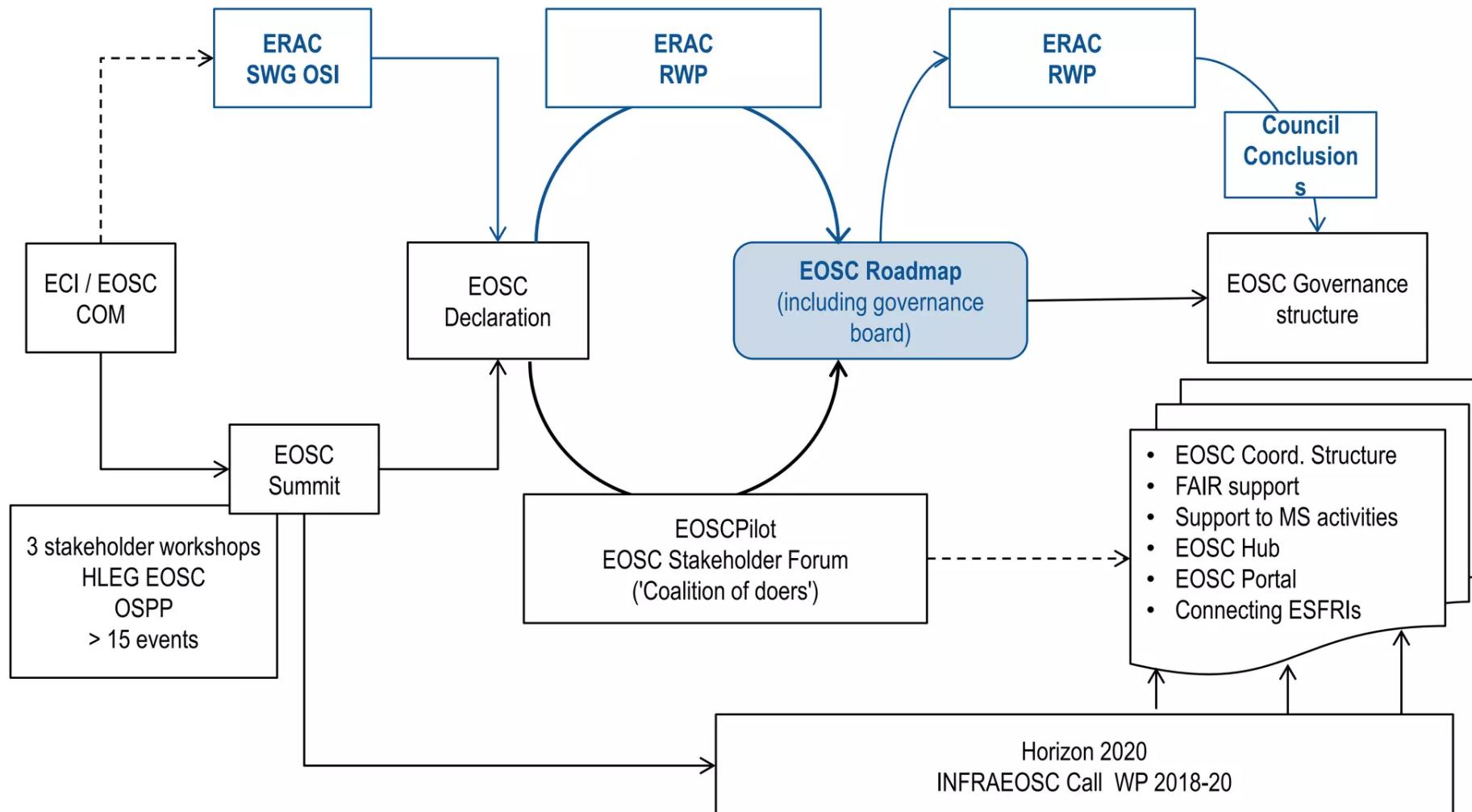


EOSC implementation process



and sure
enough...

wonderful
governance
emerged



2017

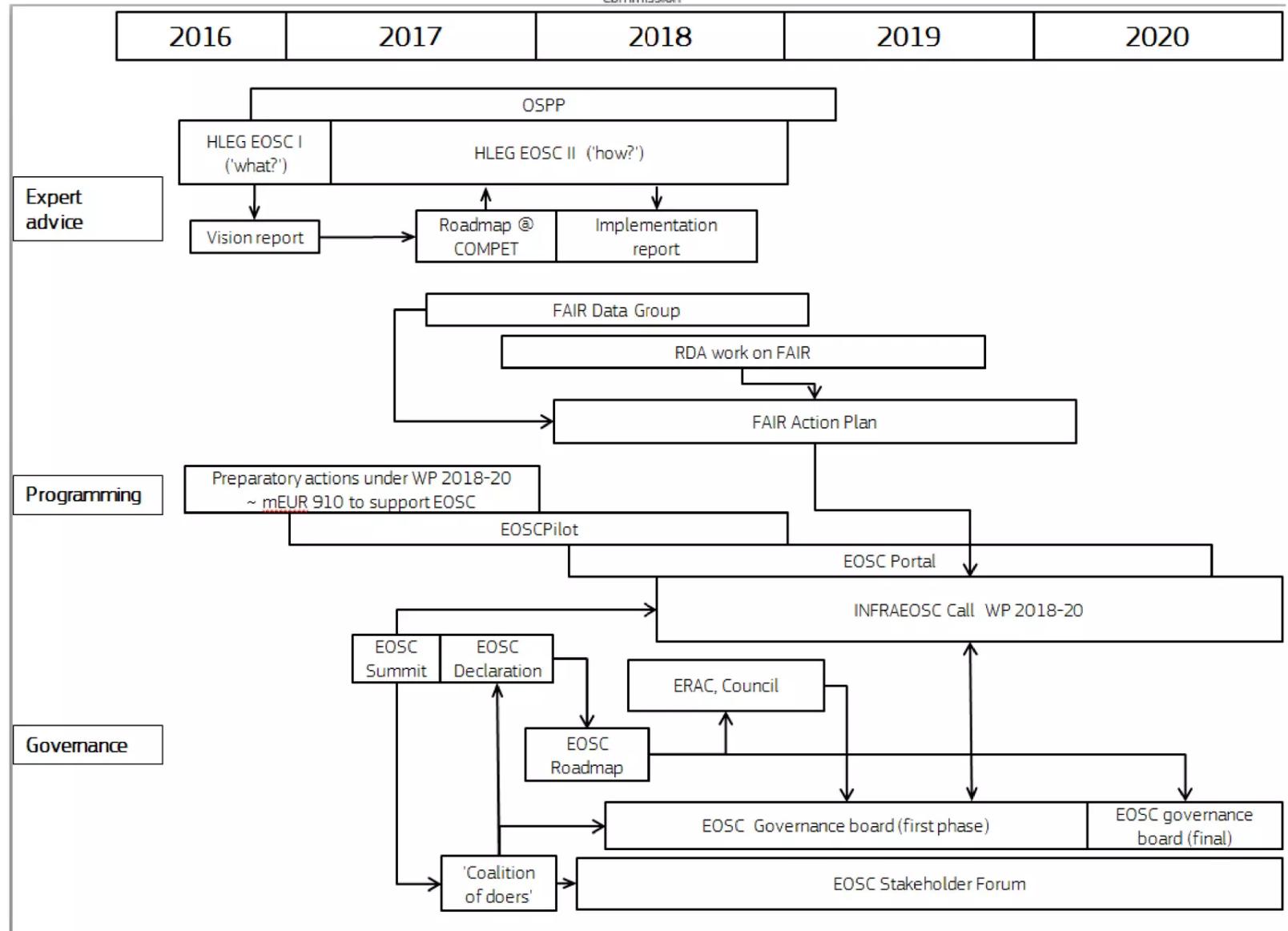
2018

2019-2020

EOSC Timeline



**splendid
roadmap
documents
were written**



**And this was the
architecture
they developed
alongside.**

...that's right: nothing.



casting the challenge as a cloud

The effects were tactically convenient: not talking about infrastructure saved the emperor's new clothes.

Strategically however we ended up with a smörgåsbord of *incompatible professorware*. Little was delivered at scale, and no synergies or re-use of underpinning infrastructure was achieved.

And meanwhile the trickle of data to Grey IT continued.

SECOND CHAPTER. EOSC, 2020-2025

All that time, us NREN operators were looking at this spectacle with wide eyes and open mouths.

- The EC said they wanted to build a cloud
- We were sitting on cloud-like resources (storage and compute facilities with web interfaces and hundreds of thousands of users in total)
- But we failed to get the EC to notice us.

CS3 Site Reports, 2019:

- **25** EFSS sites
- **400k** users
- **2.7 billion** files and directories
- **11.5 PB** storage
- Yearly storage increase: **50-250%**

Information provided on voluntary basis by CS3 site admins (January 2019)

rainy days on cloud-9, circa 2019

Meanwhile at the EC, it began to dawn that the central, high-level EOSC concept was failing to grow roots.

- hard to achieve synergies without a common architecture
- hard to scale using purely central funding (need to involve MS)

“**cloud**” was heard less, and instead the word “**federation**” began to be used, and “node” as the blocks to be federated.
Talk of “**services**” (which up to then had been seen as out of scope)



What is EOSC

“

A process

- Accelerate Open Science, FAIR data management and use of digital methods and services
- Stimulate co-operation in science and research, new insights and innovations, higher research productivity and improved reproducibility in science.

An open, trusted, federation of infrastructure

- Access existing Research Infrastructures in Europe;
- Enable circa 2 million European researchers to store, share, process, analyze, and reuse research digital objects (e.g. data, publications and software)

An evolving ecosystem

- Bringing together the European Commission, the governments and the many R&I stakeholders involved in the European Research Area
- Co-created across European, national, and institutional levels

There had been previous calls under the “INFRA-EOSC” banner, but despite the name those were used to integrate and seek synergies between research infrastructure clusters. Well intentioned but lots of sticky tape and rubber band.

Then a next round came up, called infraeosc-2019, and we got the secret signal that the EC was prepared to accept a proper infrastructure proposal, provided it was federation oriented.

so we pitched:



CS³ MESH⁴ EOSC

Connecting European Data

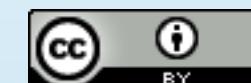


Science
Mesh

Science Mesh: Unlocking scientific collaboration through technology



CS3MESH4EOSC has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No. 863353.



13/10/2025

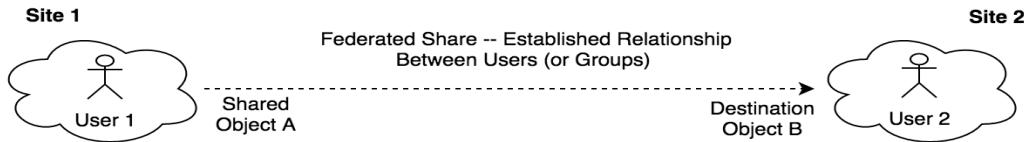
CS3MESH4EOSC

- **3-year Project**
 - Started **January 2020**
- **Objectives**
 - Building, operating and developing governance for a **Decentralized Mesh of EFSS nodes**
 - Linking users and datasets across EFSS systems, using the OpenCloudMesh protocol
 - Building on the existing userbase and data holdings of the collective of NREN EFSS services (“the CS3 community”)
 - Further developing OpenCloudMesh, so it can carry new applications and use cases.
- Split in **5 Work Packages**
- Led by **CERN**

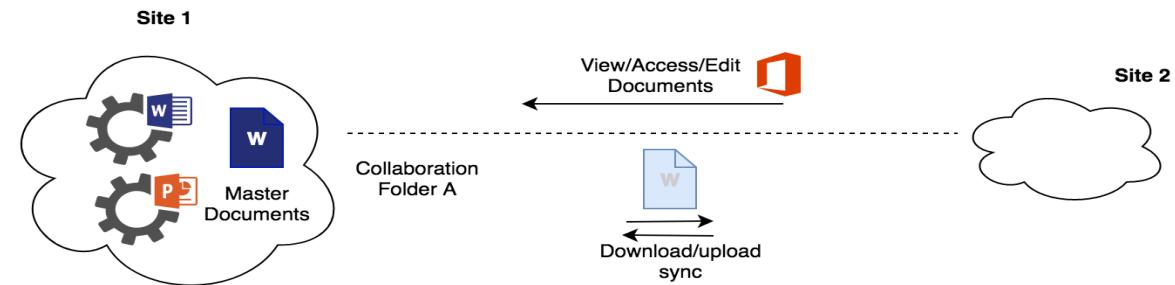


four new OCM use cases:

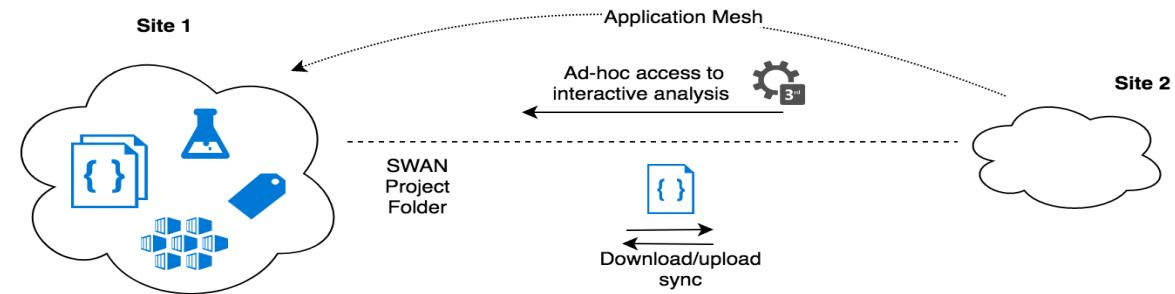
Federated Sharing



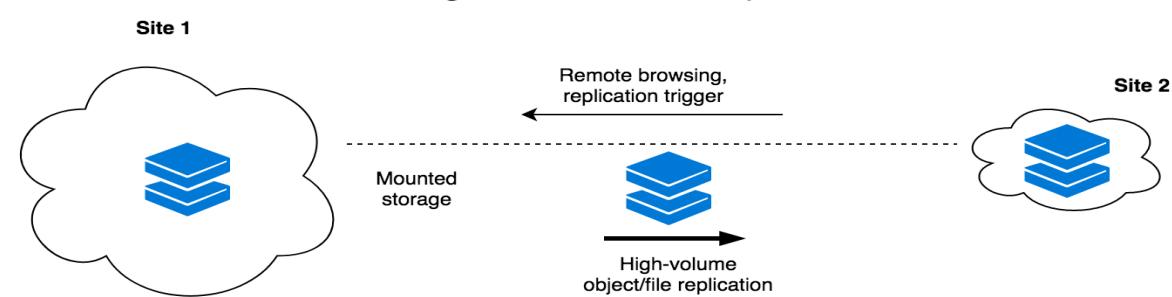
"Collaborative Documents"



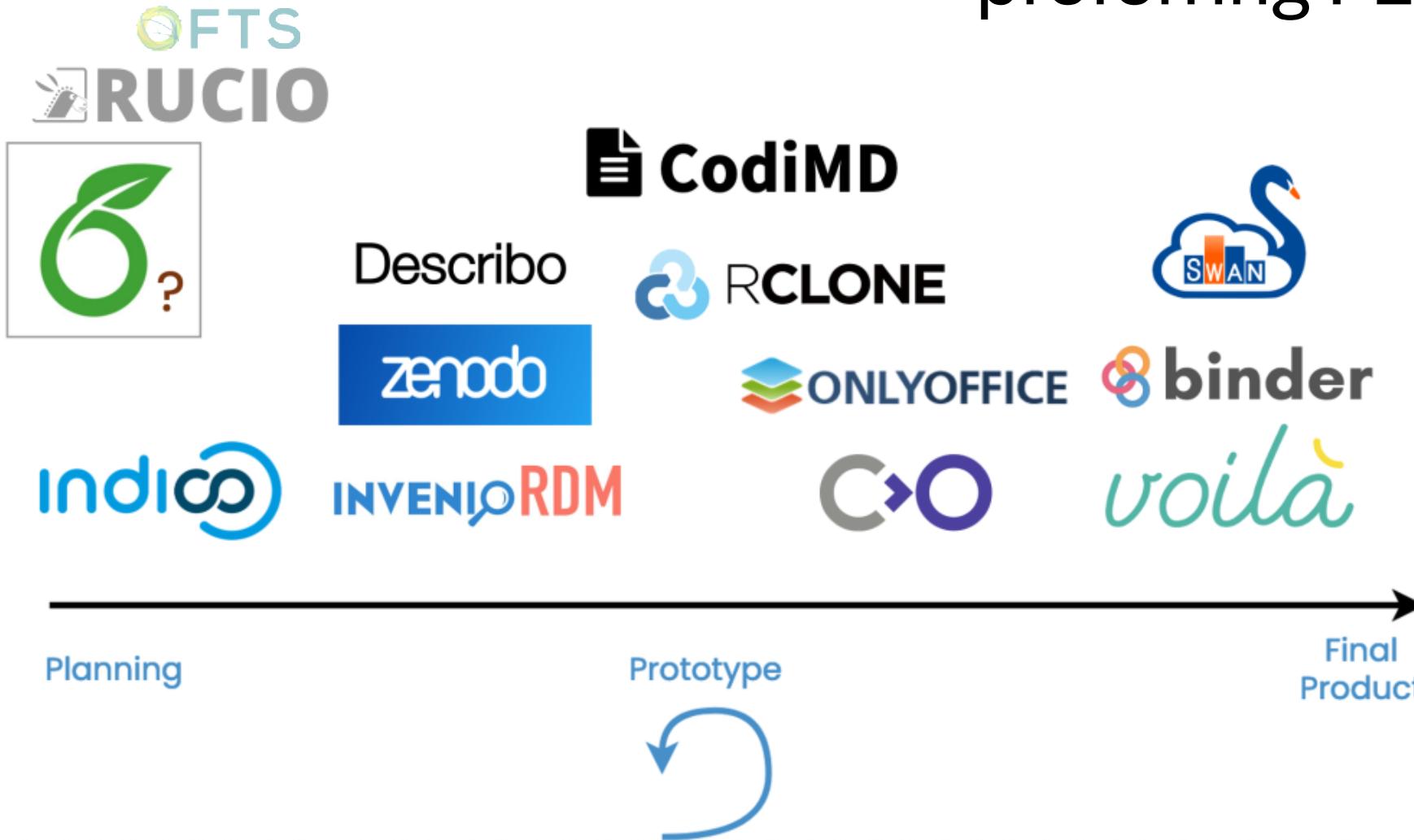
"Remote Data Analysis"



"High Volume Data Replication"



We reused wherever we could,
preferring FLOSS



EOSC 2020-2025

EOSC finally makes a start at an architecture

In 2022, the EC made a bold move.

Instead of creating a new grant scheme, they decided to tender for the next iteration of EOSC development.

Not huge: only 10% ($\sim 8 \times 10^9$ €) of what EOSC had cost so far ($\sim 85 \times 10^9$ €)

Tendering meant they had to issue a detailed specification for what they wanted. This was a break from letting the R&E ecosystem set direction.

The tender spec they issued looked *a great deal* like the outcomes of two INFRAEOSC-2019 projects: CS3mesh4EOSC and EOSC-Future.

This was fantastic news for us NREN operators, of course.

CONTEXT:

- Change of tune: EC switches to tendering as opposed to granting for EOSC services
- Tender start May 2022, end Oct 2023
- Objective: build a “pilot node” for the EOSC federation architecture
- CNECT leads effort (not RTD)
- Bidding team with SUNET involved wins Lot #2 and #3 contract, January 2024

• Structure of lots (#1-#3) of the pilot node

- **EOSC-Core: Lot 1**
- **EOSC-EXCHANGE: Lot #3**

Data centric user facing services:

- Collab platform: EFSS / Sync & Share
- Data science platform: Notebooks (Jupyter)
- User data ingest / mobility: Filesender

- **EOSC-EXCHANGE: Lot #2**

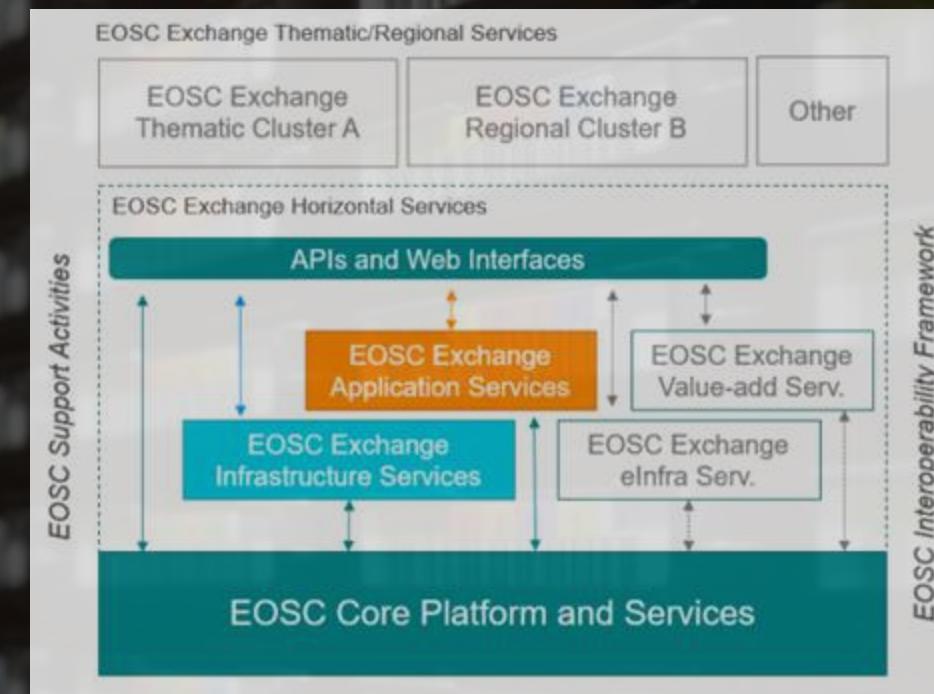
Infrastructure and platform service

- VMs (“cloud compute”)
- Container platform (OKD)
- Bulk data transfer / m2m (FTS)

- **EOSC-CORE: Lot #1**

(for the purposes of this talk, these are b2b / federative support services)

EOSC EU NODE TENDER



WHY A TENDER?

- EC used a tender model to set the bar higher on compliance:
 - Technical excellence
 - Economic viability (turnover, purchase power)
 - Service management procedures
 - Certification
 - Security handling
- **To repeat:**
 - The EC is, here, contracting on commercial terms. They are *not*, (as we are collectively more used to), making available the monies **as a grant, where they'd await the delivery of infrastructure and services as a project** deliverable towards the end of the project lifetime.

MODEL AND REQUIREMENTS



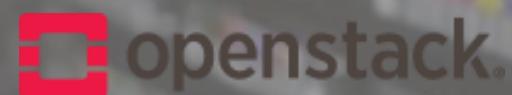
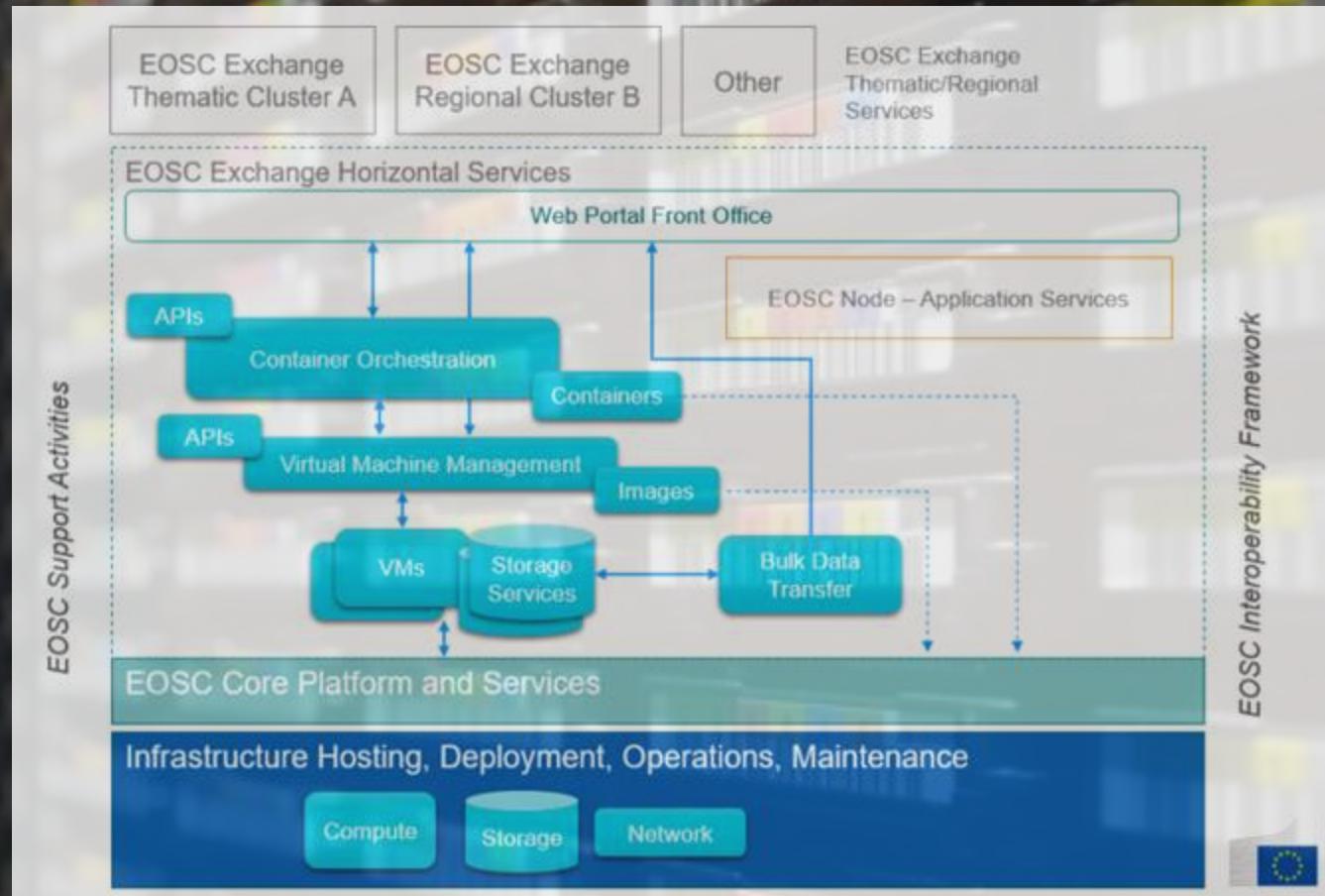
Source: <https://www.istockphoto.com/photos/challenges>

SCOPE

Lot #2:

- tender specs strongly encouraged use of open source sw and industry standard interfaces
- Solution tendered:
 - **Infrastructure (IaaS):**
OpenStack-based cloud
 - **Platform (PaaS):**
Kubernetes / OKD
 - **Bulk data transfer:**
FTS and family
 - => everything in ‘vanilla’ versions with minimum possible integration/dev. work (TRL 9)

WHAT'S INSIDE (LOT #2)

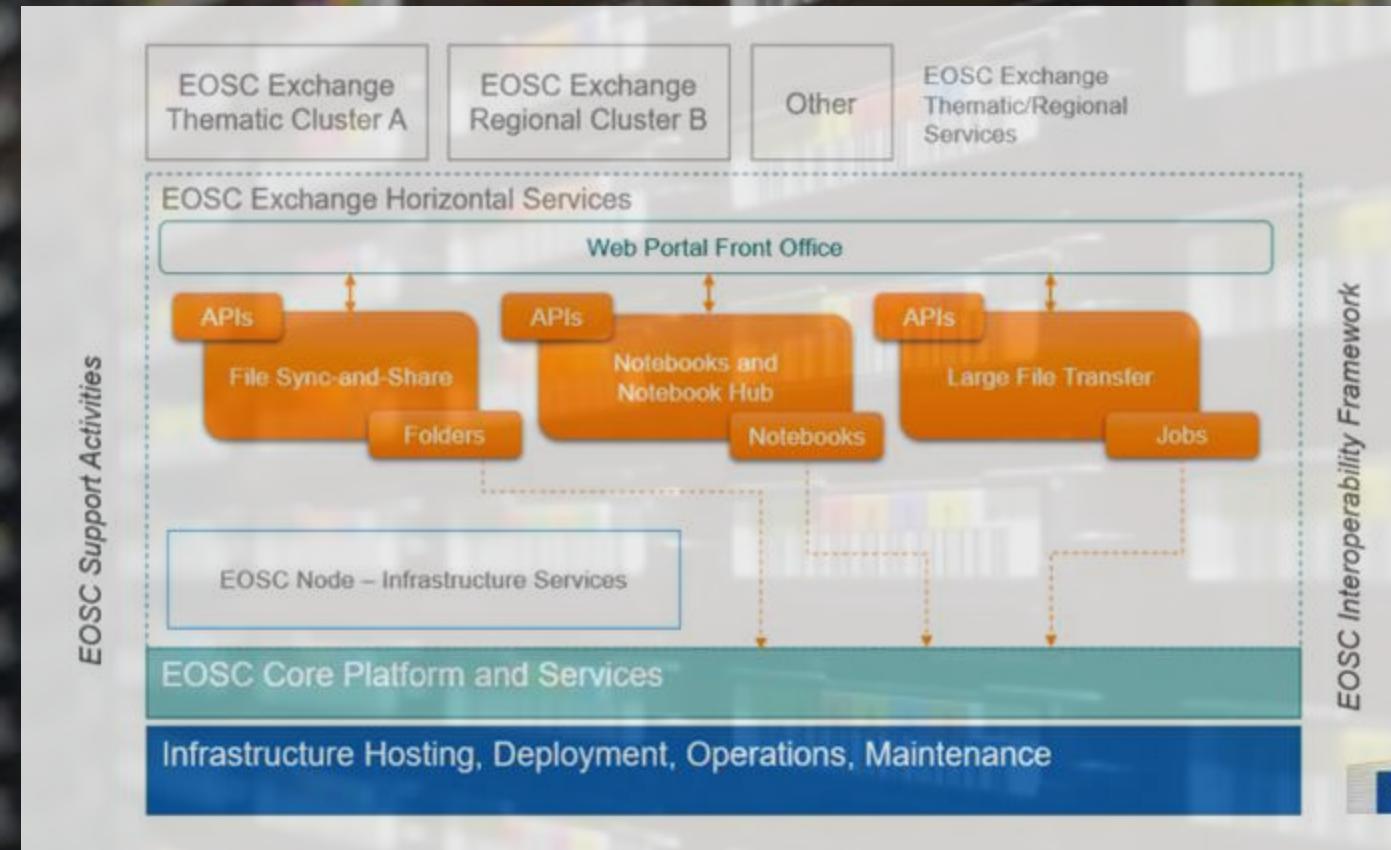


SCOPE

Lot #3

- Here too, tender specs strongly encouraged use of open source sw and industry standard interfaces.
- Data-centric services
- EFSS / synch&share: ownCloud
 - Data science: Jupyter notebooks
 - User file transfer: Filesender
- Bid group composed mostly of Not-for-Profit R&E (NRENs)
- A few companies and freelance developers onboarded into the bid group
- These are all 'usual suspects' familiar to the R&E ecosystem; no sun tzu events expected.
- **Important notes:**
 - Compared to Lot #2, the spec for Lot #3 was heavier on R&E specific requirements and value-adds. This is, therefore, also where we can delight the EC by proving we have valid use cases and by harvesting further service improvement cases from these use cases.
 - Lot #3 is also the lot where the most realistic case for service federation exists (the EFSS + the ScienceMesh recipes from cs3mesh4eosc)
 - Lot #3 is allowed to scale up using compute infrastructure provided by Lot #2 – which conveniently is "also us"

WHAT'S INSIDE (LOT #3)

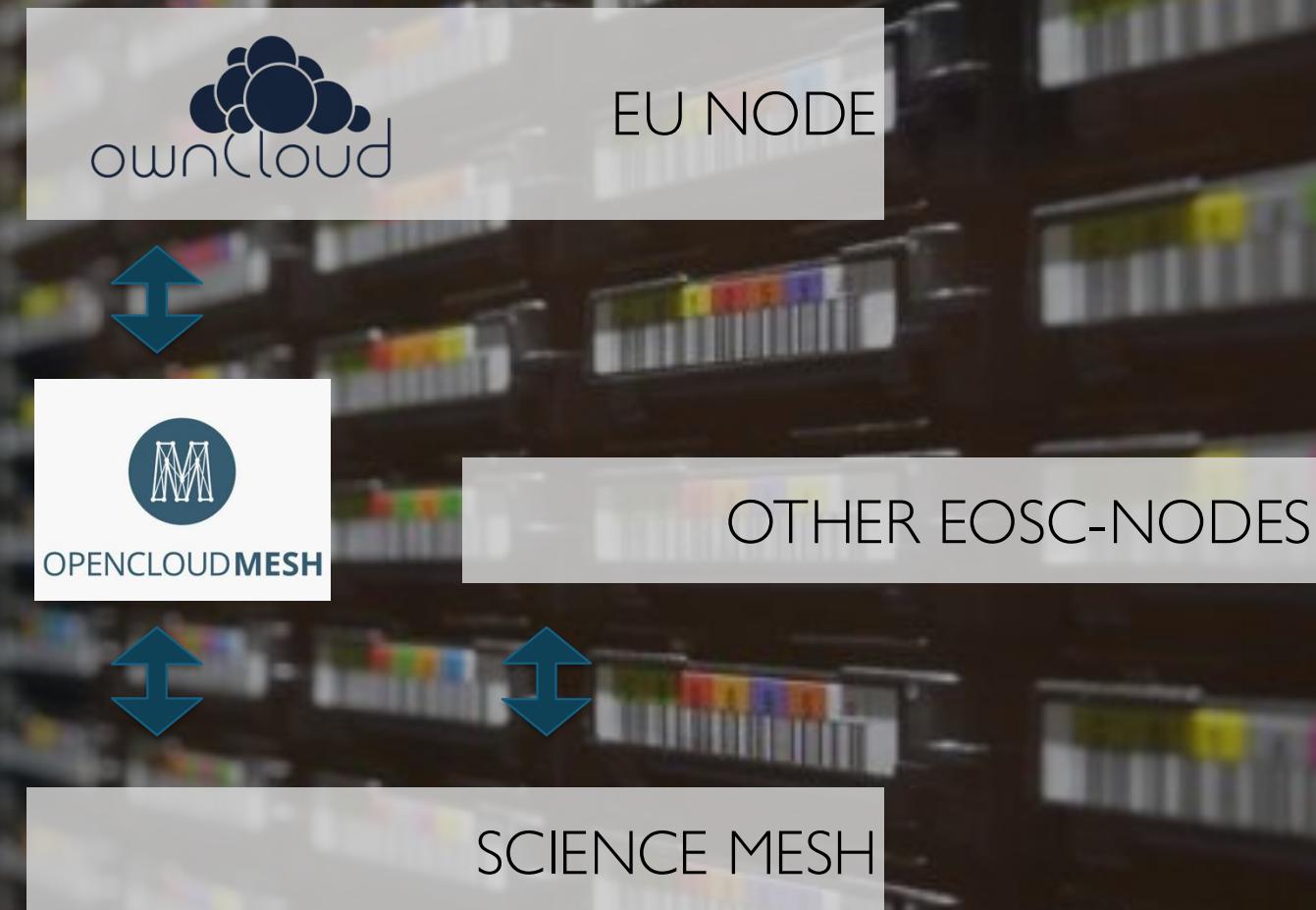


PRACTICAL FEDERATION

Lot #3:

- The specs demand that the EFSS (synch&share) collab platform provided *must be* interoperable with the ScienceMesh platform delivered for EOSC by the cs3mesh4eosc project
- Interoperable means:
[PROTOCOLS, PROCEDURES, ARCHITECTURE](#)
- We do this by offering an ownCloud based EFSS; ownCloud took part in cs3mesh4eosc.
- Its product (OCIS) has the required protocols and API built in (CS3APIs and OCM)
 - OCM is the basic building block the ScienceMesh uses **for cross-site data access and storage**
 - The idea is to turn this into a federation of upcoming EOSC nodes, with the EC Pilot node as the seed node.

LOT #3 SCIENCE MESH COMPLIANCE

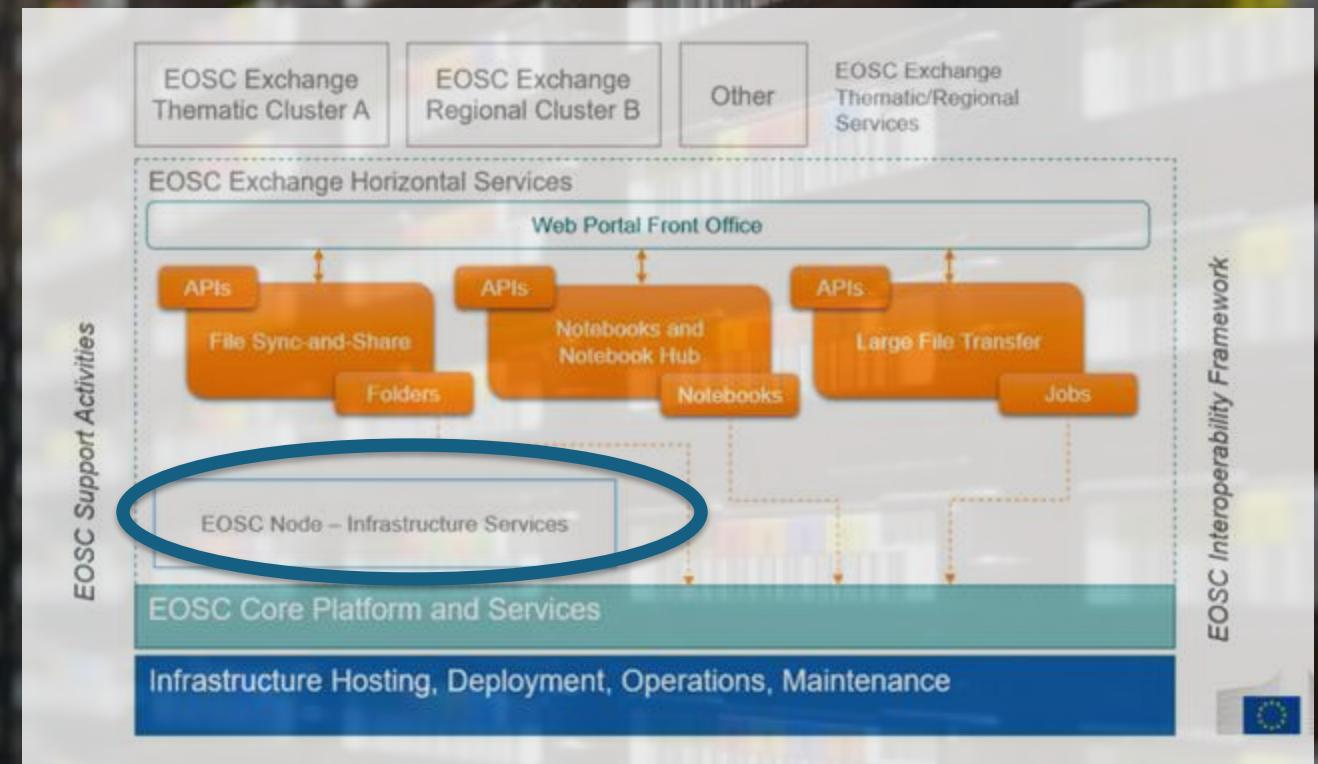


FOCUS ON

Lot #3 Integration options are plentiful:

- Vertical down:
 - scale through compute infrastructure from Lot #2
- Horizontal:
 - move raw data out of EFSS into jupyter; move result data back into EFSS
 - Direct repo deposit from EFSS (e.g. into Zenodo or OSF)
- Vertical up:
 - integrate Lot #3 services into science VREs (a tender requirement!)

OTHER INTEGRATIONS



EOSC EU Node

Objectives:

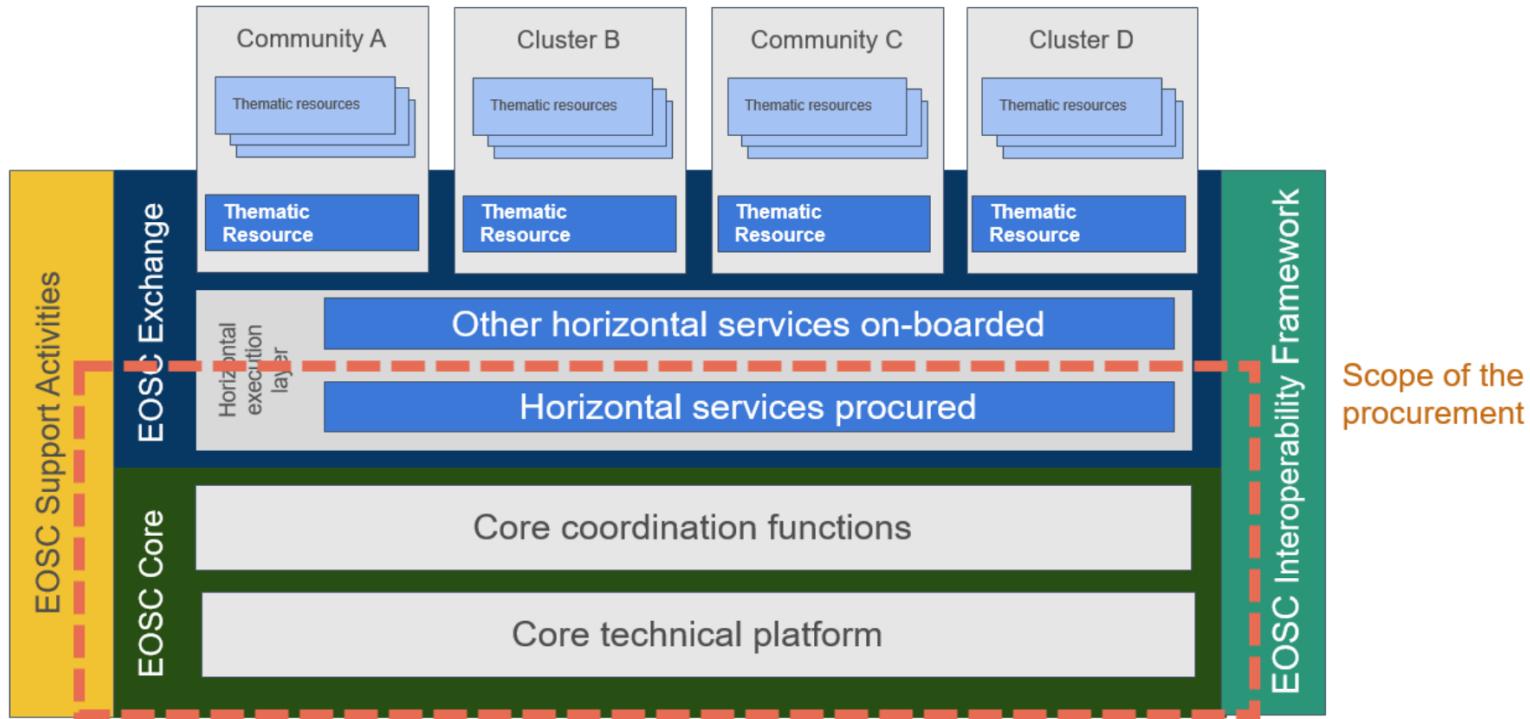
- build and deploy a **fully operational** enabling infrastructure for EOSC
- secure, cloud-based, based on **EOSC Interoperability framework**
- offering **high quality managed services** and superior user experience for a large number of users, with the functionalities **available 24/7** (*the procurement also covers operations, maintenance, and support of the EOSC EU Node for 36 months*)

Value proposition:

- **Facilitate** the creation of the “*Web of FAIR data and interoperable services*” (aka. EOSC Federation) under the Open Science Policy
- Put a “**seed in the ground**” by operationalizing the first EOSC Node at the European level for the initial 3 years
- Offer “**core services**” for scientific research infrastructures to federate (single-sign-on, catalogues, knowledge graph, application workflow, monitoring, accounting, helpdesk) and **common “horizontal services”** for end-users to benefit from (compute, containers, data transfer, notebooks, file sharing, open research data)
- Define **pathway and blueprint** (EOSC Interoperability Framework) for other potential EOSC Node operators to join the federation



EOSC – EOSC Nodes more generally



- The concept of a Node remains **F!u-ZzY©**, plenty of discussion are ongoing.
- This fuzziness is expected to begin to clear up once implementation begins, starting in 2025.



THE FUTURE OF EOSC

And whether it's relevant or
actionable to non-EU actors

EOSC – 2025 and beyond

- **2025 will be ‘the year of the design and initial deployment** of the EOSC Federation’. This is a marked break from ~7 years of whitepaper process.
- Built from “Nodes”, with the EC effectively kicking off the process.
- This EOSC EU node is *marketed* as “not a master node, but a reference node.”
- The **EOSC Federation will include multiple nodes**, each country and thematic community is expected to discuss their optimal setup.
- The **member states have a key role** in the post-2027 decision-making process.

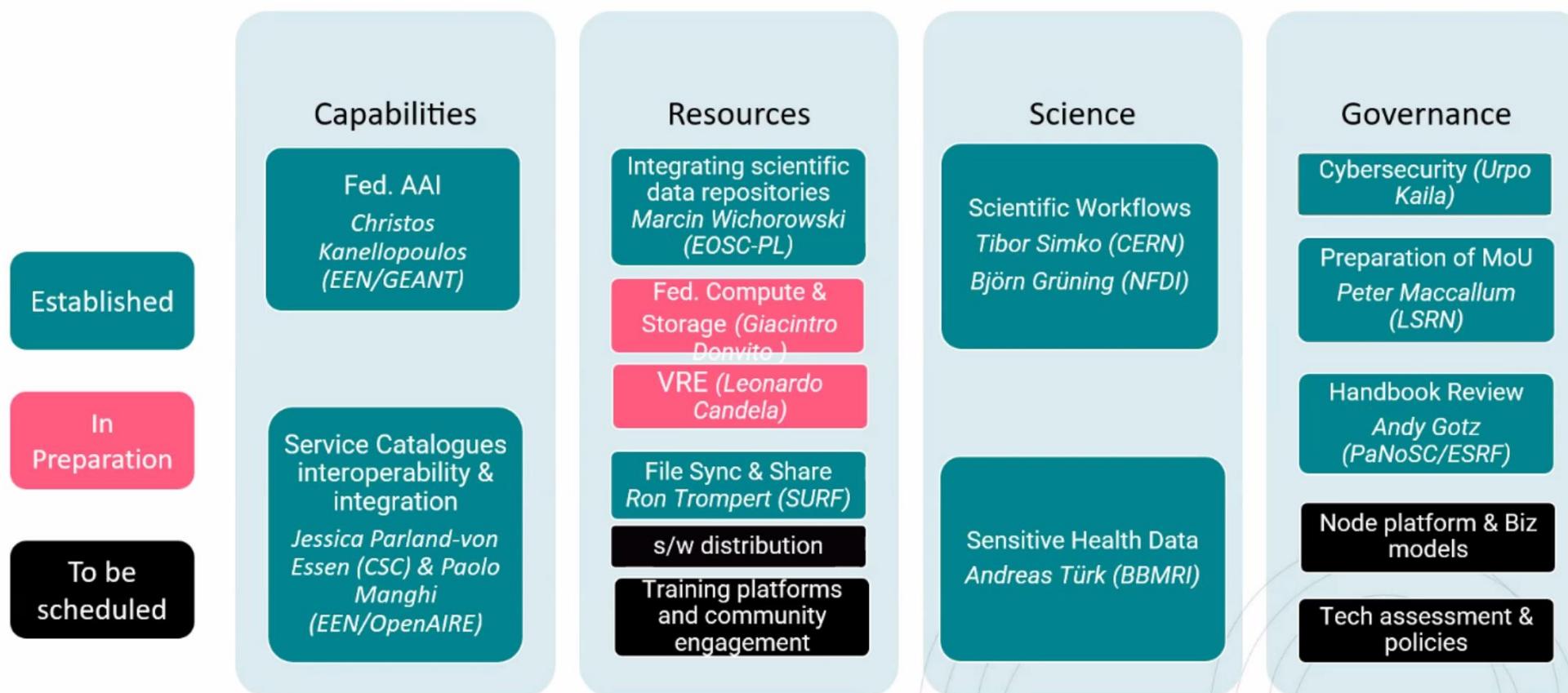
This is all complex and relatively underdefined; which implies risk. Decision making appears to be iterative. The more *involved* you are (through committees, through grants, through expert participation) the better your *grip* on things, the better you can *prepare*.

Future of EOSC: where we stand

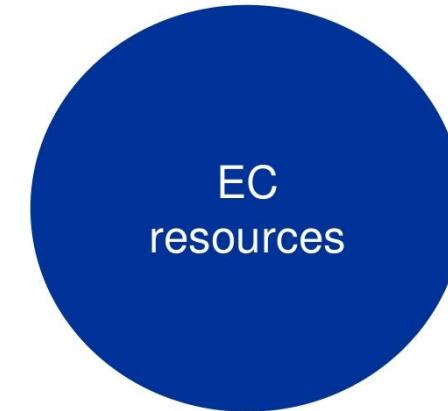
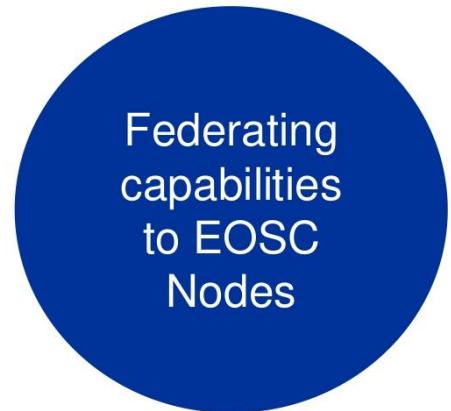
EOSC Federation build-up



Status of Subgroups



EOSC EU Node



The 'glue' of the Federation

- Authentication & Authorisation
- Services catalogue
- Accounting
- Helpdesk
- Service monitoring
- ...

Based on virtual credits

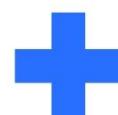
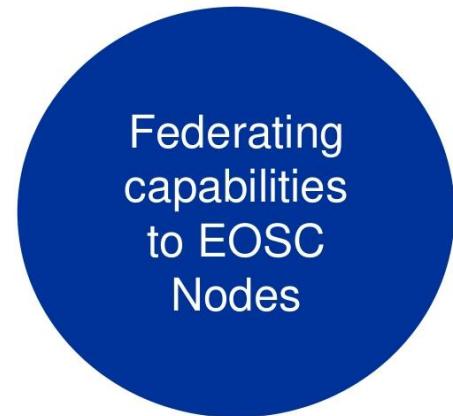
1. File sync & share
2. Interactive Notebooks
3. Large file transfer
4. Bulk data transfer
5. Virtual Machines
6. Cloud containers

Commission's 'gateway' to EOSC

- data.europa.eu (incl. JRC data)



What future for EOSC EU Node?



2025 INFRAEOSC topic to support provision of Federating capabilities also by other organisations

(Planned) 2027 INFRAEOSC topic to support provision of digital services also by other organisations

Commission's 'gateway' to EOSC



EPILOGUE

lessons learned

- In a landscape this complex, keeping definitions and expectations vague is costly. Consider taking the pain upfront.
- Kill your darlings. Be honest about the preconditions of doing what you love to do. “You can’t just build the penthouse”. Realities for commercial enterprise are still realities in R&E
- R&E operators retain an important role in this game
- Japan probably did well not getting *too* involved in EOSC so far. 2025 and onwards promise to be a less chaotic time
- If you’re confused about something, trust your instincts. It may well mean the thing you’re looking at is confusing – it’s not that you are confused ;)

Thank you!

Any questions?

guido@sunet.se

We would appreciate it if you could structure your presentation to trace the trajectory from CS3MESH4EOSC to the EOSC EU Node.

In particular, regarding CS3MESH4EOSC services, we are especially interested in the following components.

Items marked with (★) are those we consider highly relevant to our own NII Research Data Cloud (NII RDC) initiatives:

- Online office tools: OnlyOffice (★), Collabora, CodiMD
- Data science environments: Jupyter Lab (★)
- Data management and publishing functionalities:
 - + Publishing from Sciebo RDS (OwnCloud) to OSF (★), as well as to Zenodo and Dataverse
 - + Similar configurations using Nextcloud in other NREN nodes
 - + Compatibility with Invenio RDM and RO-Crate (★)
- Data transfer tools: RUCIO, FTS, RCLONE

Regarding the EOSC EU Node, as mentioned previously, we are particularly keen to hear your views on the following points:

1. The current status and roadmap of the EOSC EU Node initiative
2. Technical architecture and service composition, especially about federation based on the “system-of-systems” principle
3. Interoperability approaches with national nodes and thematic services
4. Operation and challenges of multisectoral and multilateral cooperation
5. Lessons learned that could be useful for similar efforts in Japan