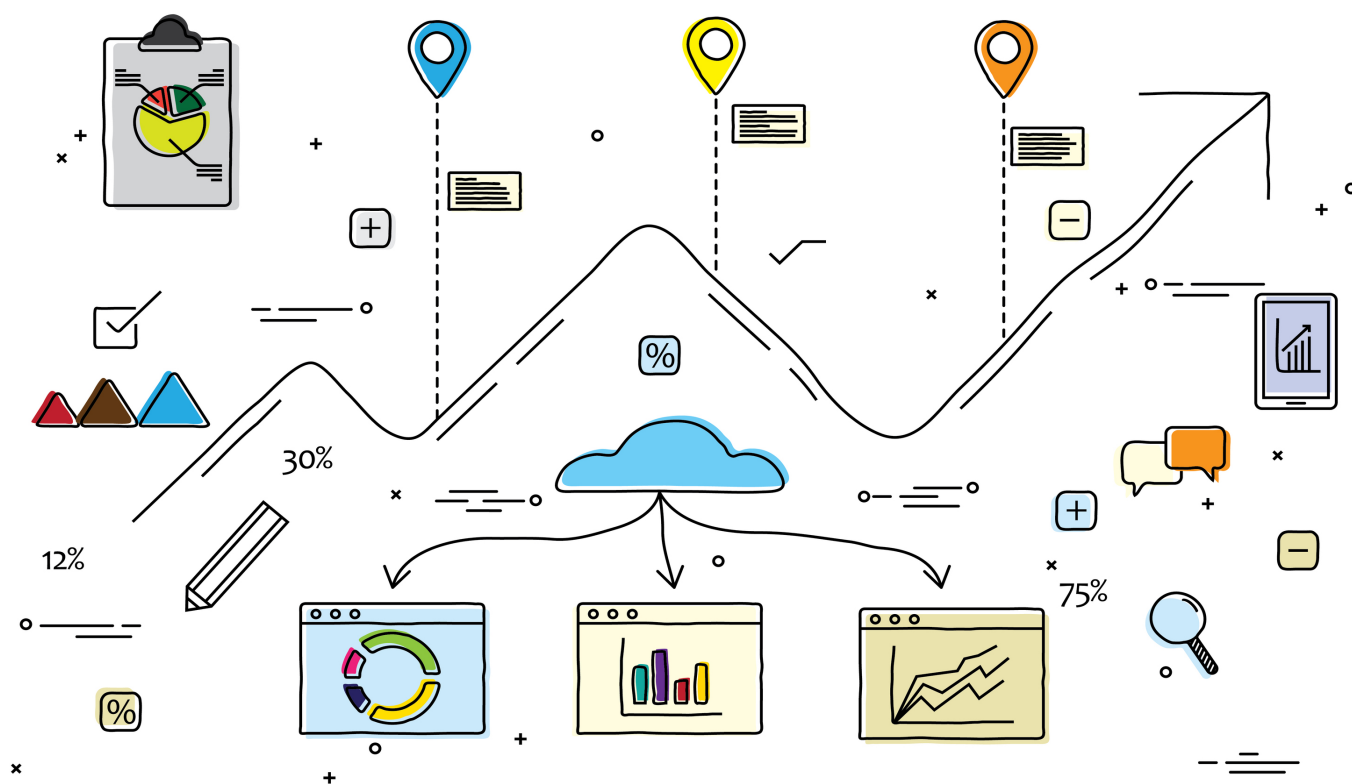


How the Science Cloud could pay its way

Business models for a sustainable European Open Science Cloud



A report of the Science|Business Network's Cloud Consultation Group

July 2018

In the first half of 2018, members of the Science|Business Network's Cloud Consultation Group met twice to discuss potential business models for the European Union's proposed "science cloud" project. This report is based partly on those discussions, but is ultimately a product of Science|Business. The views expressed herein do not necessarily reflect those of individual members.

This report is the fourth in a series that the group is working on, to gather private and public sector expertise on topics of importance to the development of the European Open Science Cloud. The other reports are available on www.sciencebusiness.net:

The case for the cloud, May 2017

Governing the Open Science Cloud, October 2017

The European Open Science Cloud: Who pays for what?, February 2018

Cloud Consultation Group members:

Amazon

Association for Computing Machinery – Europe Policy Committee

Association of Commonwealth Universities

Barcelona Supercomputing Centre

CERN

ETH-Zurich

European University Association (EUA)

European Space Agency

GÉANT

Google

Huawei

Microsoft

Research Data Alliance Europe

University of Eastern Finland

University of Twente

Rapporteur: David Pringle

© 2018 Science Business Publishing Ltd.

Avenue des Nerviens 79, Brussels, Belgium 1040

info@sciencebusiness.net

+322-304-7577

Executive Summary

This paper explores potential business models for the European Open Science Cloud (EOSC), the EU's bold plan to use cloud computing to knit together Europe's disparate research resources.

The EOSC is set to become a multi-sided marketplace, akin to Amazon.com or Apple's App Store, in which the value to each market segment depends on the participation of other market segments. Once it reaches a critical mass, the EOSC will, like other marketplaces, be subject to so-called network effects (it grows in value as it attracts more participants).

Where will the money come from?

The EOSC could achieve that critical mass by creating a compelling proposition for the 725,000 researchers working in higher education in the EU. Extrapolating from UK figures, EU governments' science ministries and the EU institutions fund approximately two-thirds of the research conducted in EU universities. If and when Member States' governments and the EU mandate that data generated by publicly-funded research is made openly accessible, the EOSC could help researchers in higher education meet that requirement. As a result, the open science cloud could initially be serving researchers responsible for about €45 billion of annual research spending (the OECD estimates that higher education in the EU spends €69 billion on research each year).

If university researchers, via their IT departments and librarians, were to spend up to 5% of the €45 billion on services provided via the EOSC, the open science cloud marketplace could be a conduit for €2.25 billion of transactions per year.

Moreover, some of Europe's major research infrastructures, known as the big data science factories, already make some data sets and analytical tools available in an open manner – in many cases, policymakers regard public usage of the data generated by research infrastructures as a key performance metric. If it can engage university researchers across the EU, the EOSC could help the big data science factories demonstrate how useful their work is to the wider scientific community and help them justify their public funding.

The Commission has said it will invest €300 million to support the core functions of the EOSC in phase 1 (up to 2020). In phase 2, after 2020, the Commission envisages the activities of the EOSC could be financed by a mix of funding, including possibly deposit fees from national funders, as well as revenues generated from users.

A freemium model to encourage early participation

Although it will ultimately need to break even, the initial priorities for the EOSC have to be driving participation and usage. Like most private businesses, the EOSC will probably need to operate at a loss (be subsidised) in its early years to ensure its proposition is appealing to both the data providers and the data users. However, once it is benefitting from network effects, the EOSC will be able to implement a sustainable business model. To do that, the EOSC will need to be highly relevant for the private sector. As outlined by the European Investment Bank Advisory Services in a recent report¹, access to relevant, commercially interesting data will clearly be needed to engage businesses in the EOSC.

¹ Financing the future of supercomputing: How to increase investment in high performance computing in Europe. Prepared for: DG Research and Innovation and DG Connect European Commission By: Innovation Finance Advisory European Investment Bank Advisory Services

To help the EOSC achieve critical mass, it would be prudent initially to give registered researchers in higher education free access to EOSC-certified data and related research tools provided by their peers in higher education. Under this “freemium” model, usage of the data and tools would be monitored to prevent abuse (only limited extracts could be downloaded), while ensuring the researchers who generated the data and tools receive recognition. This monitoring would also enable the EOSC executive body to track usage patterns and refine the overall proposition.

The EOSC will have to compensate the providers of the authentication, search and discovery tools/catalogues researchers will employ as they use the EOSC. Moreover, research communities will have to pay for their usage of commercial IaaS, PaaS and SaaS services provided by EOSC-accredited service providers¹ at commercial rates. Note, these service providers would be competing with each other within the EOSC – the research team and/or its IT department would select the service provider that offers the most attractive proposition.

Once their research has been peer-reviewed and the resulting data and research tools certified as EOSC-compliant, the research team and its higher education institution should be given public recognition. For example, their contribution could be published in an “open science index”, which ranks researchers by their contributions to the EOSC. If and when the EOSC begins to generate revenues, this index could be supplemented with credits (or cloud coins)² that can be spent on EOSC-accredited services. The credits awarded could be based on usage of the dataset.

Such a freemium model is needed to keep the barriers to usage of the EOSC as low as possible. The aim is to enable researchers in higher education to explore the potential of the open science cloud without incurring significant financial expenditure. As the number of data providers and data users increases, the value of the EOSC to both groups should rise, attracting greater participation. Once it has secured critical mass and has become the preferred means of making scientific data open and accessible, the EOSC will have the scope to adapt the freemium model to ensure the open science cloud is self-sustaining.

From 2020 (once the EOSC Executive Body has a legal status that enables it to manage funds), private companies and commercial spinouts from universities and SMEs could pay the EOSC a subscription to access the services that are available for free to public sector researchers. At this point, the EOSC could also begin charging mainstream researchers working in higher education a low subscription fee to access EOSC-certified data or research tools, if the initial freemium model proves to be unsustainable.

Other potential source of revenues (as identified by the EOSC High Level Expert Group) could be fees paid by conformant commercial service providers to offer their services in the marketplace, a transaction tax on EOSC credit vouchers (i.e. a service provider is taxed a small percentage on each voucher that is redeemed against their services), perhaps combined with a transaction tax on private sector users.

The use of EOSC credits would help drive competition between cloud providers: the researcher’s IT specialists would spend the credits with the provider offering the best value. Such a system would also allow for measurement of activity (since the EOSC would see where the credits were getting redeemed) and help ensure that European science isn’t becoming overly dependent on a single cloud vendor.

¹ The EOSC High Level Expert Group has suggested a funding and payment model that is based on a certification programme for commercial and non-commercial providers of computing services that can provide scientifically useful services, that will meet EOSC-defined standards to ensure minimum levels of access and interoperability and that will accept specific, EOSC-defined financial transactions in payment for these services.

² The EOSC High Level Expert Group envisages the EOSC and member states would certify one or more brokers to manage the acquisition, distribution and payment for EOSC vouchers. These brokers could be government agencies in member states, entities within member states, transnational governments or private firms.

Introduction

Drawing on the input from the Science|Business European Open Science Cloud (EOSC) consultation group, this paper aims to answer the following question: How does the EOSC and its stakeholders plan to generate revenue to recover their costs? To generate revenue, the EOSC needs to create value for Europe's scientific researchers.

As outlined in the Science|Business report, The European science cloud: Who will pay?, the EOSC aims to give European researchers access to a multitude of platforms, software tools, algorithms and data that they can't access today: the availability of computing resources should no longer be a bottleneck. Moreover, the EOSC is designed to create a safe and seamless environment for sharing research data. As a result, researchers in both the public and private sectors may ultimately be able to conduct new kinds of experiments and research, with a lower level of risk, which could ultimately yield major economic benefits. The net effect would be to breathe new life into existing investments and draw new money into European science, creating a virtuous circle that fuels investment in innovative businesses and new public services.

The first step towards that goal is to bring together both users and providers of research data and related services, tools and analytical models. In other words, the EOSC will be a multi-sided marketplace, akin to Amazon.com or Apple's App Store, in which the value to each market segment depends on the participation of other market segments. As the primary appeal of the EOSC will be straightforward access to research data and related tools, a high priority for the EOSC is to encompass key datasets that will attract participants. Once it reaches a critical mass, the EOSC will, like other marketplaces, be subject to so-called network effects (it grows in value as it attracts more users).

This paper contends that the EOSC should focus initially on achieving that critical mass as quickly as possible by creating a compelling proposition for researchers working in higher education.

Why should the EOSC focus initially on the higher education sector? The top priority for the EOSC is to improve the efficiency and effectiveness of public-funded research in the EU. Of the approximately one million scientists (full-time equivalents) working in the public sector in the EU, approximately three quarters are employed in higher education, according to EuroStat (see Figure 1). In some cases, these researchers may also be supporting commercial university spin-outs.

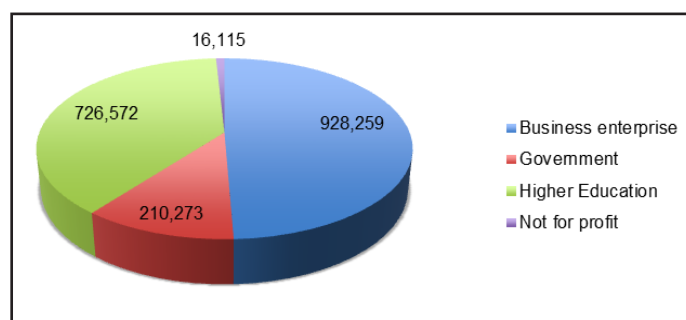


Figure 1: Breakdown of EU researchers by sector in 2016 (FTEs)
Source: EuroStat

In phase 2 of the EOSC (post-2020), the platform should also seek to engage with researchers working in private companies. This is potentially a major market for the EOSC in the medium to long term: In 2015, there were 907,000 researchers (full-time equivalents) in the business sector in the EU, according to EuroStat (2016 figures aren't yet available).

Of these, 518,000 or 57% were employed by large companies with at least 500 staff, implying they have significant resources. A further 82,000 or 9% were working for companies with 250-499 employees. The remaining 34% or slightly over 300,000 were employed by SMEs, of which:

- 155,000 or 17% were working for medium sized enterprises (50-249 employees)
- 107,000 or 12% were working for small enterprises (10-49 employees)
- 43,000 or 5% were working for micro-enterprises (with 1-9 employees)

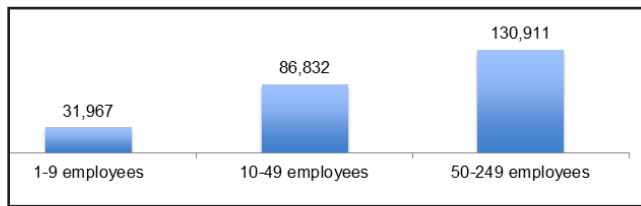


Figure 2: Number of EU researchers working in SMEs in 2015 (FTEs)
Source: EuroStat

In terms of R&D spending in the EU, the private sector accounts for the lion's share, spending more than €193 billion in 2016 (see Figure 3), according to OECD figures, compared with €69 billion in higher education and €33 billion in higher education and €33 billion by other public sector bodies. Again, these figures highlight how the private sector could be a major source of revenues for the EOSC in the medium to long term.

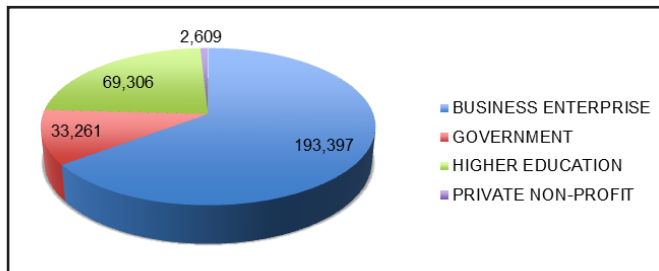


Figure 3: R&D spending in the EU by sector 2016 (millions of euros)
Source: OECD

However, in phase 1 (pre-2020), the biggest and most important market for the EOSC will be researchers working in higher education. This group receives funding from a range of different actors, spanning the public, private and charity sectors. Although there are no EU-wide figures available showing the sources of the R&D funding conducted by higher education, the Royal Society has published a breakdown for the UK. In 2013/14, the UK government's science budget funded almost 55% of the £7.04 billion spent by UK universities on research (see Figure 4). Other public agencies were responsible for a further 13%, while the EU was the source of almost 10%. Charities accounted for 16% and the private sector for 7%.

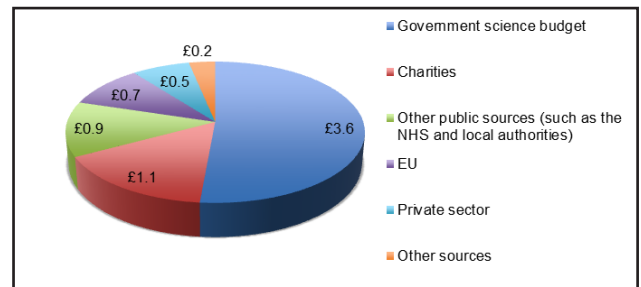


Figure 4: Sources of funding for research conducted in UK universities (billions of pounds)
Source: Royal Society, figures for 2013/14

If the pattern shown in Figure 4 is replicated across the EU, that would suggest central governments' science departments and the EU fund approximately two-thirds of the research conducted in EU universities. Member States' governments and the EU are beginning to mandate that data generated by publicly funded research is made openly accessible. If the EOSC offers the best way for researchers to meet that requirement, the open science cloud could initially be serving researchers responsible for up to €45 billion of annual research spending, which should be sufficient to give the EOSC the critical mass it needs to gain network effects and expand to address the needs of scientists working in public sector agencies and the private sector.

If university researchers, via their IT departments and librarians, were to spend up to 5% of the €45 billion on services provided via the EOSC, the open science cloud marketplace could enable up to €2.25 billion of transactions per year. That should be sufficient to motivate service providers from the private sector to support the EOSC with appropriate IaaS (infrastructure-as-a-service), PaaS (platform-as-a-service) and SaaS (software-as-a-service) propositions.

By way of context, the European Commission estimates that the EU (including the Member States) already spends approximately €10 billion a year on data infrastructures for science conducted in universities and other publicly owned facilities. However, much of that funding is provided by Member States, which are likely to support the EOSC with varying degrees of enthusiasm.

The value proposition for researchers in higher education

At a high level, the immediate appeal of the EOSC for researchers in higher education revolves around convenience – if they can quickly find relevant data and tools, they will be able to do better-informed research, much faster¹. If it fulfils its mission, the EOSC will provide a single reliable shop from which researchers can source data organized on FAIR principles², together with related analytical models and tools. Note, the data, tools and models available via the EOSC will come from a mix of sources, including researchers in other universities, commercial service providers, publicly-funded e-infrastructures and specialised research infrastructures. By providing cloud-based collaboration tools and working spaces, the EOSC should also make it easier for researchers to partner with their peers working on similar projects at other institutions.

Figure 5 shows how the EOSC could federate different entities, depicting the necessary horizontal federation services, such as authentication, accounting and billing, and the different kinds of tools that would be available to researchers using the EOSC.

Note, the cloud services shown in the pink boxes in Figure 5 will involve a mixture of IaaS, PaaS and SaaS.

EOSC services

In its March 2018 staff working paper on the EOSC, the European Commission suggested that the EOSC will provide the following services:

- Identification and authentication to enable access to EOSC resources (free of charge).
- A protected and personalised work environment/space (e.g. logbook, settings, compliance record and pending issues) (free of charge).
- Relevant service information (status of the EOSC, list of federated data infrastructures, policy-related information, description of the compliance framework) and relevant guidelines. (free of charge).
- Catalogues of datasets (free of charge).
- Data mining, analytics, fusion and processing services, which enable users to find, access, re-use and analyse research data generated by others (may incur a fee).
- Services to make data FAIR, to store it and ensure long-term preservation (costs to be covered by EOSC business model).

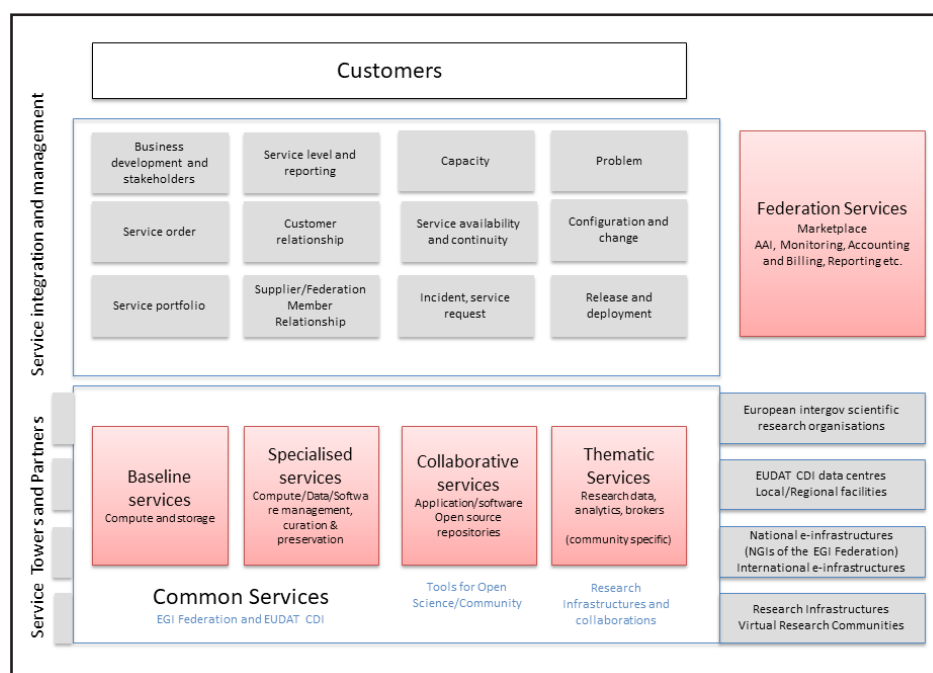


Figure 5: How the EOSC will bring together various services and tools for researchers
 Source: *The EOSC pilot*

¹ For a discussion of how the EOSC could make research easier, see the Science|Business report, the Case for the Cloud: <https://sciencebusiness.net/report/case-cloud>

² FAIR: findable, accessible, interoperable and reusable

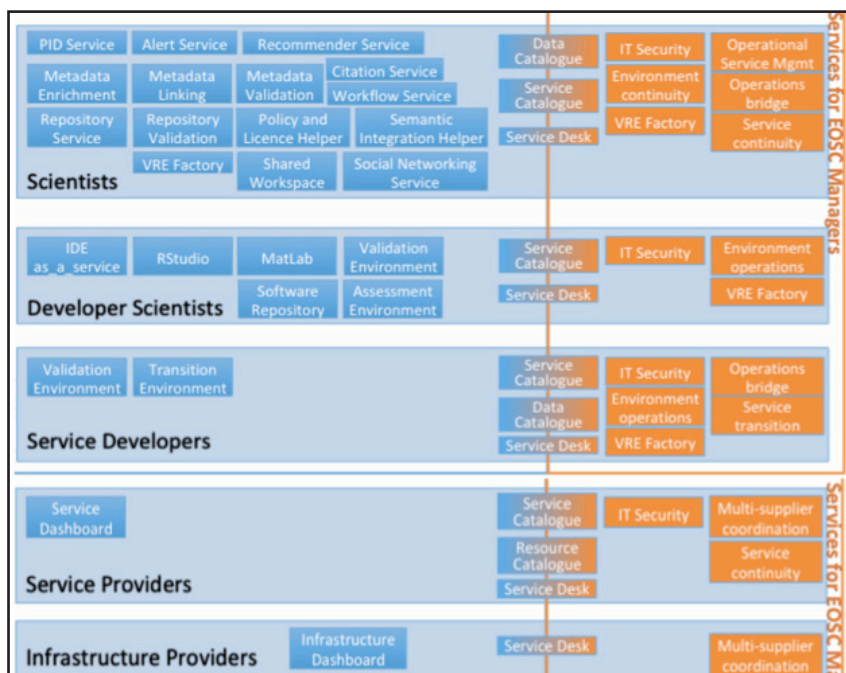


Figure 6: Overview of the services that the EOSC could provide to each group of users
 Source: The EOSCpilot

The EOSCpilot has further segmented these services, as shown in Figure 6. It defines the roles of the various actors as:

- Scientists: use functions and tools to perform their domain-specific research activities and to collaborate.
- Developer scientists: develop new analytical models, new processes and tools to analyse data and derive knowledge
- Service developers: develop services for scientists by understanding their requirements and the Open Science vision.
- Service providers: operate and provide access to their own portfolio of services according to declared service level agreements .¹
- Infrastructure service providers: provide access to computational, storage and network resources according to declared service level agreements .²
- EOSC managers: facilitate the operation, assistance and quality assurance of the EOSC system and coordination among different stakeholders.

¹ The nature of such agreements and how they are enforced will depend on whether the provider is in the public or private sectors.
² The nature of such agreements and how they are enforced will depend on whether the provider is in the public or private sectors.

Funding these services

In phase 1 (up until 2020), the Commission has said it will invest €300 million to support the core functions of the EOSC, while it has asked Member States to flag the national initiatives that they want to federate into the EOSC and the resources they are willing to provide in kind.

Moreover, funding agencies will start making research grants conditional on the recipients making the data generated by their research projects open and accessible. For example, EU grants awarded through the forthcoming Horizon Europe framework programme will require researchers to store their data in cloud-compatible repositories.

In phase 2, after 2020, the Commission envisages the activities of the EOSC could be financed by a mix of funding, including possibly deposit fees from national funders, as well as revenues generated from users. During Phase 1, the EOSC Governance Framework will be tasked with producing a full cost estimate for the running of the EOSC, which will be the basis of a financial plan for Phase 2, addressing both scalability and legacy.

The initial service charging model

To help the EOSC achieve critical mass, it would be prudent not to charge registered researchers in higher education to access EOSC-certified data and related research tools provided by their peers in higher education.

Under this “freemium” model, usage of the data and tools would be monitored to prevent abuse (only limited extracts could be downloaded), while enabling the generators of the data to receive recognition. This monitoring would enable the EOSC executive body to track usage patterns and refine the overall proposition. This freemium model could be applied indefinitely or it could only apply up until the end of 2020, after which researchers working in higher education could pay a low subscription fee to access EOSC-certified data or research tools.

In any case, the EOSC will have to compensate the providers of the authentication, search and discovery tools and catalogues researchers will employ as they use the EOSC. Moreover, researchers themselves will have to pay for their usage of commercial IaaS, PaaS and SaaS services provided by EOSC-accredited service providers¹ at commercial rates.

Note, these service providers would be competing with each other within the EOSC – the research team and/or its IT department would select the service provider that offers the most attractive proposition.

In time, they may also benefit from a volume discount negotiated between the EOSC executive body and the service provider once the EOSC is widely used. These services should, of course, be available from multiple providers to ensure that users of the EOSC have both choice and access to competitive pricing. (Note, some providers of PaaS and SaaS may source IaaS from another entity).

¹ The EOSC High Level Expert Group has suggested a funding and payment model that is based on a certification programme for commercial and non-commercial providers of computing services that can provide scientifically useful services, that will meet EOSC-defined standards to ensure minimum levels of access and interoperability and that will accept specific, EOSC-defined financial transactions in payment for these services.

In the case of publicly funded cloud service providers, the transactions will need to be compliant with procurement regulations, which can restrict their ability to participate in commercial tenders, offer service level agreements and penalties, and receive more than 20% of their total revenue from outside their main funding source.

To illustrate how this charging and service model would work in practice, here is a high-level description of how researchers in higher education could engage with the EOSC in the start-up phase between 2018 and 2020. Note, the subsidies inherent in this process could be tapered off once the EOSC has achieved critical mass.

- **Step 1:** A university research team would register for access to EOSC resources and would then be authenticated online by an EOSC-accredited service provider. Registration would be free of charge. The research team will then have free access to a protected and personalised work environment/space and relevant service information provided by an EOSC-accredited service provider.
- **Step 2:** Before undertaking a new project, the research team would consult the EOSC catalogue to help identify any relevant data and research tools that already exist. Provided by an EOSC-accredited service provider, this facility is available for free.
- **Step 3:** During their project, the team would spend a portion of their research grant on IaaS and PaaS or SaaS services to access and analyse data generated by relevant projects they identified via the EOSC catalogue. In most cases, the research team would be prohibited from downloading or replicating the data, thereby enabling the EOSC to monitor usage.
- **Step 4:** During their project, the team may also spend a portion of their research grant on the IaaS and PaaS services required to access research tools and models developed by scientists elsewhere and made available through the EOSC.
- **Step 5:** During their project, the team would also spend a portion of their research grant on IaaS and PaaS or SaaS services made available via the EOSC to manage and analyse data generated by their own research.

- Note, these services could include collaborative tools that enable researchers to work remotely with other teams of scientists.
- **Step 6:** During their project, the team may also spend a portion of their research grant to use IaaS and PaaS services made available via the EOSC to develop new research tools and models.
- **Step 7:** At the end of their project, the team spend a portion of their research grant to employ IaaS and PaaS or SaaS services to organise the resulting data in line with FAIR principles, and store it where it can be accessed via the EOSC for a duration appropriate to the discipline in which the research is conducted. Note, where data needs to be stored long-term, the cost could be significant. They could also use IaaS and PaaS services to add any research tools they developed for the project to the EOSC, where they can be accessed by other researchers.

The role of EOSC credits/cloud coins

Once their research has been peer-reviewed and the resulting data and research tools certified as EOSC-compliant, the research team and its higher education institution should be given public recognition. For example, their contribution could be published in an “open science index”, which ranks researchers by their contributions. This index could be supplemented with credits (or cloud coins)¹ that can be spent on EOSC-accredited services.

These credits would be shared in a transparent and proportionate manner between research teams and institutions.

Some kind of assessment process would be required to determine the value of the credits awarded. This process could be based on usage of the dataset.

The EOSC High Level Expert Group says: “Ideally, the marketplace would keep track of how frequently a data set is used and the provider of that data set compensated accordingly, similarly to how YouTube pays people who upload videos based on how many times they are viewed,” it adds.

However, such a performance-based approach would require the EOSC executive body to contract service providers to monitor usage of individual data sets. While monitoring usage adds complexity and additional cost to the EOSC, publicly-funded researchers can justify their research budgets by showing to their funding agencies the impact and re-use of their results. Such monitoring will provide an incentive for data providers and service providers to participate in the EOSC. However, this model implies that researchers will not be able to download the data, which then creates a form of exclusivity for EOSC, which may encourage participation, but has to be managed carefully to ensure that the open science cloud is a genuinely open and competitive marketplace, rather than a de facto cartel.

Note, these credits may not be redeemable until after 2020, once the EOSC executive body has a legal status that enables it to manage funds and is beginning to generate revenues. As well as funding these credits, the EOSC would also need to finance the management of the marketplace itself. How much this relies on fixed subsidies, how much from grant conditions and how much on actual value will need to be carefully monitored.

From 2020 (once the EOSC Executive Body has a legal status that enables it to manage funds), private companies and commercial spinouts from universities and SMEs could pay the EOSC an annual subscription to access the services that are available for free to public-sector researchers.

¹ The EOSC High Level Expert Group envisages the EOSC and member states would certify one or more brokers to manage the acquisition, distribution and payment for EOSC vouchers. These brokers could be government agencies in member states, entities within member states, transnational governments or private firms.

Other potential source of revenues (as identified by the EOSC High Level Expert Group) could be an annual subscription fee paid by conformant commercial service providers to offer their services in the marketplace, a transaction tax on EOSC credit vouchers (i.e. a service provider is taxed a small percentage on each voucher that is redeemed against their services), perhaps combined with a transaction tax on private sector users. A transaction tax has the advantage of removing upfront charges for participation by a service provider.

The High Level Expert Group says these revenues could contribute to funding the operational costs of the EOSC gateway/marketplace and the provision of key open data sets.

Although researchers working in the public sector could, at their own discretion, use some of their research grant to access EOSC services and make their data and tools available via the EOSC, the introduction of EOSC credits could allow for alternative mechanisms, such as:

- Researchers could apply for a specific data management budget when they apply for a research grant. This budget would be paid in EOSC credits.
- Researchers are paid in EOSC credits each time their data set, analytical model or research tool is accessed. However, as discussed above, this would require the EOSC or nominated agents to monitor usage of each dataset, analytical model and research tool.
- EU and Member States' funding agencies insist that a specific proportion of a research grant is used to pay for the storage, curation and management of research data via the EOSC. In effect, this could mean top-slicing research grants so that the research team must spend 2-3% of its received funding on data stewardship services from its chosen (and qualified) provider. The funding agency could make these funds available as EOSC credits that can be spent on any accredited cloud service.

- The funding agencies could link research funding to the identification of a capable agency that will provide the data stewardship services and products that will carry and manage the data long term. In most cases, this would be the research institution itself (as an IT or librarianship function), but researchers should have the flexibility to employ other service providers that meet the minimum standards specified by the EOSC. Note, a rigid top-slicing mechanism may face opposition from some researchers who want to retain the flexibility to spend all of their grant on conducting their research, if that proves necessary.

The use of EOSC credits could help drive competition between cloud providers: the researcher's IT specialists would spend the credits with the provider offering the best value. Such a system would also allow for measurement of activity (since the EOSC/Commission would see where the credits were getting redeemed) and help ensure that European science isn't becoming overly dependent on a single cloud vendor. EOSC credits could only be redeemed for services from accredited cloud service providers. This would prevent money being spent on facilities that don't meet the required standards, while helping to ensure privacy and security are respected in line with the EU's regulations including the General Data Protection Regulation.

Accredited cloud service providers will also need to have the capacity and capabilities to serve a lot of groups suddenly taking a close interest in a particular "hot" dataset and wanting to run workloads against it when a significant discovery is announced, and thus requires and attracts heavy scrutiny.

Moreover, an accredited cloud service provider should be able to deliver a pricing model, which is transparent about availability levels and data durability, and - importantly - is a function of total factor costs. That would reduce the risk that a university-built facility, for example, under charges for a service, because it is unaware that it is building on implicit or explicit subsidies (e.g. they don't factor costs of power or cooling or land into their "sell" price).

Such subsidies could become unsustainable if a university's low-priced cloud services become wildly popular, since the hidden mechanism in the back paying for the subsidy won't be ready for the onslaught (e.g. the campus estates department gets a bill for electricity they simply couldn't predict).

This could result in the long-term safety and stewardship of the data being imperilled. Regardless of whether the service provider is a private company or a publicly funded institution, total cost/price calculations should not rely be distorted by sunk costs and subsidies.

The EOSC should allow for business model innovations and dynamic competition in terms of pricing and proposition. For example, as they compete for EOSC-related usage, commercial cloud providers may wish to use hybrid models, involving so-called freemium models, in which the basic service is free, but users pay for premium features. Note, only a single copy of each dataset needs to be stored in the cloud (assuming the IaaS provider has appropriate back-up mechanisms to ensure data isn't lost in the event of a disaster): users pay for resources to process data and store outcomes, rather than paying to host their own copy.

Engaging the big data science factories

The success of the EOSC will depend on the active engagement of the major research infrastructures, such as CERN, EMBL, ESO, ESRF, and ILL, and e-infrastructures, such as GÉANT.

Some of the research infrastructures, known as the big data science factories, already make some data sets and analytical tools available in an open manner – in many cases, policy makers regard public usage of the data generated by research infrastructures as a key performance metric. For that reason, many of these organisations are likely to want to incorporate their existing propositions into the EOSC if the science cloud will make it easier for researchers to access their data.

For data that is already available on FAIR terms, the process of making this data accessible through the EOSC should not be too arduous or expensive. In short, the EOSC could help the big data science factories demonstrate how useful their work is to the wide scientific community and help them justify their public funding.

CERN, for example, already has two services that are designed to enable the wider world to access subsets of the data its experiments generate: the open data portal and Zenodo.

The data providers (LHC collaborations in the case of the open data portal or individual users in the case of Zenodo) have taken the effort to make their data understandable to others and agreed to make it open. However, in the case of the LHC experiments, only a fraction of their data is visible through the open data portal because of what is commonly known as an 'embargo period' and also because of the significant effort it takes to prepare it for re-use by people outside the experiment. Zenodo, is different in that it is open to everyone, including individual researchers, and hosts datasets from a multiple of research disciplines, not just high-energy physics. If CERN were to make these services compatible with the EOSC, it could extend their reach and help it further demonstrate the usefulness of its data to researchers.

However, some research infrastructures will need to be able to recover the cost of making their data and tools available through the EOSC. Once it is generating revenues, the EOSC could provide these research infrastructures with credits they can spend on the cost of using SaaS, PaaS and IaaS services to make their data and tools available through the EOSC.

Note, the EOSC High Level Expert Group suggests the Commission could, through the European Data Infrastructure (EDI) initiative, provide economic incentives for research infrastructure providers to use and co-develop shared facilities and data repositories through the EOSC.

Although these economic incentives would need to be temporary, they could be sufficient to persuade the big data science factories to make their existing open data services compatible with the EOSC.

Conclusions

Although it will ultimately need to break even, the initial priorities for the EOSC have to be driving participation and usage. Like most private businesses, the EOSC will probably need to operate at a loss (be subsidised) in its early years to ensure its proposition is appealing to both the data providers and the data users.

However, once it is benefitting from network effects, the EOSC will be able to implement a sustainable business model. To do that, the EOSC will need to be highly relevant for the private sector. As outlined by the European Investment Bank Advisory Services in a recent report¹, access to relevant, commercially interesting data will clearly be needed to engage businesses in the EOSC.

The private sector has already shown strong interest in the data being created by some of Europe's leading research infrastructures, such as the ESA and CERN. To ensure these big science data factories make their open data services compatible with the open science cloud, the EOSC has to provide these players with the distribution reach and the monitoring tools they need to demonstrate to their funders the value of their experiments and research. It can do that by making it easy for the data factories to engage with the 725,000 researchers working in higher education across the EU, as well as those launching commercial spin-offs.

To ensure researchers in higher education engage with the EOSC, the freemium model outlined in this paper is designed to keep the barriers to usage as low as possible.

The aim is to enable researchers to explore the potential of the open science cloud without incurring significant financial expenditure. As the number of data providers and data users increases, the value of the EOSC to both groups should rise, attracting greater participation. Once the open science cloud is established as the preferred means of making scientific data open and accessible, the EOSC will have the scope to modify the freemium model to ensure it can become self-sustaining.

¹ Financing the future of supercomputing: How to increase investment in high performance computing in Europe. Prepared for: DG Research and Innovation and DG Connect European Commission
By: Innovation Finance Advisory European Investment Bank Advisory Services



The Science|Business Cloud Consultation Group

In November 2018, the EU will launch a bold initiative to pull European science into the 'cloud.' The €6.7 billion European Open Science Cloud promises to change the future of research, cloud services and data technologies in Europe - changing the research game for industry, academia and governments. But how?

To track this important initiative, Science|Business runs a platform for companies, universities and research agencies. The group meets regularly in private and public to share intelligence and facilitate dialogue in Brussels. Topics include:

'Rules of Engagement'

Under what terms can private vendors, and public users, access the science cloud and the huge market – upwards of €50 billion – it can create?

Finance

Should public research grants include budgets to buy cloud services? Will labs and universities pay centrally? When corporate labs use it, will they be charged?

Infrastructure

How will broadband networks, data centres and existing research infrastructures be affected by the fast-growing new services?

Science data

How can total quality management of data be enabled throughout its lifecycle?

New technologies

How should the science cloud harness ground-breaking advances in computing and connectivity, such as AI, quantum computing and 5G?

Global strategy

How will Europe's science cloud be integrated into global research and on what terms?

SCIENCE | BUSINESS[®]

NETWORK

Bringing together industry, research and policy

Industry

Amazon	Microsoft
Amgen	Nickel Institute
Dow Europe	Novartis
Frontiers	Pfizer
GE	Sanofi
Google	Total
Huawei	Toyota

Academia

Aalto University	Trinity College Dublin
ESADE Business & Law School	TU Berlin
ETH-Zurich	University College London
Karolinska Institutet	University of Amsterdam
KTH Royal Institute of Technology	University of Birmingham
KU Leuven	University of Bologna
Norwegian University of Science and Technology (NTNU)	University of Eastern Finland
Politecnico di Milano	University of Luxembourg
Polytechnique Montréal	University of Pisa
Sorbonne University	University of Twente
Tallinn University of Technology (TalTech)	University of Warwick

Public organisations

Barcelona Supercomputing Center	Innovate UK
Business Finland	Innovation Norway
CERN	Republic of South Africa - Department for Science and Technology
Centre National de la Recherche Scientifique (CNRS)	European Investment Bank
The COST Association	European Space Agency
Eureka	Fraunhofer

Associations

ACM Europe Policy Committee	European University Association (EUA)
ATTRACT	Photonics21
The Association of Commonwealth Universities (ACU)	The Guild

Group members

Trimbos Instituut	Tataj Innovation
Simmons & Simmons	GEANT
Deusto International Research School (DIRS)	ICHOM - International Consortium for Health Outcomes Measurements
Sant Joan de Deu - Barcelona Hospital	