D12.1 Procurement requirements and demand assessment

Lead Partner:	EGI Foundation
Version:	1
Status:	Approved by EC
Dissemination Level:	Public
Document Link:	https://documents.egi.eu/document/3466

Deliverable Abstract
This document presents a demand-side market research to understand the need for and level of demand of digital services for research, the manner in which such needs and demand are currently satisfied, the challenges presently being faced in respect of analysis workflows, data management and related infrastructure and services. The document also describes a business model analysis in the EOSC context, with a specific focus on proposing mechanisms for acquiring digital services in a business-to-business type of relationship. The document closes with a number of key recommendations for different types of stakeholders and draws the next steps for the work.

EOSC-HUB RECEIVES FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT NO. 777536.

COPYRIGHT NOTICE

DELIVERY SLIP

Date	*Name*	*Partner/Activity*	*Date*
From:	Sergio Andreozzi Hilary Goodson Bob Jones Anita Bens Christophe Veys Michelle Williams Phil Matthews Daniel Wustemberg Daniel Mallmann Hylke Koers Angele Giuliano Annamari Magri Pantea	EGI Foundation EGI Foundation CERN CERN CERN GEANT GEANT GEANT JUELICH SURFsara AcrossLimits AcrossLimits	26/06/2019
Moderated by:	Małgorzata Krakowian	EGI Foundation	
Reviewed by:	Tiina Kupila-Rantala Dale Robertson	CSC, WP2 JISC, T2.3	
Approved by:	AMB		26/06/2019

DOCUMENT LOG

Issue	*Date*	*Comment*	*Author*
v.0.1	16/03/2018	Table of content – first version	EGI Foundation
v. 0.2	08/05/2018	First draft version of demand-side market analysis	EGI Foundation
v. 0.3	10/07/2018	Second draft version of demand-side market analysis and related methodology	AcrossLimits
v. 0.4	02/10/2018	Agreed methodology for demand-side market analysis	AcrossLimits based on WP12 partners feedback
v. 0.5	21/02/2019	Initial results from market analysis	AcrossLimits based on WP12 partners feedback
v. 0.6	28/03/2019	Improved table of content based on face-to-face meeting outcome	CERN, EGI, FZJ, SURFsara, AcrossLimits
v. 0.7	08/04/2019	Findings section	AcrossLimits
v. 0.8	09/04/2019	First draft of Insight sections	CERN, EGI, FZJ, SURFsara
v. 0.9	12/04/2019	First draft of recommendations, lessons learnt, and business models based on the face-to-face meeting	AcrossLimits
v. 0.10	30/04/2019	Refactored section on business models	GEANT

v. 0.11	23/05/2019	Various improvements and additions for a complete document ready for external review	CERN, EGI, FZJ, SURFsara, GEANT, AcrossLimits
v. 0.12	03/06/2019	External review received	
v. 0.13	19/06/2019	Added shorter summary for findings and methodology and moved the main text in the annex	AcrossLimits
v. 1	26/06/2019	Final changes to address the comments from the external review	CERN, EGI, GEANT, SURFsara

TERMINOLOGY

https://wiki.eosc-hub.eu/display/EOSC/EOSC-hub+Glossary

Terminology/Acronym	***Definition***
ARCHIVER	Archiving and Preservation for Research Environments project, funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 824516.
AWS	Amazon Web Services
DI4R	Digital Infrastructures for Research Conference
EIRO	
ESCAPE	European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures (ESCAPE) project
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findable, Accessible, Interoperable and Reusable
GDPR	Regulation (EU) 2016/679 (General Data Protection Regulation)
GEANT	Through interconnections with its 38 national research and education network (NREN) partners, the GÉANT network is the largest and most advanced R&E network in the world
H2020	Horizon 2020 is the financial instrument implementing the Innovation Union, a Europe 2020 flagship initiative aimed at securing Europe's global competitiveness.
HPC	High Performance Computing
HTC	High Throughput Computing
IaaS	Infrastructure as a Service
LOFAR	The Low-Frequency Array or LOFAR, is a large radio telescope network located mainly in the Netherlands, completed in 2012 by ASTRON, the Netherlands Institute for Radio Astronomy and its international partners, and operated by ASTRON's radio observatory, of the Netherlands Organisation for Scientific Research
NREN	National Research and Education Network
OCRE	Open Clouds for Research Environments project, funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 824079
PaaS	Platform as a Service
PaNOSC	Photon and Neutron Open Science Cloud, is one of five cluster projects funded under the European H2020 programme
SaaS	Software as a Service
SLA	Service Level Agreement
SSO	Single Sign On

Contents

List of Figures

List of Tables

Executive summary

The EOSC-hub project (Grant Agreement ID: 777536) has the mission of bridging major European research infrastructures and the core e-infrastructure community to develop a common catalogue of data, services and software for research. It also aims to create the first implementation of the EOSC Hub, a European-level delivery channel connecting the demand-side (the EOSC Customers) and the supply-side (the EOSC Providers) to allow researchers to conduct their work in a collaborative, open and cost-efficient way. The goal of EOSC-hub Work Package 12 (WP12) is to contribute to the design of future business models and procurement frameworks for acquiring digital services from both publicly funded and commercial providers. The activity is split into phases, and this document presents the results of the first phase, structured in three main parts.

Demand-side market research: the first part consisted of a demand-side market research to understand the need for and level of demand of digital services for research in the context of the EOSC; understand the manner in which such needs and demand are currently satisfied; understand the challenges presently being faced in respect of analysis workflows, data management and related infrastructure and services; identify current and preferred delivery models for such services; identify funding streams and procurement strategies (e.g., drivers, constraints approaches); identify areas of improvements for business models. To address these questions, the task conducted market research consisting of in-person interviews (N=32), an online survey (N=33) targeting a broad range of roles as well as different types of organisations.

The market research confirmed that there is a growing demand of digital services for research. From all the services that are being used by researchers, the most popular are data repositories and data registries, followed by a large variety of other services which are also currently being used (e.g. analytics, data management tools, collaboration services and general computational services).

Among the challenges of using digital services, researchers raised ease of use and ease of access as main areas for improvement (e.g., availability of a single sign-on facility). Also, the configuration of digital services to be used with own data sources was a major concern, which indicates challenges with the interoperability and integration of services.

Researchers also expect e-infrastructures to support data management planning with trusted repositories, provide collaboration tools, and offer training and support, when necessary. Further, several interviewees stated that they are missing a complete EOSC data and service catalogue, and that there is a low uptake of FAIR principles due to lack of effort, explicit funding and support for using tools. Finally, researchers expect access to e-infrastructures to be free at the point of use.

On the IT support side, a high-level of FAIR readiness was reported, which indicates a gap between the perception of FAIR readiness from the providers' and users' points of view. Some service providers reported more demand than they could handle, thus indicating an urgent need for adequate business models and financial incentives to expand services. They also reported that most of their services are based on a hybrid of local, national and international resources.

As for the motivation to use e-Infrastructures, the main reasons are the need for federation of resources, specifically commodity compute-at-scale and long-term simple storage. On the other side, reasons for not outsourcing are internal skills development, retention of control (especially

over the full lifecycle), survival during financial squeezes, risk reduction (buyout, shut down, etc.), flexibility to integrate into the researchers' workflows, as well as low latency and high bandwidth network requirements.

With regards to the delivery models, most of the expected growth is for IaaS, followed by PaaS and SaaS. The usage patterns range from months to decades. With regards to the sourcing process, there was not a single restriction that stood out: respondents report a broad range of conditions that limit their choice of supplier, including geographical ones and certification requirements (e.g., ISO or CoreTrustSeal). Other factors that were mentioned include reliability, availability and quality of the service; integration with the researchers' existing workflows (requiring flexibility and a partnership approach); and open source / open standards to prevent vendor lock-in.

For users, the important factors that would make the EOSC services attractive are a short lead time from order to access, meaning that it is important for the acquisition process to be efficient and transparent in the way that the resources are allocated; ease of use; trust that data will be secure, even if the data are stored off-site. It should be noted that other projects have articulated the need to ensure data can be repatriated easily, and this should be included as a requirement.

Business model analysis: the second part of the task consisted of a business model analysis in the EOSC context, with a specific focus on proposing mechanisms for acquiring digital services, based on the use cases where a researcher may look to acquire services 'on demand', a departmental manager may look to acquire services for its constituent researchers to use, and where resources can be procured in 'bulk'. The assumption in the analysis was that the customer/user always operates in a business-to-business (B2B) type of relationship and not as a consumer (as defined by the EU Consumer Rights Directive) for personal purposes. This assumption is important, as it rules out the need to comply with consumer rights regulation.

The business model analysis identified three main models that will be considered for the second phase of WP12 activities. The first model is called 'Cloud Coin' or Voucher Access Model and applies when an individual researcher or a small research group needs limited-scale access to commercial services on an ad hoc basis. This model foresees that users will receive services free at the point of use while service providers are financially rewarded indirectly by funders. The second model is called 'sponsored use' where institutional usage would be fulfilled via specific purchase-supply contracts between the service provider and institution. It analyses the explicit terms that the EOSC Portal would need to capture to address any issues concerning contract privity, EU Procurement Directives compliance and other statutory legislative/regulatory implications of such a business relationship. The third business model is called 'aggregated procurement' and aims to support demand aggregation and joint purchasing to collectively exploit existing pan-European mechanisms and remove the need for repeated and costly procurement processes.

Recommendations and Next Steps: the third part of this work highlights a number of recommendations for different stakeholder categories and defines the focus for the second phase of WP12 to be carried forward by Task 12.1. The main work will be to further develop the identified business models and consider public-to-public, private-to-public and public-to-private scenarios also in relationship to the wider funding environment in order to identify opportunities to enable a virtuous EOSC ecosystem.

1 Introduction

The European Open Science Cloud (EOSC) initiative aims at creating a fit for purpose pan-European federation of digital services for research and move from the current fragmentation, to a situation where data is easy to store, find, process, share and re-use. As such, the EOSC will be a fundamental enabler of Open Science and of the digital transformation of science, as well as for "Building the European Data Economy" initiative[1].

EOSC aims to support research and innovation by offering the possibility to access, reuse and further analyse (including through HPC and HTC services) all publicly funded research data in Europe. This access should operate across disciplines and borders; it should also operate throughout the value chain, delivering scientific, societal and industrial benefits.

To this end, EOSC is being developed as a pan-European federation of digital infrastructures built around a set of core services, governed by statutes developed in partnership with its users, which will provide access to a wide range of publicly funded services supplied at international, national, regional and institutional levels. Complementary commercial services will also be added.

Establishing the conditions for such an ecosystem is a challenging goal and requires a proper understanding of the needs of the prospective user segments and the way those needs are currently satisfied. It also requires exploring what are the current challenges in delivering digital services for research and the related funding streams and procurement strategies, including the constraints under which the current ecosystem operates. Finally, it also requires clarity on the possible opportunities and desired areas of improvement for the overall business models.

In this context, the EOSC-hub project (Grant Agreement ID: 777536) started in January 2018 with the mission of bridging between major European research infrastructures and the core e-infrastructure community to begin the development of a common catalogue of data, services and software for research. The project is expected to provide also a European-level delivery channel connecting the demand-side (the EOSC Customers) and the supply-side (the EOSC Providers) to allow researchers to conduct their work in a collaborative, open and cost-efficient way. This is made up of several key initial elements for EOSC [R1, R2]: the EOSC portal, where the researchers come and find services, either thematic or horizontal; the Federating Core, that is the part of EOSC which allows the research targeted services for the EOSC Portfolio to operate; this is supported by two components: 1) shared resources, access to central research enabling services such as data sets of broad interest, centralised repositories, data processing facilities or other key resources; 2) compliance framework, to include the rules for participation, how to meet FAIR data principles, how services can be integrated, address security and privacy and include governance practices and risk controls. The procurement and purchasing models can be seen as possible component of this compliance framework, unless captured in a business layer to be introduced.

The EOSC-hub Work Package 12 (WP12) was designed with the goal to contribute to future business models design and procurement frameworks underpinning the mechanisms for acquiring digital

[1] https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy

services from both publicly funded and commercial providers. The activity was split in three main phases.

The first phase covered by the WP12 task 12.1 is the analysis and includes a demand-side market research to understand the requirements from different stakeholder groups that will be prospective customers and users of EOSC services, including current service delivery models, procurement strategies and constraints and funding models. This phase also includes the analysis and identification of business models underpinning the procurement of digital services.

The second phase covered by the WP12 task 12.2 focuses on developing the identified business models by also considering the legal aspects and agreement framework. This phase will also focus on strengthening the collaboration with the new-born initiatives that are relevant to this work, like the procurement related projects OCRE and ARCHIVER, or the EOSC governance bodies and related working groups and supporting the EOSC Secretariat project.

The third phase will perform an evaluation of the developed structures and will draw a roadmap for activities to be performed beyond the life of the project.

This report focuses on describing the results of the first phase and is structured in three main parts. The first part presents the demand-side market research describing the proposed methodology, how this was implemented, the findings from the data collection phase and the insights that could be derived. Based on the learning from the market research and from the existing knowledge, the second part provides a business model analysis focusing on aspects of procuring services in the context of the EOSC. Finally, the third part, highlights recommendations for different stakeholders including opportunities for future activities in the context of the EOSC-hub project.

2 Demand-side market research

This section summarises the methodology, timeline and findings from the demand-side market research and details the insights we developed as a result of conducting the market survey. During the first part of this work, the task partners agreed to engage an experienced market research consultancy to assist in the work of scoping, implementing and analysing the results.

2.1 Introductory Summary

We quickly discovered that there is very little published information on the current level of demand for, and spending on, e-infrastructures and FAIR data management in the EU Member States. Therefore, we needed to develop for ourselves a detailed understanding of the needs and challenges faced by research organisations and communities regarding:

- Their sourcing strategy for e-infrastructures and services.
- How associated issues could be better addressed in the context of the EOSC.

We determined, therefore, that the most appropriate approach to undertaking a demand-side market analysis was one driven by primary research.

With the objective of gaining a detailed understanding of e-infrastructure and service needs of users in the European research community, our primary research began with the development of a methodology (see Appendix I). The methodology commenced with desk research of the little information that currently exists, in order to develop a baseline understanding. From this baseline understanding we developed an initial set of questions during a workshop that included members of the consortium and contracted experts. These initial questions were designed to fill in the knowledge gaps we had identified in the desk analysis. Using this initial set of questions, we tested our base line understanding during preliminary interviews with notable members of EU research communities. The results of these interviews were used to update our understanding and to revise the question bank, which was carried out during a second workshop. All participants contributed to the development and agreed the finalisation of the question bank. A timeline (see Appendix II) of the actions required to achieve our objective was then mapped out and an associated plan was then devised and executed.

Notably, two rounds of user community engagement were carried out. First of all, a set of face to face interviews was conducted, this was then supplemented by a public survey. Both engagements used the same set of questions. Beyond this, the results were collated, aggregated and analysed to expose the findings that emerged from the user community engagement (see Appendix III). From these findings, the insights that follow were developed.

2.2 Insights

2.2.1 Current usage and foreseen demand of digital services for researchers

There is clear demand for digital services from researchers, as can be seen from the responses that focused on demand for different types of digital services in both the interviews and the online survey. The answers to questions relating to the assessment of current demand, demonstrate that

it is the current increase in service usage which is driving procurement at the moment; therefore, looking at current usage is important in understanding how demand for existing services is currently being satisfied in order to prepare the initial EOSC offerings to match and supersede them.

From all the services that are being used by researchers, it seems that the most popular are repositories with just over 80% of usage. Registries are quite popular (above 60% overall), followed by a large variety of other services which are also currently being used; these include web server analytics, data management tools, search services, collaboration services and general computational services.

From the answers obtained there is also some evidence of unmet demand and future needs. Half of the respondents (from both interviews and surveys) have explained that their current e-infrastructure was adequate for the foreseeable future; however, the other half of the respondents were either uncertain that this is the case, or they clearly feel that they already need additional services. Cloud services were definitely recognised as an acceptable or attractive addition to the existing services by the great majority of those surveyed. One can therefore confirm that there is a clear present and future demand for cloud services to be provided to researchers.

2.2.2 Current challenges in using digital services by researchers

The interviews and the online survey identified major challenges in using e-infrastructures. In particular:

- Researchers expect digital services to be easy to use and easy to access. This includes the availability of single sign-on and that security and privacy requirements are taken care of.
- The configuration of digital services for their own data sources was a major concern expressed in several interviews, which indicates challenges with the interoperability and integration of services.
- Access to e-infrastructures is expected to be free at the point of use.
- Researchers expect that e-infrastructures will support them in data management planning, that collaboration tools are available, repositories are trusted, and that e-infrastructures offer support and training where necessary.
- Several interviewees stated that they are missing a complete EOSC data and service catalogue, and that this poses a challenge to them.

2.2.3 Current challenges in offering digital services to researchers

A strong theme from the responses centred on a lack of guidance regarding procurement of future services in relation to the FAIR principles. Particularly, that these principles are an underlying philosophical and policy base for EOSC; but how these principles are practically translated into data requirements, and how services are able to capture and implement these in delivery, is not understood. Where it was understood by interview participants, it was in regards to specific components (i.e., a provider had made their data findable through open publication of datasets), and not understood in terms of how and to what extent this would be needed to inform future procurement of services, particularly as it related to future funding of projects.

The fact that the FAIR principles are not a common practice yet is due to numerous reasons. Some concern the lack of awareness in the research community about how to share data, in which format, what information or metadata should be provided etc. Others touch upon existing cultures and behaviours in conducting research, from a lack of incentive for researchers to share their data to researchers not even considering that the data they produce can be valuable for others, the lack of attention given to the preparation of a data management plan, missing metadata, various competing standards for research data and metadata, and the lack of persistent identifiers for data, datasets and metadata. Moreover, many researchers and organisations anecdotally in response showed reluctance to apply the FAIR principles and share their datasets because of real or perceived costs, including time investment and money. Hence, challenges arose over issues of FAIR data and FAIRification in the context of procuring services. All respondents in the interviews conducted, cited awareness of FAIR. However, its importance was varied. Some respondents stated that FAIR was not a significant consideration when assessing procurement needs, and this was due to lack of practical guidelines as to what it means in application.

Anecdotally, a specific definition of what FAIR looks like in today's practice did not exist. This demonstrates all too frequently a need for provision of practical guidelines. Knowledge of FAIR in practice ranged from uncertain to mature. There is concern over a lack of definition and guidelines for implementation. The majority of respondents were not implementing FAIR principles. The main issues were cost, and lack of centralised data management plans. These two responses were recognised as interconnected by those who were interviewed. The need to preserve data was recognised as central to FAIR, where several respondents cited reusability as key. However, it is felt that preservation is not funded by the EC, and the priorities of the researcher are more concerned with compute and analysis needed for their work. This is reflected in the large number of responses citing facilities expecting users to import data but export results back to localised storage. In this sense, FAIRification is seen as a 'post-project' consideration. Outside of the larger entities, data access is seen as under the control of the researcher. This presents a challenge as the practice points to being 'bottom up' whilst the policies on FAIR and open science are top down. There is a clear recognition of the importance of FAIR and the potential benefits, particularly strong across the Humanities and Social Sciences, but a lack of simplicity in implementation, and in funding, to implement from the beginning of the research lifecycle. Long term domain agnostic data storage was demanded, and training was also requested by many communities. An exception is the medical and life sciences community where there is a history of needing to discover and reuse, and services have been procured and architected to support this in complex iterations.

The demand for FAIR, therefore, needs much more definition. Respondents also noted the domain and research-discipline specifics of FAIR, most notably with access, and the need to understand openness within the context of other legislative demands (such as privacy with health data; provenance with biodiversity data etc). It was noted as a valuable concept, by several of the respondents, that the Rules of Participation have domain-specific recommendations.

FAIR was a 'stretch goal' - currently 'always on' was identified as the main need of the user as 'most needed'. This was closely followed with ease of use, access and configuration, coming as a close second priority. Other factors include the availability of SSO, security and privacy requirements taken care of, access to training and the availability of collaboration tools. This highlights the gap in

demands needed by researchers; more immediate, project relevance; and the challenge of service providers offering services that satisfy FAIR policy demands. As far as expectations are concerned, the users (researchers) of e-infrastructure expect Data management plans issues to be taken care of and for repositories to be trusted, only in one case was data-pipeline design and deployment seen as more important than data management; the primary concerns were use- and configuration-driven.

When asked to further specify services required by users, a slight majority (approx. 65%) considered automated processing to be the priority, other factors were evenly demanded at between 35%-40%. Other desired attributes were data and data as content. Some communities also wanted data analysis as a service. A perspective not yet considered for EOSC is that of allowing research leaders to access data that illustrates how their teams use the services, which assumes that a set of instruments are put in place to measure such parameters. All of this must be made available at an affordable price. The new services should be easy to adopt and not result in changes to existing workflows.

Users' main needs of their e-infrastructure were less evenly distributed in comparison with the focus group surveys. The ranking of "does not require ICT expertise" and "help when needed" as qualities of high importance, demonstrated the significant expectations of the users of e-infrastructure to have data management and processing issues to be taken care of, and for the repositories to be trusted.

Considering the important services required by users, everyone (100%) considered free (interpreted as *free at the point-of-use*) and easy access to e-infrastructure services to be the priority. The next highest demand was for easy integration of legacy data (62.5%). The three other factors declined from 37.5% (automated processing [focus group interview highest-scored]) to 12.5% (cross-disciplinary FAIRness) in popularity. Other desired attributes were: Big Data compatibility and metadata integration. There is some interest (50%) in accessing other data services via e-infrastructure but no further details were given.

Some notable discrepancies emerge when asked about ease of use and required expertise. Almost all interviewed participants noted ease of use as important to services. However, this was contradicted further in the interview set, when discussing the main needs of the user base vis-à-vis research e-infrastructure, in response to 'ICT expertise required', 80% answered with 'not important'. Help at hand, however, was 'Essential/desirable.' There is a likely assumption here that the EOSC will provide support for implementation. This makes sense when the EOSC is understood as a single market to aggregate services and research lifecycle needs. It is apparent that it is assumed EOSC could or should offer a degree of overarching technical management. Technical and operational management of services contracted across multiple entities for components.

Other main challenges in offering digital services to researchers become visible as the data exposes motivations for procurement. We see that on average, in 3 out of 4 cases external procurement is sought when the project is aggregated, i.e., involving multiple researchers, consortiums, users, and/or instruments. This is almost always carried out at disciplinary and interdisciplinary levels. Each disciplinary or thematic project has its own mechanism for aggregating funding from partners and allocating use of whichever services had been procured, based on its own rules and frameworks.

This adds a layer of complexity in determining a singular procurement model that is fit for purpose in the most general sense across Europe. Consideration needs to be given to the funding mechanisms and cycles specific to each country, as well as the big European-led funding programmes.

2.2.4 Current and preferred delivery models of digital services for researchers

In terms of current delivery models, IaaS (+58%) was the most heavily procured type of services followed by SaaS (+45%) while PaaS is rarely considered (-70%[2]). This preference for IaaS is probably influenced by the majority of interviewees and survey respondents coming from well-structured research communities where IT expertise is readily available to manage and operate low-level services. The same outcome was broadly repeated, though with a slight increase of SaaS and PaaS to the detriment of IaaS, when it came to which delivery models would be chosen in the future, with most looking to acquire these new services within the coming year, if the procurement process was not already underway.

Researchers had more urgent and short-term needs while service operators were looking further ahead and wanted to establish longer-term relationships with service providers. Presumably this is because end users are focussed on the immediate needs of their current research activities while service operators prefer stable relationships with reliable service providers.

In terms of external service providers, respondents are using a mixture of public sector e-infrastructures (European Commission funded initiatives as well as national providers and research community structures federated to support ESFRIs) and commercial market leaders (such as Amazon Web Service, Microsoft Azure and Google Cloud Platform). There may be a lack of understanding of how many external providers are involved in supporting particular services. This situation is not unusual for IT services where the lower-layers should be invisible to the end-user.

External service providers are used to supplement in-house service provisioning which remains popular and a mixture of these delivery models indicates some level of flexibility and experimentation to satisfy demand. The responses also indicated a sense of frustration where service providers recognise that they are not fulfilling user demand, but their scope for action is constrained by resources and procurement practises.

2.2.5 Funding streams and procurement strategies

2.2.5.1 Funding streams and access policies

Public funding streams, either nationally or internationally, are the predominant funding stream for the organisations that were interviewed (92%). For many, this form of funding is supplemented by

[2] The negative percentage describes a compounded negative response. Across the three related questions over half of respondents **did** procure IaaS, over two thirds **did not** procure PaaS (this is not the same as +30%) and nearly half **did** procure SaaS. The three questions were structured as Yes/No to each type and combining the responses any other way made less sense than this method. -70% was chosen because an active decision was made by respondents to vote "no", it does not arise out of the passive result of no one voting for this selection.

national or EU funds obtained through a competitive process (66%). Private donations or other sources of revenue, on the other hand, appear to be an exception.

The most common model for organisations to provide access to their services is membership-by-affiliation - i.e., anyone affiliated with the organisation is granted access (56%) - while a substantial amount of institutes also provide open and free access for all (26%), or have some form of excellence-driven selection based on peer-review (21%).

2.2.5.2 Why procure?

While about half (52%) of the respondents believe that their current e-infrastructure will meet their needs for the foreseeable future, there are also many who believe that is not the case (22%) or who are unsure (26%) about this.

This naturally raises a question on how to best ensure that the needs of researchers continue to be met in the future, which includes a consideration to extend and develop in-house solutions or rather procure services externally. The main drivers to outsource e-infrastructure services are seen to be flexibility (including but not limited to scaling up/down) and cost-efficiency (32%); followed by (temporarily) insufficient capability or capacity in their own organisation (28%). Funding constraints and accounting pressures (e.g. CapEx vs. OpEx) are seen to play a much more limited role, as may also be appreciated in the figure below.

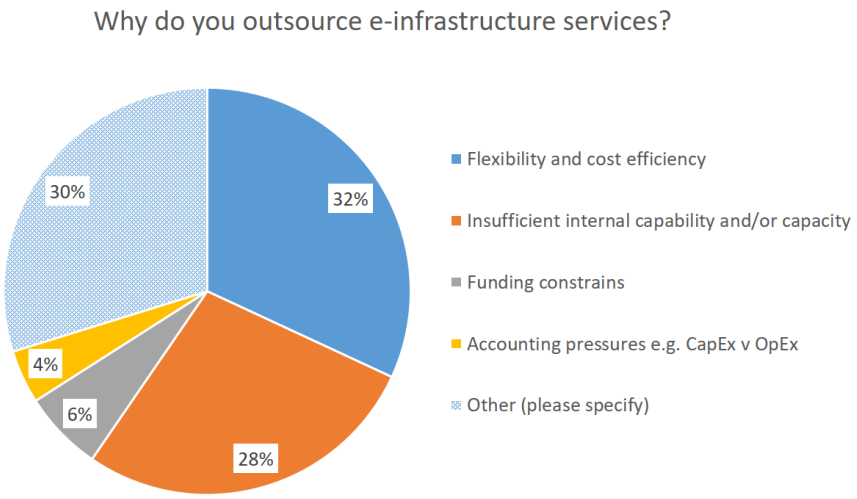


Figure 1: Motivations to outsource e-infrastructure services

In addition to the above, some respondents expect that the services provided by a 3rd party will provide more functionality and enable them to benefit from innovations more quickly.

Interestingly, flexibility and integration into the local workflow was also mentioned by some as a reason to develop and maintain services in-house rather than procure them externally. Other reasons for *not* wanting to procure are a desire to develop expertise in-house to be able to support

researchers better; and a desire for long-term control over the service to nullify the risk of a service being discontinued or the service provider being acquired by a non-EU business.

2.2.5.3 Scope of procurement

Type of services

The majority (71%) of respondents are currently procuring some of their e-infrastructure services externally. For those that are currently procuring services, the medium number of external providers is two, but that number can rise significantly: three organisations (out of 22 answering this question) reported 5+ external providers.

Services in scope for procurement span a broad range of functionality. Some commonly named challenges that organisations hope to solve with the services they procure are federation and/or distribution of data sources; long-term preservation; and computing at scale. Other challenges are listed in the diagram below.

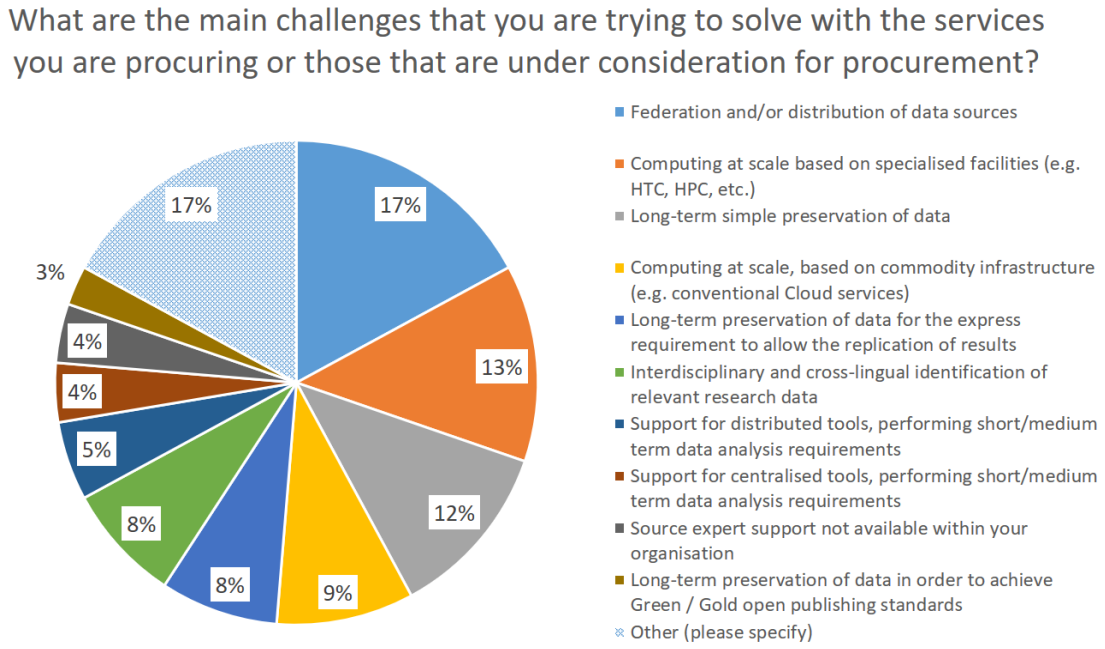


Figure 2: Challenges driving procurement

Delivery model

With those organisations already procuring cloud services, IaaS is the most commonly used (56%), followed by SaaS (39%). Interestingly, the PaaS model appears to not be commonly used amongst the respondents at present (selected by just 1 out of 18 respondents).

2.2.5.4 Current procurement practices

Type of contract or agreement

The most commonly used types of contracts are a framework contract / agreement (36%) - with terms and conditions agreed with the selected suppliers for the duration of the agreement - and a multi-annual contract (21%). This would seem to indicate a preference for longer-term commitments with well-defined conditions and minimal risk. Framework contracts/agreements where price is determined by a "mini-competition" between the selected suppliers, or single-task contracts were mentioned by one of the 14 respondents for this question, and then in the context of hardware procurement. The GÉANT framework is mentioned by three respondents as a useful tool to provide guidance and simplify the procurement process.

Payment models

Current payment models vary considerably. The most commonly selected model was invoicing for bundled services (27%) and fixed periodic payments (27%), while casual and ad-hoc payments were also mentioned (each 13%). Finally, there are some early indications of performance-based models being discussed or adopted, but this is clearly not yet a common practice.

Procurement procedures

Various procurement procedures are used without a single option clearly standing out. The reported procedures include open, restricted, negotiated, competitive dialogue, tendering, and other methods. The GÉANT framework and institute-specific procurement rules were also mentioned (by one respondent each).

Challenges

All but one respondent (9 out of 10) find the existing fragmentation of the research data service and e-infrastructure landscape to be a problem for their organisation. What this means at a practical level is that a lot of time and effort needs to be put into identifying and selecting the right provider for a desired service. Apart from just investing a lot of time in this process, current strategies to manage this challenge include collaboration and, for one, working with a consultancy company to help select the right service provider.

Another challenge that is reflected in several comments is about the need to match *a priori* independent timelines between research projects, data preservation periods, budget cycles, and service procurement. For example: a research project may be funded for three years; the data that is produced in the project needs to be preserved for 10 years; the IT support organisation has yearly budget allocation; and they have contractual agreements spanning a 5-year window with 3rd-party service providers. It is clearly a challenge to manage these different timelines in such a way that guarantees uninterrupted service to researchers while maintaining cost-effectiveness.

2.2.5.5 *Procurement strategies; preferences, constraints and other considerations going into the procurement process*

Delivery Model

Cloud-based solutions are widely seen as a viable option to meet current and future e-infrastructure needs (62% "yes" and an additional 33% "maybe"). IaaS is seen as the most attractive delivery model (73%), followed by PaaS (50%) and SaaS (42%). The choice is seen to be driven by the needs and maturity of users - with SaaS models especially suited for the "long tail" of researchers that are looking for an easy and readily usable solution, whereas "power users" and research communities that are used to developing bespoke software themselves are seen to benefit more from IaaS-type services. Given this heterogeneity of the user base, it may be expected that multiple delivery models will continue to coexist in the future.

An interesting point to note is that, while respondents rank the attractiveness of PaaS models to be in between IaaS and SaaS, it is actually not reported to be used much in today's practice.

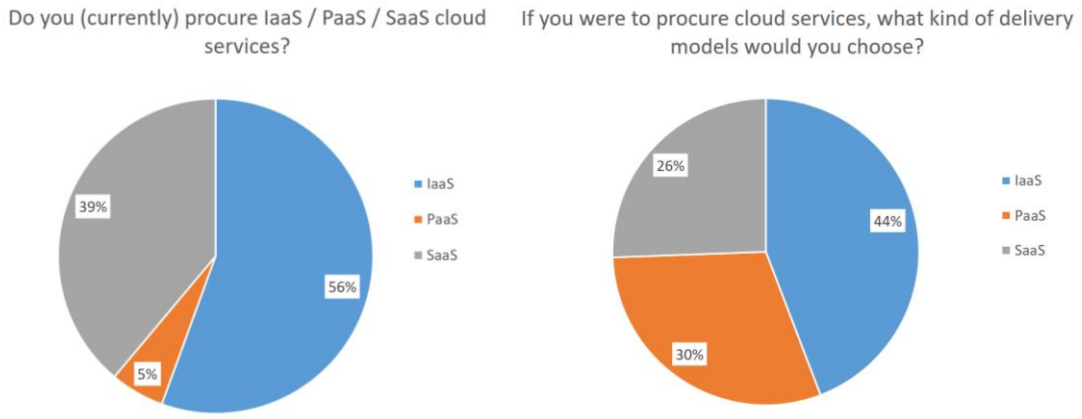


Figure 3: Current (right) and preferred (left) cloud delivery models

Financial considerations

In terms of costs, purchase, migration, running, maintenance, and exit costs are all taken into consideration in the procurement process. Life-cycle costing is often considered when conducting procurement of e-services by a majority of respondents (64%).

Duration

Longer-term partnerships are clearly preferred (63%); predominantly because this is seen as to allow for the design/development and support of a customised solution, but also to sustain an SLA for a defined service offer for many years. In addition to this, there is a smaller but non-negligible demand (21%) for short-to-medium solutions to support a specific project.

Rules and regulations

A large majority of respondents (10 out of 12 answering this question) indicates that their current procurement rules allow them to buy their own e-infrastructure and other digital services. These existing rules include both national and European regulations including but not limited to tendering

and GDPR. The tender process, which under some conditions is mandatory, is seen by many respondents (who, it should be noted are often not procurement professionals) as onerous and time-consuming.

Other conditions and restrictions

While there is not a single restriction that stands out, respondents report a broad range of conditions that limit their choice of supplier, including geographical ones and certification requirements (e.g., ISO or CoreTrustSeal). Other factors that are mentioned include reliability, availability and quality of the service; integration with the researchers' existing workflows (requiring flexibility and a partnership approach); and open source / open standards to prevent vendor lock-in.

With regard to research data, respondents mention both guarantees for data safety (e.g., to prevent that data is shared with competitors when commercial entities are involved) as well as requirements to make the data openly available (to secure long-term availability of the data to the public and to decision-makers).

2.2.5.6 Areas of improvements for business models

From the market research, various factors could be identified that should be considered when designing the future business models underpinning the EOSC. The focus of the interviews and surveys carried out attempted to learn about those organising principles that closely relate to the procurement and purchasing of digital services for research, but limited information was offered by the respondents.

Research organisations often act also as digital service providers towards their research community and may fulfil the need for services either by organising in-house capabilities or by outsourcing the service provision to supplier organisations. Such supplier organisations can be closely linked to the research organisation (e.g., via a federated governance model) or they can come from the free market (depending on the type of service and constraints). Such research organisations, therefore, act as buyers on behalf of the served community. In the limited responses received on the topic of procurement from organisations that consider themselves to be service providers, common amongst the responses was the interest in accessing mechanisms that allow achievement of economies of scale in the purchasing process. Also, it emerged that developing long-term relationships with suppliers is an important factor.

For users, the important factors that would make EOSC services attractive are: 1) short lead time from order to access, meaning that it is important for the order process to be efficient and transparent in the way that resources are allocated; 2) ease of use; 3) trust that data will be secure, even if the data are stored off-site. It should be noted that other project activities have articulated the need to ensure data can be repatriated easily, and this should be included as a requirement.

Concerning the role of an intermediary organisation, when inquiring about how EOSC-hub could better facilitate or support the procurement process, the following aspects were mentioned:

- Expert support in selecting digital services for research
- Make procurement simpler, more cost transparent and faster (especially for compute)

- Support demand aggregation (unlikely to engage in a procurement process in isolation)
- Validate that providers meet pre-defined requirements, e.g. from research funders (this could link to the evaluation of meeting the Rules of Participation for the EOSC)

2.2.6 Learning from co-hosted World Cafe session

During the 2018 edition of the Digital Infrastructure for Research Conference, the projects EOSC-hub and EOSCpilot co-hosted a session titled 'Shaping the EOSC service roadmap: what users need'[3] with a view to extracting high-level requirements for services and their prioritisation from the users attending the conference. In order to maximise opportunities for audience participation a 'World cafe[4]' format was selected, where small, facilitated discussion groups were formed and asked to respond to specific questions, the output of which were collected and summarised in a live questionnaire tool. The questions asked are as follows:

- Question to understand the audience (answered by the moderator only):
    - How is your group composed? (i.e., researchers/users, service providers, service developers)
    - What is the role of each person in the group?
- Questions for response by the group:
    - What research data do you find difficult to discover today?
    - What capabilities should EOSC provide to support your research workflow?
    - What are the top 2 wishes that the EOSC could fulfil over the next year?

The collective response called for improved accessibility and discoverability of services, rather than for specific services that could be procured as part of the EOSC Hub Work Package 12 activities. The session found that there are numerous barriers to discovery relating to various topics, such as:

- Some disciplines do not have indexes of data/data repositories
- Completion of metadata, versioning or documentation to assist reproducibility is difficult to enforce and as a result can be difficult to find and use even within an organisation.
- Metadata 'standards' are diverse and discipline-specific
- Handling of large or fragmented datasets is difficult
- Lack of funding for long-term storage and archiving have limited FAIRness
- Data is often outside systems with single-sign-on facilities, making access and sharing more difficult.

The following themes were generated in response to the question "What are the top 2 wishes that the EOSC could fulfil over the next year?" that focused more on what the EOSC itself should be, rather than a demand for any specific procurable services that it should contain:

[3] https://indico.egi.eu/indico/event/3973/session/27/?slotId=0#20181010
[4] http://www.theworldcafe.com/wp-content/uploads/2015/07/Cafe-To-Go-Revised.pdf

- Identity & security
    - Single sign-on
    - Service catalogue (at least a medium size list with single sign on for what is published)
- Services for data
    - Simple data access
    - Support data sharing (simplify metadata procedure)
    - More services around data sharing
    - Storage near computing facilities
- Services for computing, such as easy-to-use near data compute capability
- Services for network, such as on-demand networks (including the last mile)
- Business models should be sustainable and should promote:
    - Long-term service and resource availability
    - Simplified contracting frameworks with standardised legal agreements
- There should be a clear architectural design (which separates the generic layer from the domain specific layer) and there should be an easy way to integrate services
- There should be a value proposition targeted at each user group that includes a clear definition/description of the EOSC and its benefits, with a unique selling point that resonates nationally
- There should be a focus on the quality of onboarded service providers, including requirements for:
    - More training material and tutorials on how to use services
    - Adherence to well-defined and documented good practices
    - Simplified and accelerated application processes to request resources
- To support service provider onboarding it would be necessary to define and document the minimum set of criteria to qualify for service registration in the catalogue
- The Portal design should consider:
    - Easier discoverability of services/resource matching specific requirements
    - How to scale up (and out) services and resources provisioned via EOSC
    - Automated/machine-generated metadata (auto-tagging concerning funders, projects, ORCID)
    - Ability to report on patterns of user behaviour (which tools, services and data users at institutions have been accessing, using and contributing to)

3 Business Models in the EOSC context

This section of the deliverable will address the 'Value Capture' aspect in the area of mechanisms for acquiring digital services. From the market research, we can extract the following assumptions that will be considered in the context of the business model analysis: 1) researchers expect services free at the point of use while service providers expect financial compensation for opening up their offerings; 2) it would benefit the research community if procurement were to become easier, more cost transparent and faster; 3) capabilities of demand aggregation are a viable option to achieve economy of scale and to increase the shared understanding of community requirements.

The business model analysis will be based on the use cases where 1) a researcher may look to acquire services 'on demand', 2) a departmental manager may look to acquire services for his/her constituent researchers to use , and 3) where resources can be procured in 'bulk' and made available to the use cases in 1) and 2) above. The detailed analysis of these scenarios will be carried out in Task 12.2, but in order to ensure that the EOSC meets its objective to offer services 'free at the point of use' in a manner that allows sustainability and recovery of the costs incurred for the supply of resources and services, it is imperative that the analysis considers how it could provide for a transactional relationship with its customer base. In doing so, it must set out clear definitions of its vision of the concept of 'customer' and 'supplier'.

In order to contextualise the framework that will be necessary to deliver 'free at the point of use' services, it is important to set out the fundamental principles of offering products and services to customers. The term 'business model' is used to describe various strategic concepts concerning the manner in which an organisation operates and exploits its products, and how organisations create, deliver, and capture value. Investopedia describes the term as something that '*identifies the products or services the business will sell, the target market it has identified, and the expenses it anticipates*'[5], and in these terms, the EOSC business model is in part defined by its objective to [R1]:

"give the Union a global lead in research data management and ensure that European scientists reap the full benefits of data-driven science, by offering '1.7 million European researchers and 70 million professionals in science and technology a virtual environment with free at the point of use, open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines'."

In essence, the offered products or services are those of the many European service providers who, in turn, offer their services to the European research community; the target market is the researchers, and wider 'consumers' of the services. This market comprises European researchers looking for services to use in making their data FAIR, European researchers looking to re-use data in their own research, service providers that have relevant services to promote and offer, and service providers that could also act as consumers. The term 'service', in this context, also includes the data produced by the research community. The EOSC fills a gap where there was, until now, no single platform for researchers to use in order to discover, promote and consume data and services. The EOSC business model combines a Portal with a series of IT service management processes to deliver

[5] https://www.investopedia.com/terms/b/businessmodel.asp

value to its customers, and section 4.1 in [R2] sets out the various themes that the EOSC must address in the implementation of its business model.

3.1 The EOSC Market

It is important to further describe the EOSC market in order to illustrate the primary drivers of business and its wider environment, before then zoning in on the characteristics of the customers, and suppliers, within that market. In a traditional business model, the primary objective would be to generate sufficient sales (or attract sufficient investment) to cover costs incurred in creating and operating the product or service, as well as any overhead and R&D costs, by targeting a specific area of the market: the Oxford English Dictionary defines the term business model as "A plan for the successful operation of a business, identifying sources of revenue, the intended customer base, products, and details of financing."[6]. In creating its plan, an organisation would typically target either the 'business-to-business' (B2B) market or the 'consumer' (B2C) market, each having unique characteristics in its behaviours, expectations, and related mandatory laws and regulations. B2B transactions fall within the scope of corporate laws and corporate taxation[7] frameworks, while B2C transactions are within the scope of the EU Consumer Rights Directive 2011/83/EU which places additional statutory obligations on the organisations serving the consumer. A consumer is defined as "a natural person, who is acting outside the scope of an economic activity (trade, business, craft, liberal profession)." [R3]. Therefore, this distinction between B2B and B2C is imperative in order to frame the manner in which the target market of EOSC will interact with the EOSC Portal. For example, in all 'buy-sell' transactions there is a contract between the buyer and seller, whether written or implied, and the most common contractual relationship is the one that creates a relationship with the buyer and seller only, excluding any '3rd party relationship'. Contractual obligations apply regardless of whether there is a financial transaction between the parties, as contracts lay out fundamental conditions that consider more than just financial elements (such as limits of liability, termination terms, start dates, end dates, guarantees, etc).

The various aspects of the business model themes discussed in [R2] are being addressed by several Work Packages, as well as other projects and initiatives, and fall into the topics of 'Value Creation' and 'Delivery of Value' as outlined in Figure 4. The primary objective of WP12 'Business Models and Procurement' is to establish mechanisms for interested organisations to acquire digital services from either publicly-funded infrastructures or commercial providers, where Task 12.1 'Business Model Analysis' was targeted to 1) analyse the demand deriving from the research community, and 2) identify typical scenarios and propose appropriate business models.

[6] https://en.oxforddictionaries.com/definition/business_model

[7] It is important to note, however, that in the context of VAT regulations, *an organisation that is not VAT registered would be defined as B2C*, although still falling into the definition of a B2B transaction so far as the EU Consumer Rights Directive is concerned.

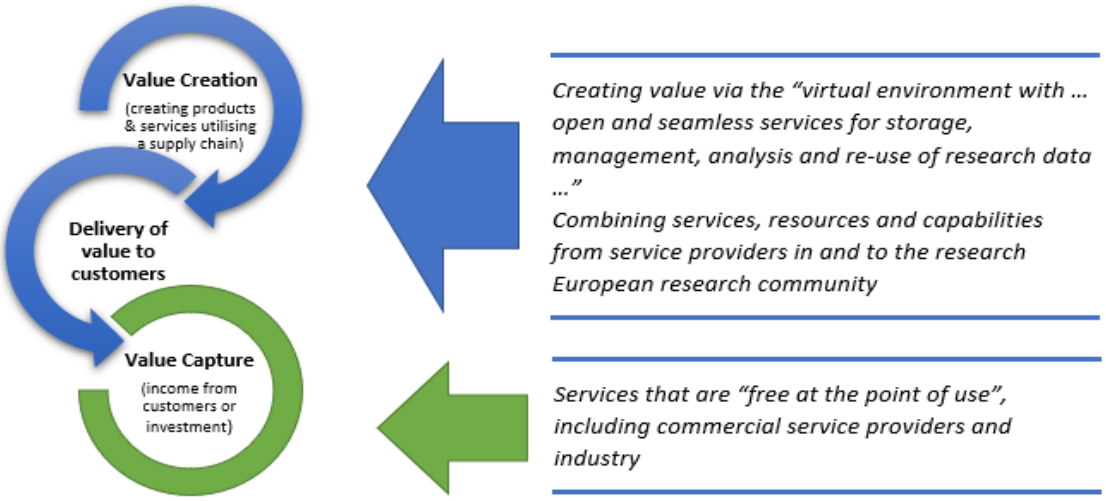


Figure 4: Value Transformation (value creation and value delivery are dealt with outside WP12)

3.2 Characteristics of EOSC customers

In order to describe the types of business transaction the EOSC must provide for, it is important to define the nature of the typical EOSC customer. In terms of service access/consumption, the EOSC customer will be the **EOSC End-Users**[8], and it has been stipulated that they should be given access to services that are 'free at the point of use'. A key feature of the Minimum Viable Ecosystem [R2] is that a possible End User action would be to either pay for a service or be sponsored to use a service as a solution to providing market-driven access. Therefore, confirming that the legal status of an EOSC End-User is that of a business-to-business end user, and not that of a business-to-consumer user is a fundamental and defining principle that influences the functionality that must be offered, and relates to the additional contexts set out below:

1. If intended for use by Consumers (as defined by the EC Consumer Rights Directive), services to be offered to EOSC End Users in the EOSC Portal would need to be procured with Consumers in mind; it is unlikely that public Central Purchasing Bodies would be authorised to procure pre-paid/ringfenced services for use by Consumers as a general principle.
2. In order to comply with the Consumer Rights Directive, specific functionality and additional business processes would need to be delivered by the EOSC Portal above and beyond those features and processes already planned.
3. In order to comply with EU-wide VAT regulations, it would be necessary to distinguish between VAT-registered and non-VAT registered organisations.

[8] "EOSC End-user Definition: an EOSC System User consuming EOSC Resource(s) by means of EOSC Service(s) (e.g. the EOSC Portal) to accomplish a task. She can be a Researcher, a Research Administrator, or a Third-party Service Provider". EOSC Glossary https://wiki.eosc-hub.eu/display/EOSC/Glossary

It is equally important to note that, whether an EOSC End-User is:

- buying resources 'on-demand' for a specified purpose for a one-off project,
- buying resources for a department or institution's researchers to use,
- a researcher or user at a not-for-profit organisation, or
- a researcher or user at a commercial organisation within industry,

in these cases the EOSC End-User will always contract under a business-to-business transaction (as far as the EU Consumer Rights Directive is concerned) as an employee, representative or associate of its home institution or employer, regardless of the ultimate source of the funds used to pay for that specific service. It is, therefore, expected that the EOSC End User will, in all transactions, 'contract with' the service provider as a representative of his/her home institution or organisation, and not as a 'Consumer'. Furthermore, in procurement terms the EOSC End-User's home institution is a legal entity with commercial capacity to enter into a contract and which also often has the status as a Contracting Authority as defined by the EC Procurement Directives 2014/24. On some occasions, one Contracting Authority may aggregate demand for a number of Contracting Authorities. Where a Contracting Authority is not subject to the EC Procurement Directive, there is often an obligation to evidence value for money by adopting criteria that test for the most economically advantageous tender rather than simply choosing the lowest price. However, any solutions in EOSC should aim to satisfy the governance requirements of both of these scenarios.

The Minimum Viable Ecosystem does not anticipate that the EOSC End-User will be characterised as a Consumer[9] (as defined by the EU Consumer Rights Directive), and the EOSC governance should consider whether the EOSC should eventually be made available to consumers when determining its future strategy for the EOSC after the initial phase (2019-2020). It should be noted that, unless they can be associated with a specific organisation capable of creating business relationships, Citizen Scientists would fall into the definition of consumer, both in terms of EU Consumer Rights Directive and EU VAT Directives.

As an example, Amazon is primarily a B2C platform, but it also offers accounts for business users. The following user journey illustrates the difference between a consumer and an organisational user as they interact with Amazon and the features it offers:

"As an Amazon customer, I must choose which account is the most suitable for my purposes: as a consumer I have my own account attached to my own credit card, where Amazon is bound by the standard terms and conditions that meet the requirements of the EC Consumer Rights Directive. I can return my purchases, I can receive refunds, and I benefit from *statutory* consumer rights. As a business user, I have to use separate credentials to log in to an account that my organisation has created, and I can use my corporate credit card or a shared payment method. I can utilise approval workflows and other business functions, but my employer signs the terms and conditions when the account is created, and my employer pays for my purchases; this relationship is subject to corporate law. So regardless of my demand or of the size of my business, I continue to be a corporate user. As

[9] In terms of the functionality offered by the Portal; non-VAT registered organisations would be taxed, where applicable, according to B2C VAT regulations.

a consumer, I might choose to use my own credit card to buy something for work purposes through my personal account – if I do so (even if my expenses are reimbursed), I'm still a consumer and the Consumer Rights Directive applies."

In summary, and in recognition of the statutory legal and regulatory parameters, all functions offered within the EOSC-hub in the Minimum Viable Ecosystem of the EOSC Portal support only business-to-business transactions (although any flow of funds must also account for VAT according to the EU VAT Directive); in this phase, there will be no functionalities to support a business-to-consumer model. It is recommended that specific analysis is carried out in order to understand the necessary business-to-consumer functionalities that would support the Citizen Scientist[10] and other consumer-type use cases.

3.3 Characteristics of EOSC suppliers

To further explain the types of business transaction the EOSC must provide for, it is important to define the nature of the typical EOSC supplier and how services will be provided. It is understood that EOSC Suppliers[11] will provide services to be accessed based on various access modes. The EC identified three base modes: Excellence-Driven, Market-Driven or Wide Access (see [R4]). A further access mode not covered by this taxonomy, but currently used by publicly-funded suppliers is the policy-based access, where an organisation would allow access to services if certain policies are met (e.g. nationality of the user, membership in a specific scientific community). The report in [R2] points out that *"Private providers willing to provide resources within the EOSC framework will envision a Market-Driven approach to support users."* The scenarios explored within this deliverable assume that every service will only be accessed via one of the three modes, and that the EOSC System will not be expected to accommodate a variety or combination of access models for each service.

Where Market-Driven Access applies, whatever transactional mechanism is used to ensure that the service provider is paid for the service, mandatory European regulations and legislation will apply to the transaction, and it will be necessary to consider, for example (this is not a limited or exhaustive list):

1. Whether 3rd-party 'contractual' obligations (and therefore liabilities) arise or become implied by an EOSC supply chain.
2. How procurement directives and financial regulations might impact EOSC's interaction with non-European service providers and users.
3. Where VAT responsibilities lie.

[10] "Broadly defined, Citizen Science is "scientific work undertaken by members of the general public, often in collaboration with or under the direction of professional scientists and scientific institutions." (ref: https://ec.europa.eu/research/openscience/pdf/citizen_science_recomendations.pdf), therefore, when Citizen Scientists are acting as agents of an organisation or Institution, they fit the EOSC End-User definition; when not affiliated to an Institution, Citizen Scientists are defined as 'Consumers'

[11] "EOSC Supplier Definition: an EOSC System User responsible for the provisioning of one or more EOSC Service Component(s) enacting an EOSC Service within the EOSC System. It is further specialised in the following sub-roles: EOSC Service Component Supplier, Data Service Supplier, and EOSC Service Developer." EOSC Glossary, https://wiki.eosc-hub.eu/display/EOSC/Glossary

4. Where exemptions in the Directives (such as supply of proprietary services) would enable a contract or transaction to take place that would not require a procurement (therefore, informing the RoP)
5. Where one Contracting Authority intends to offer or provide services to another Contracting Authority, the EOSC Hub should provide visibility to the end users and Institutional buyers in respect of how the requirements of Article 12 of the Procurement Directive have been taken account of in order to make compliant procurement (ultimately the liable party is the buying organisation so the EOSC Hub needs to demonstrate that services available from its "EOSC suppliers" are "procurement compliant" or the buyer will not buy).

EOSC Suppliers shall be private organisations (of various types of incorporation) or public institutions and it is assumed that the EOSC Suppliers are expecting to engage with organisations, rather than consumers (as explored above), i.e., in the context of a profession and not with private persons. Where an EOSC Supplier is a commercial service provider, that service provider will most likely expect to charge a fee for their services. Adopting the standard methods of commercial supply will attract a larger and more diverse set of service providers to the EOSC. Creating non-standard or complicated methods to remunerate commercial service providers will most likely deter commercial service providers from engaging with the EOSC.

Where an EOSC Supplier is a publicly-funded organisation, there is a need to understand how to incentivise the opening-up of services. Many publicly-funded organisations are not able to invoice for services, or if they can do so, they may be limited in the share of market-driven activities they can conduct as a proportion of their overall activities. They may also lack understanding of how to legally establish a fee-based public-to-public service provision above thresholds that would require a public tendering process. There are existing examples of money flowing for public-to-public service provision: 1) EGI has conducted a pay-for-use pilot activity to understand which affiliated publicly-funded service providers would be able to provide fee-based access to services (as opposed to the more traditional free-at-point-of-use access with indirect cost recovery)[12]; 2) the GEANT Association NREN membership offers services where the EC rules concerning public-to-public procurement are met, 3) the EOSC-hub project is trialling the virtual access mechanism from the Horizon 2020 framework program to recover marginal operational costs. These examples show different approaches with different levels of maturity for establishing a fee-based public-to-public service provision. Identifying a solution that would work in the context of the EOSC across borders would be an essential building block to enable the ecosystem to flourish and will be analysed further during WP12 Task 12.2, and considerations around this aspect are discussed further in the following section.

[12] EGI Pay-for-Use Initiative: https://wiki.egi.eu/wiki/Pay-for-use

3.4 Key considerations for the EOSC business model

In terms of the mechanisms to acquire or access services, the report in [R2] makes the following statements and recommendations:

"Conventional grants and contracts could be a logical way to manage the fixed costs of an EOSC, for providing resources to academic institutions and the private sector, and to also include models such as pre-commercial procurement for the development of new services for the development of capabilities or making large public data sets available on widely accessible infrastructure. Cloud coins are most useful for managing the marginal (or operating) costs of the EOSC, providing individual investigators with the ability to effectively move among the variety of resources that constitute the EOSC. They are also convenient when it comes to monitoring."

and

"Considering the approaching transition between H2020 and Horizon Europe a number of instruments should be considered, besides what has already been planned, to ensure a continuous support of resources to the implementation phase, to reduce the risk of interruptions in completing the operational vision of EOSC. For instance, an EOSC-PPP (Public Private Partnership) could be envisioned; additionally, incentives for pan-European infrastructure development programmes could be designed as well as capacity building programmes for people's competences."

Therefore, this deliverable shall assume that the manner in which the ultimate funding arrives into the open science arena will continue to have a bias towards central funding and need not be described in detail in this document, for example services that fit existing and traditional funding models, i.e., services developed/operated by public institutions for use by public and private sector research users would continue to be funded via grants or centrally procured structures, and funds flow from a public, grant-style funding model, rather than from a commercial-style revenue stream (where income derives from the customer of the service). Similarly, large-scale institutional usage of publicly funded services, even when discovered via the EOSC, would most likely continue to be fulfilled via specific purchase-supply contracts between the service provider and institution, and any related membership fee, subscription fee or similar charge would be levied outside of the EOSC Portal would also be outside of the scope of the EOSC System.

As such, the innovation will need to focus on the manner in which the funding flows in terms of providing sustainability for the services offered, particularly in terms of services where individual users require access to services on an ad hoc basis, therefore, 'cloud coins', 'vouchers' or other such credit mechanisms will require investigation as a method to represent financial transactions in specified situations, such as where usage is on a small scale but total contract values might exceed EC Procurement Directive thresholds.

Integral to the business model to be provided for innovative 'cloud-coin' style credit systems, is the need to carry out detailed investigations of the ways in which these models would operate, such as:

- How does the EOSC make services that are chargeable at their source 'free at the point of use'?

- Who would issue the vouchers/credits? How could the vouchers/credits be monitored for a) ensuring they are not under/over-utilised, b) use in future aggregated procurements.
- Does the proposed model comply with procurement, taxation (and other regulatory) directives?
- Which policies and guidelines do we need to capture in order for the participants (both researchers and service providers) to be clear on their obligations to each other, and for them to understand how to engage?

3.5 Opportunities for the EOSC in design of its business model

While the survey has highlighted the possible procurement needs and concerns of service providers and researchers, it gathered insufficient information about quantitative data related to the market potential as seen by those responsible for procurement decision making in the public research sector. Consequently, the models described by this deliverable consider:

1) EOSC End-Users consuming commercial services free-at-the-point-of-use (via market-driven access), exploring a voucher-based approach in order to establish whether vouchers/'cloud coins' or other credit-style mechanisms would be a fit-for-purpose long-term solution to enabling 'free at the point of use' access to commercial services.

2) 'Sponsored use' where institutional usage would be fulfilled via specific purchase-supply contracts between the service provider and institution, analysing the explicit terms that the EOSC Portal would need to capture that would address any issues concerning contract privity, EU Procurement Directives compliance, Teckal compliance and other statutory legislative/regulatory implications of such a business relationship.

This deliverable summarises the approach at a high level, with the intention of exploring and analysing in further detail under WP12 Task 12.2.

3.5.1 'Cloud Coin' or Voucher Access Model (Market-Driven Access)

A regularly referenced use case is that of the individual researcher or a small research group who need limited-scale access to commercial services on an ad hoc basis. Initial experience with a voucher access model has been documented by the recently completed HNSciCloud project [R5, R7] and is being actively expanded in the OCRE project[13]. This type of user expects free at the point of use services, simplicity, ease of use and ease of access. The services must also be easy to find and distinguish (as relevant to address their needs) from other services. The nuance of this model for chargeable services is that a party other than the end user will be expected to pay for utilisation, and the challenge is to ensure that a user can freely access services to which he or she is entitled, no matter how they were originally funded. The voucher could be used to distribute pre-paid, ring-fenced or discounted commercial resources. In fact, whether the mechanism allows for pre-payment or payment in arrears, it will be necessary for a user to prove that it is 'calling off' capacity from a specified commercial arrangement, potentially by using central AAI systems (and funding in the case of pre-paid services) specified during the procurement process, and this also impacts the

[13] https://www.ocre-project.eu/

procurement model as pre-paid means the supplier knows the sum they will be receiving when signing the contract, while post-paid means they only know when the vouchers are redeemed. Vouchers have already been successfully trialled in the context of incentivisation or introductory schemes, but there is an argument that 'cloud coins', vouchers or credits in some form could be a legitimate mechanism to provide a long-term solution to providing services that are freely available to entitled users. This deliverable assumes that when a researcher receives a grant, the financial commitments against that grant are paid by the institution (via research support managers, finance departments, etc), and that the researcher doesn't have specific authority to actually raise a payment against a commitment of this nature. Therefore, the voucher could be allocated to a particular research project, but the resources the voucher relates to would only ever be paid for by the institution in receipt of the grant. This requires further analysis and investigation with commercial providers and institutions open to assisting in this analysis.

Decisions would need to be made about various policy matters, such as:

- Who would be entitled to issue the vouchers? Who would monitor their use? There may be a need for some form of regulator/broker/hub to administer and manage the voucher allocation, and the terms of their involvement would need to be defined.
- It could not be assumed that the Contracting Authority reserving (and thus paying for) the resource would be the same organisation distributing the entitlement to use that resource, therefore, the policy and process for its implementation would need careful consideration.
- Whether it would be sensible to limit the amount of resource that could be called off by each institution (in order to share the resources fairly and equally across entitled parties).
- What would happen when the voucher limit is reached?
- What would happen when the ring-fenced/pre-paid resources run out?
- Would the vouchers be time-limited?
- What would happen if the resources are not consumed in their entirety before the funding project/deadline expires?
- A dispute resolution process will be required in case the service level agreement is breached
- Would the vouchers provide access to the full range of features and functionality? If not, how would the researchers 'top up' their allocation to gain unlimited access?
- What would happen in the event that the funding body implements a policy and/or contractual obligations that requires its grantees to use the funds for the purchase of approved EOSC-compatible e-infrastructure services only?
- Where the goal is to discourage researchers from buying local storage solutions, would it be necessary to implement policies that explicitly forbid the funds from being spent on local resources or unapproved services?
- Would EOSC vouchers be valid only for those services listed in the catalogue which satisfy the Rules of Participation?

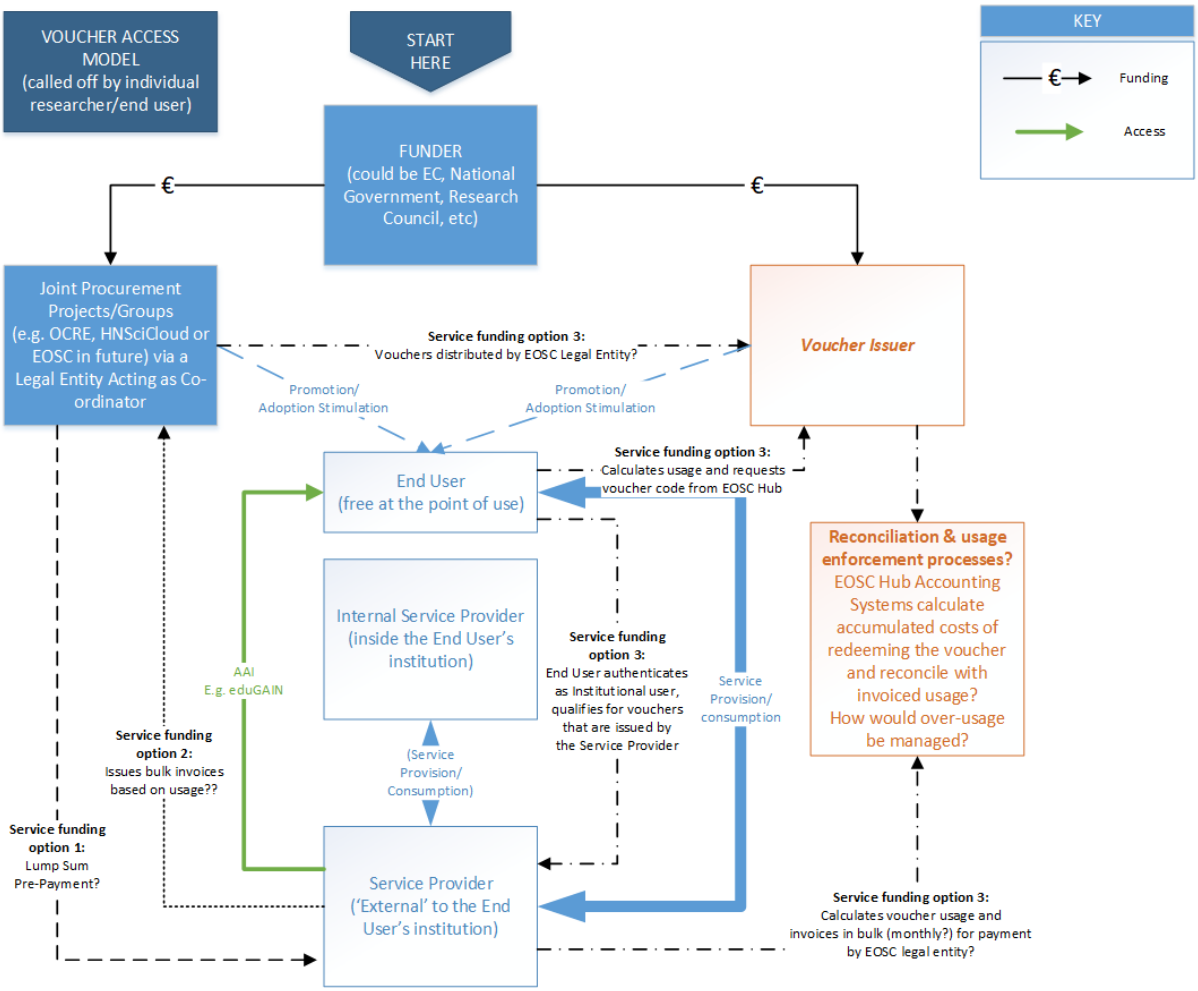


Figure 5: 'Cloud Coin' or Voucher Market-Driven Access Model

3.5.2 'Sponsored User' (also Market-Driven Access)

This use case is illustrated by scenarios where institutional usage would be fulfilled via specific purchase-supply contracts between the service provider and institution, where any usage limits and payments are managed between the institution and the service provider on a 1:1 basis. As the EOSC Portal brokers services, it would be vital to ensure that the terms of utilising such services, having discovered them via the EOSC Portal, are made explicit in order to avoid any implied 3rd-party relationships between the EOSC, Service Provider and buyer. This would need to capture and address any issues concerning contract privity, EU Procurement Directives compliance, Teckal compliance and other statutory legislative/regulatory implications of this nature of a business relationship.

It is important to note that in this model, it is still expected that the user enjoys free at the point of use services. However, the differentiating factor of this model is that the entity that the user is associated with contracts with the service provider directly to payment terms agreed and fulfilled between the two entities. As the nature of contractual arrangements is likely to differ from buyer to buyer, there would be little to no opportunity to utilise voucher/coin/credit schemes. It will continue to be necessary for a user to prove that it is 'calling off' capacity from its specified

commercial arrangement, however, such access mechanism would not require association with a larger entity (as necessary in the case of the Voucher/Coin model), therefore, it would be necessary for flexible AAI systems to be in place for the service providers and buyers to select from.

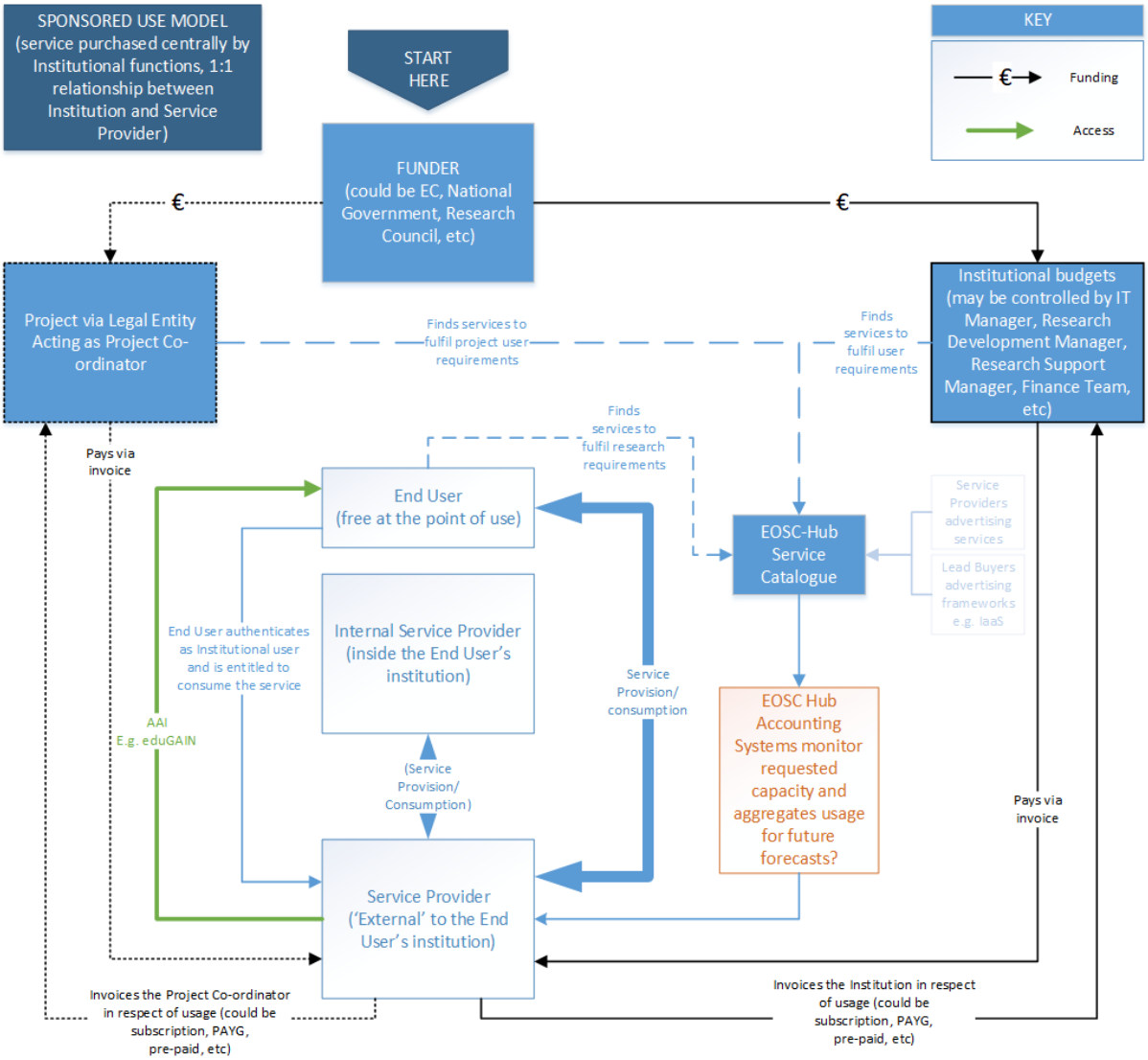


Figure 6: 'Sponsored Use' Market-Driven Access Model

3.5.3 Aggregated Procurement Opportunities

Aggregating the committed demand of Contracting Authorities trading with EOSC Suppliers, has the potential to bring economies of scale and process efficiencies. Procurement law allows buying organisations to establish commercial and contractual models that reduce the need for repetitious tendering for similar goods and services. These may take the form of framework agreements, for example, where one Contracting Authority (operating as a Centralised Purchasing Body - CPB) establishes framework agreements for use by a number of other Contracting Authorities. Other examples include Pre-Commercial Procurement (PCP), Public Procurement of Innovative solutions (PPI), Public Private Partnerships (PPP), or Dynamic Purchasing Systems (DPS).

Within EOSC, subject to satisfying the various governance requirements of interested EOSC-End Users and their home institutions (Contracting Authorities), a buying organisation could establish such agreements, albeit the same organisation might not actually need to commit any expenditure with suppliers despite being entitled to do so. Contracts under such frameworks or DPS' are typically established by the Contracting Authorities/entities described within the tender documentation during the relevant procurement activity.

When committing expenditure with suppliers in instances where one Contracting Authority/entity acts on behalf of itself and others, then the terms/licenses established during the tendering stage should be complied with. For example, many suppliers operate a channel model, and this may preclude one entity buying on behalf of a number of others and *reselling* that capacity to them. To overcome such barriers, a group of Contracting Authorities has far more commercial or contractual leverage with the supply base than acting alone.

Within an established framework, either the CPB or other organisations representing a group of buyers (e.g. for a certain scientific discipline with similar needs) could act as Lead Buyer[14] in surveying groups of EOSC End-Users and related service providers to assess the scope for collaborative purchasing. Careful planning is needed in respect of managing collaborative procurement exercises in order to achieve genuine commitment from the buying organisations, which in turn gives the market comfort that their effort in responding to the tender will not be wasted. This is key as to achieve preferential rates and terms for those participating, suppliers need to really see the value in taking part. The Lead Buyer will need to provide 'added-value' services to support the initiation and execution of joint procurement activities in order to ease the process of planning and implementing the tendering activity. For instance, in the context of the recently funded H2020 cluster projects in the context of the INFRAEOSC-04 call, the projects PANOSC, EOSC-Life and ESCAPE have envisioned a procurement activity aiming at better investigating how sourcing of services could be improved. In this context, joint procurement is believed to be viable, albeit how it is to be taken forward is yet to be defined. Cooperation with OCRE as a pilot procurement vehicle for EOSC is expected for these projects.

This could lead to the frameworks resulting from the aggregated procurement exercises either being published/promoted or accessed via either the EOSC Portal Catalogue, or via a dedicated area of the Portal that could also provide an opportunity for customers, funders and policy makers to post their requirements (via surveys published in the same dedicated area) and for EOSC to publish its policies for utilisation of these frameworks.

In practical terms, where framework agreements designed to leverage economies of scale through aggregation can offer preferential rates and terms to the end user, a mechanism within the EOSC Portal should be provided in order to ensure that end users can benefit from the improved rates and terms as uptake and aggregate spend increases; this will be explored in Task 12.2. A mechanism may also be necessary to monitor and enforce the performance of the SLA against the parameters of SLA attributes in this model. Clear distinction, however, would need to be made in order to establish where 'commercial risk' lies in the resulting frameworks. For example, would the Lead

[14] A lead buyer is acting as contracting authority and is responsible both for the project management (e.g. coordination with contracting authorities) and the execution of the procurement procedure.

Buyer/Funder procure pre-paid services, therefore, taking a risk in paying for more services than eventually consumed? Or if preferential rates and terms are based on committed utilisation volumes, would the Lead Buyer/Funder be exposed to penalties? In the event that the Lead Buyer acts as an aggregator of utilisation, does it become a creditor of the service providers? And how would such concepts be implemented in the context where many commercial IT service providers operate channel models that only allow certified 'value added resellers' to behave in this way?

It is worthy of note that the formation of Central Purchasing Bodies must be centrally co-ordinated. While the concept of aggregated procurement as a whole implies that the community would collectively benefit from increased buying power, unless it is carefully managed there is a risk that multiple Central Purchasing Bodies might be formed unwittingly in competition with each other, which would weaken the combined buying power of the community as a whole. Central co-ordination could prevent this from happening and could instead ensure that Central Purchasing Bodies focus on 'thematic' demand aggregation, whilst representing a 'single voice to market' in its engagement with the supply market.

When inquiring about how an EOSC hub could better facilitate or support the procurement process, the following aspects were mentioned (and it may be that an intermediary organisation behaving as aggregator or Lead Buyer could improve the user experience):

- expert support in selecting digital services for research and assessing services against key requirements, such as data security.
- make procurement simpler, more cost transparent and faster (especially for compute)
- support demand aggregation (unlikely to engage in a procurement process in isolation)

Were EOSC to become a legal entity capable of behaving as a Lead Buyer, or aggregator of demand, it would strengthen its stated ambition to become the Number 1 source of services to the research community, therefore, increasing its ability to attract commercial services providers to collaborate. However, many aspects of this concept require further detailed evaluation, as outlined above and below.

4 Recommendations and Conclusion

4.1 Recommendations

Based on the work presented in this document, this section provides recommendations for different types of stakeholders potentially interested in this deliverable. The structure follows:

- General Recommendations
- Digital Service Provision Recommendations
- Future EOSC entity Recommendations
- EOSC Executive Board Recommendations

From the analyses, one can draw different conclusions and develop some recommendations for the future of EOSC and of the e-infrastructure community in general. Most of these recommendations can be taken on board either by the service providers themselves or else by the higher EOSC-related bodies. However further actions will be suggested in order to continue building the momentum towards sustainability.

4.1.1 General Recommendations

It can be implied from the results of the survey and interviews that many non-specialist users do not know what they want from EOSC in the future. Rather than plan the services they need, many users will assess what is offered in relation to how they find it useful at the time they require a service. It is unlikely that this type of user will ask for a new service. However, users would like to be involved in the specification and design of nominal new services. This approach has been used successfully at the national level and can be detected at a regional level in reported resource pooling activities and should continue to be used here. IT specialists and heavy e-infrastructure users do develop plans to meet their future requirements, these users require domain-centric co-design approach if an e-infrastructure service provider is going to satisfy their future requirements through developing its own service catalogue in a manner that is meaningful to this type of user.

Two clear profiles describe service needs in relation to EOSC services: Simple and Complex profiles. These profiles are summarised below, and fulfilling their needs drives the recommendations which follow.

Service characteristics of simple initial need

- Simple services that:
    - Are easy to select (in relation to all other services, similar or not)
    - Are easy to access (including SSO)
    - Are easy to use
    - Are always on
    - Are free to use (at point of use)

- Deliver secure and confidential services
- Enable data privacy
- Provide trusted repositories and similar data services (e.g. registries)
- Offer data interoperability and include robust data management systems (without changing existing workflows)
- Provide long term domain-agnostic data preservation
- Offer support services and training
- Expose comprehensive data and service catalogues

Service characteristics of complex initial need

- More complex services that incorporate and build on many of the simple service characteristics, but which:
  - Can be co-designed on a domain by domain basis
  - Have well designed interfaces
  - Enable (user) supervision and monitoring services to help research leaders understand service usage
  - Are inexpensive/affordable when deployed

4.1.2 Digital Service Provision Recommendations

Considering the findings and insights concerning the researchers' demand for data management services, a range of services are recommended below that could offer added value of EOSC to researchers.

- First, there is a strong need to further develop and deploy awareness-raising, advocacy, training and support activities to enable researchers in achieving FAIR standards for their data. Also, the use of FAIR standards should be exploited as a label to demonstrate that a quality standard has been achieved regarding particular data or data sets.
- Based on our experience, for their daily work, researchers would benefit from a secure enterprise file synchronisation and sharing service (EFSS[15]), typified by Dropbox and similar offerings. Such EFSS services would allow researchers to securely share and synchronise data with colleagues and collaborators across multiple devices and using a combination of on-premise and cloud storage depending on their security or compliance needs. When publishing their results and data, researchers would need a reliable and trusted digital repository, typified by Zenodo[16] and similar offerings. The EFSS and repository services should work together seamlessly so that, once a researcher decides it is time to publish, it should be a

[15] https://en.wikipedia.org/wiki/Enterprise_file_synchronization_and_sharing
[16] https://zenodo.org/

simple operation to select material from the EFSS service and have it published via a digital repository.

- It is primarily at the time of publishing that researchers consider the FAIR data principles for scientific data management and stewardship[17] and so the digital repository, more than the EFSS service, should implement these principles and ensure long-term data preservation to a level that satisfies the requirements of the researcher's funding agency/agencies and can be included in the data management plan (DMP) for their research.
- Researchers would also like access to compute facilities to process their data. The compute facilities should interoperate with the data management services, so no manual intervention is required by the researcher to use the compute facilities to process data held in an EFSS or a repository.

4.1.3 Future EOSC Entity - Recommendations

The recommendations in this section are directed to the future EOSC entity that is envisioned to be established. As this does not exist yet, they can be considered in the design phase:

- Decide on the key target groups of the EOSC entity and invest in developing beneficial, meaningful relationships. Take into account what the individual researchers and small research groups are saying since the e-infrastructure community needs to be providing these kinds of services. Although these users do want such services at no end-user cost, an appropriate business model can be chosen to meet this important requirement. The larger research groups and larger research institutions need to be engaged in a more interactive manner in order to continue meeting their needs through a co-design process and the relationships here already exist, so they should be easier to maintain and develop.
- Ensure there is a procurement and 'standards' body, de-risking the procurement of commercial services, providing easy access to all services (regardless of whether they are public specialised services or commercial commodity services) and ensuring the quality of all services. Do not attempt to become the all-in-one provider of all services, or to focus only on publicly supported e-infrastructure services. Procurement remains an open issue for EOSC and needs to be addressed with activities explicitly aimed at the procurement community.

4.1.4 EOSC Executive Board Recommendations

The recommendations in this section are directed at the EOSC Executive Board:

- Consider EOSC as an intermediary organisation rather than as a specialised service provider and aggregator organisation. EOSC should not try to compete with large commercial service providers. As an intermediary organisation it can coexist effectively alongside the commercial service providers; EOSC will still be able to aggregate services but these no longer need be exclusively public services. As an intermediary organisation, EOSC should also be able to:
    - Reduce service costs for researchers by bulk buying and de-risking commercial service provision

[17] https://www.nature.com/articles/sdata201618

    - Support joint procurement groups for larger research entities
    - Set, monitor and enforce standards of service delivery
    - Offer awareness-raising, advocacy, training and support services
    - Monitor and manage FAIRification efforts
- When planning landscaping analysis, include in objectives the understanding of the intention, feasibility and constraints of publicly-funded service providers to sell services across borders and disciplines. The output should help designing frameworks that simplify cross-border public-to-public service provision.

4.2 Conclusion and Next Steps

This work contributed to the essential activity of designing future business models and procurement frameworks that support the sourcing of digital services from both publicly-funded and commercial providers in the EOSC. It started with a demand-side market research to understand the need for and level of demand of digital services for research, the manner in which such needs and demand are currently satisfied, the challenges presently being faced in respect of analysis workflows, data management and related infrastructure and services. It also supported the identification of current funding streams and procurement strategies, related constraints and opportunities, and areas of improvements for business models. The demand-side market research based on interviews and online survey was also enriched with a world cafe style session run at the DI4R conference in Lisbon, 10 October 2018.

Based on the demand-side market research, a number of insights were identified. Part of these insights have been used to conduct a business model analysis in the EOSC context, with a specific focus on proposing mechanisms for acquiring digital services in a business-to-business (B2B) type of relationship. Three main models were considered: 1) 'Cloud Coin' or Voucher Access Model and applies when an individual researcher or a small research group needs limited-scale access to commercial services on an ad hoc basis; 2) 'sponsored use' where institutional usage would be fulfilled via specific purchase-supply contracts between the service provider and institution; 3) 'aggregated procurement' aiming to support demand aggregation and joint purchasing. The insights were also used to identify a number of key recommendations for different types of stakeholders.

Considering the next phase of the EOSC-hub Work Package 12 activities to be carried forward by Task 12.2, based on the results of this work, it is recommended that:

- **The ''Cloud Coin' or Voucher Market-Driven Access Model' is analysed further in Work Package 12 Task 2 'Procurement and purchasing framework'.** This develops the concept of 'cloud coins' that are proposed by projects as an incentive to adopt cloud services, or as introductory schemes (see the recent paper "Voucher Schemes for Accessing Commercial Cloud Services in the Research Environment"), into a potentially more long-term solution to enable free use of paid commercial services.

A voucher scheme for providing access to commercial cloud services for public sector researchers has been developed, tested and the results documented by the HNSciCloud project. A voucher scheme is currently being piloted by the EOSC-Hub Digital Innovation Hub for providing access to publicly funded services for industry/SMEs. A voucher scheme is also being proposed as one of

several procurement channels within the OCRE project. There is merit in investigating how vouchers can be practically utilised as a long-term solution to distributing free-at-the-point of use services to end users. There are contractual and taxation issues associated with voucher payment, and it would be a valuable exercise to set out practical and concrete guidance to EOSC that sets out 1) the end-to-end process necessary to make vouchers a viable method of proving entitlement and monitoring usage, 2) the specific financial regulatory parameters and rules that must be adhered to when offering services via vouchers, 3) determining which parties have 'contract privity' (the contractual relationships between the parties that mean that the parties subject to a contract do not have contractual obligations to 3rd parties) in such a model, and looking at other types of service provision (e.g. software licensing) that may be incompatible with this model.

- **The 'Sponsored Use' Market-Driven Access Model is analysed further in Work Package 12 Task 2 'Procurement and purchasing framework'.** Research organisations often rely on services provided by publicly funded infrastructures. It is likely that as well as services being consumed directly by the researcher, there will be requirements for institutions at a departmental level and projects to procure services from service providers to the EOSC.

Typically, supply-side organisations from the public sector provide services free-at-point-of-use or with indirect cost recovery. Sometimes the demand-side organisations pay for the services. In this scenario, there is a need to identify what would be required to ensure that institutions can buy or consume services via the EOSC. There would be a need to investigate and analyse the end-to-end process and the agreements, etc, that would need to be provided in order to facilitate the process. It may also be necessary to investigate potential Teckal implications, the various roles in EU Procurement and their characteristics and whether procurements via the EOSC would be compliant with EC directives.

- **Both models must consider public-to-public, private-to-public and public-to-private scenarios and must consider the wider funding environment in order to identify opportunities for new ways of working**.
- The task should attempt to illustrate the process, working with an example/typical institution and a commercial cloud provider, in terms of issuing purchase orders and invoices.
- **A continuing dialogue will be maintained between the EOSC-Hub and EOSCsecretatiat.eu projects** to ensure that business model recommendations are included in the consultation of the Sustainability Working Group that will focus on determining a fit-for-purpose business model for the EOSC.
- **Work Package 12 Task 2 will work in close collaboration with the OCRE and ARCHIVER projects**, both projects having been initiated during the course of 2018 and commenced in 2019. These projects present ideal case studies that will identify the challenges that the EOSC business model will need to address. OCRE will procure commercial cloud services, utilising EC funding intended to increase adoption of cloud services, and ARCHIVER will procure the R&D of solutions to address long-term storage and data preservation, in order to assess how cloud solutions and short-term project funding can be made more compatible

via a Pre-Commercial Procurement. Practical guidance can be offered to the OCRE and future projects.

- **Work Package 12 Task 2 should propose the terms of reference for future EOSC Central Purchasing Bodies or intermediary organisations acting as demand aggregators,** including the characteristics and criteria that any such group must meet in order to comply with EC Directive and taxation regulations, and competition law, and the design of a process to establish such groups. Such recommendations could be made to the EOSC Executive Board Working Group on Sustainability that will consider suitable business models, governance structures and legal entity.
- It would be helpful to find a way to identify the main websites and services being accessed by users; the goal is to identify which providers are being accessed and to try and understand why users are making their selections.
- Framework contracts/agreements and long-term contracts are already in common use, indicating a preference for long-term stable relationships. An intermediary body could continue to deliver this level of confidence to its service users; however, the intermediary body should be prepared to renegotiate bulk deals from the supply chain as often as is required in order to maintain the best terms and to fulfil the needs of the research community, and should consider the position of such a function in the context of national initiatives that also provide pre-negotiated framework agreements.

5 References

No	*Description/Link*
R1	Implementation Roadmap for the European Open Science Cloud, Commission Staff Working Document, 14 March 2018, https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf
R2	Prompting an EOSC in practice, Final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC), 2018 https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy
R3	The notion of 'consumer' in EU law, Library Briefing, Library of the European Parliament. 6 May 2013. https://epthinktank.eu/2013/05/23/the-notion-of-consumer-in-eu-law/
R4	European Charter for Access to Research Infrastructures https://ec.europa.eu/research/infrastructures/pdf/2016_charterforaccessto-ris.pdf
R5	B. Jones, J. Fernandes, M. Devouassoux. Voucher Schemes for Accessing Commercial Cloud Services in the Research Environment. 29 March 2019. https://zenodo.org/record/2615456
R6	EOSC-hub Week 2019 https://www.eosc-hub.eu/events/eosc-hub-week-2019/programme
R7	EGI-Engage D2.11 Cross-border procurement of e-Infrastructure services: Opportunities, Barriers, Use cases, Best Practices, 6 March 2017 https://documents.egi.eu/document/3013

Appendix I. Demand-side market research: Methodology

I.1 Scope

Developing and offering a technical ecosystem of data and services is not enough to ensure that research data will start to be made FAIR by the data producers and actively reused through various computational analysis services by other researchers and data users. Such an ecosystem will be effectively used only if an underlying FAIR data culture prevails amongst researchers and there are adequate incentives, skills and support (including financial) to support it. Active and FAIR research data management and reuse is in fact not currently general practice in many scientific communities, and it is important to understand the diversity of situations and obstacles.

The demand for the EOSC e-infrastructure and the services it will deliver will materialise and reach its full potential only if sufficient awareness, interest and ability to use them exists amongst researchers and their support communities. Therefore, the level of readiness and product–market fit, as well as any gaps that need to be bridged to ensure the translation of a perceived need into actual demand, specifically for the EOSC federated services, have to be assessed.

In this overall context, the EOSC-hub project identifies that collaborative and cross-border access to such an e-infrastructure and its services is an essential capability to be organised, as it has the potential to stimulate demand by increasing efficiency and reducing time, effort and risk devoted to resource acquisition.

I.2 Approach

There is very little published information on the current level of demand and spending on e-infrastructures and FAIR data management in the EU Member States[18]. Therefore, an in-depth understanding of the needs and challenges faced by research organisations and communities vis-à-vis e-infrastructures and services, and in their sourcing strategies of such services, should be developed. Additional understanding about how these issues could be better addressed in the context of the EOSC also needs to be developed. In this context, the most appropriate approach to undertake a demand-side market analysis is one driven by primary research[19]. We structured our methodology in the phases presented in Figure 1

[18] Hodson, Jones et al. (2018) Turning FAIR data into reality. Interim report of the European Commission Expert Group on FAIR data

[19] Market research can be undertaken as either primary or secondary research, according to the focus on the gathering and analysis of new data that has not been collected before, versus analysis of existing data from general sector studies already having been produced. However, should such studies and their data be insufficient or too general compared to an organisation's information needs vis-à-vis the market, the undertaking of primary research is necessary. This is largely based on field research, involving for example surveys using questionnaires and/or interviews with groups of people in a focus group, and is tailored to an organisation's particular needs. As such primary research delivers more specific results than secondary research, which is an especially important consideration when launching a new service.

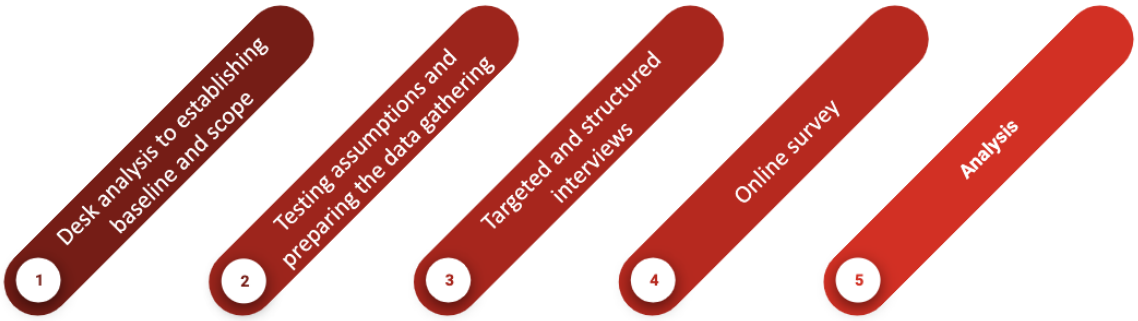


Figure 7: Methodology applied in the deliverable

During the first phase, a desk analysis of any existing information relating to potential users was undertaken. This was based on both the available public information, and the internal information available within the EOSC community, especially from those entities already providing access to the e-infrastructure and services. The exercise provided a baseline and initial assumptions on the market demand characteristics.

During the second phase, preliminary interviews were undertaken. The purpose of these interviews was to: 1) scope out the level of need for computing and data intensive services in experimental analysis and the level of exposure and expertise of the interviewee as regards them and, 2) discover the breadth and depth of understanding of the current service supply model and the possible future service supply models. This phase was intentionally fluid and unstructured, requiring the interviewer to explore narrative answers to questions with the interviewee.

In the third phase, a structured interview template was developed for delivery through remote conferencing tools (or face to face if possible, with phone being a last resort). The template comprised both multiple-choice and open-ended questions (see Appendix IV). Interviewers were nominated by each task partner and a training was organised to ensure a common approach in conducting the interviews[20]. A total of around 50 interviewed organisations was planned in both phases 2 and 3.

In the fourth phase, based on and informed by the overall results of the interview phase, an online survey of potential users was pursued. In the initial plan, it was proposed to have a shorter, self-applied questionnaire focusing on the most important aspects which had emerged during the interview phase. As the number of collected interviews was lower than the expected target, the approach for this phase was changed to keep a rich questionnaire and distribute it via the various promotion channels of the task partners.

The final phase was devoted to the analysis of the collected data. Initially, the consulting company extracted the findings from the data and offered them to the task partners as domain experts for comments. Following the feedback that helped to better interpret the data, the narrative on the

[20] Face-to-face meeting to train the interviewers: https://indico.egi.eu/indico/event/4251/

findings was developed. Finally, the task partners developed the insights section, providing an interpretation of the data.

I.3 Specific objectives

The main objectives of this exercise were the following:

- Understand the need for and level of demand of digital services for research in the context of the EOSC
- Understand the manner in which such needs and demand are currently satisfied
- Understand the challenges presently being faced in respect of analysis workflows, data management and related infrastructure and services
- Identify current and preferred delivery models for such services
- Identify funding streams and procurement strategies (e.g., drivers, constraints approaches)
- Identify areas of improvements for business models

I.4 Target audience

I.4.1 Organisations

The target audience for this demand-side market analysis were representatives of research and support communities that require and/or organise a service offering and either need to currently source it from outside, plan to source it from outside, or may source it from outside if the circumstances benefit them in some way.

The intention was to target a broad range of potential buyers of digital services both inside and outside the core European Research Infrastructures (RIs) and RI Clusters, also including federated and independent research communities, institutes and organisations. A subsequent aim was to include private industry and public sector potential users if time and funds allowed.

Given the limited number of organisations that can realistically be interacted with in an interview format, those targeted for phase 3 were deemed to be representatives of potential early-adopters and customers with significant market influence and bargaining power (e.g. main research organisations/communities with significant interest and budget for research support services). This approach was agreed as it would allow their needs, constraints and requirements to be taken into account early in the business modelling process, hence reducing the risk of developing a business model misaligned with market demand. It was envisaged that the Pareto rule of 80/20 would apply here, anticipating that 80% of the market demand will come from 20% of the overall client base. Therefore, a particular attention was given in the selection of the research and support communities to be involved in the primary market research, with the intention of moving on to the “long tail” of science subsequent to understanding the core requirements. It was considered important that the interviewed organisations should be selected not only or mainly from amongst those communities with which EOSC partners are already involved as service providers, in an attempt to avoid engaging only with the “converted”. The list of organisations that participated in the interviews is presented

in Appendix V. It was agreed that the demand analysis would cover a balance of the main types of research stakeholders, presented in Table 1.

1. Organisations serving large organised research communities
    - EIRO
    - Big lab
    - ESFRI
        - RI Projects / RI Clusters / RI Landmarks
            - Especially those funded under INFRAEOSC-04-2018
    - International research collaborations
2. Organisations serving small research groups / individuals
    - Universities
    - Research centres
3. National Research and Education Networks (NREN)
4. H2020
    - H2020 project
    - EOSCpilot demonstrators (*providing the means of accessing informed subjects).*
5. Scientific-service provider or consumer institutes
6. Private sector/business organisations with significant (initially pre-competitive) research focus (accepting that there may have been some advantage to exploring those entities conducting more advanced industrial research later)
7. Public sector organisations with significant research focus

I.4.2 Representatives

In line with the typical stages involved in the process for defining and organising a complex service procurement, it was envisaged that three types of interviewee would be engaged (with specific questions within the overall set):

1. Those working with the researchers who will use the services:
    a. Those establishing the user requirements, being leaders of research groups/departments.
    b. Those specifying the requirements for the protection and storage of research results, being research librarians.
2. Those establishing the technical specifications, being the organisations' IT facilities managers.

3. Those organising the procurement and required financing, being managers or procurement leaders that are responsible for defining the sourcing strategy and understand the funding streams.

The interview was structured to accommodate these three categories. Whilst the interviews could have been undertaken separately with representatives of each of these target segments, a focus group interview approach was pursued as it was considered to be more effective, both from a time management perspective, and to allow for cross-answering of questions, as well as for internal discussion, correlation and validation of answers.

In order to entice the contacted organisations to participate in such an interview, a concise introduction to the scope of the exercise, the EOSC concept and the FAIR data drive of the European Commission was provided. It was also made clear to the invited interviewees that the collected information would be treated with highest confidentiality, including within the EOSC-Hub consortium, and in compliance with GDPR provisions. Only aggregated and anonymized information are included in this final report with the summary of the findings of demand-side market research and to inform the future business models of EOSC.

For the wider survey phase, the questionnaires were structured along the same main sections and with clear instructions as to how and by whom the various questions to be answered. It was agreed that the main audience to be addressed should be the research communities' IT facilities managers, with a view to their liaising internally with their colleagues responsible for procurement, finance and research when answering the questions. A similar introduction and clarification as to the confidential treatment of the data collected was provided. Other type of incentives or rewards for filling in the questionnaire were considered, but no incentive was offered beyond the opportunity to receive some 'consultancy' to assist later.

I.4.3 Interview structure and questions

The interviews were structured in three main parts – each section targeted the three role types (researchers/users, IT Managers and Procurement Managers). Some questions were pertinent for more than one of the roles, as each of them would bring forward a certain perspective and together give a more complete picture of the demand situation in their organisation. It was understood that the questions were ambitious, and that it might not be possible to generate the engagement of all necessary parties at each targeted organisation. However, it was determined that an ambitious approach would gather richer responses than a narrower approach might have yielded.

It was suggested that ideally the interviews would be performed in a focus group format, with the various roles all present; however, it was recognised that this may not be possible to achieve and that in that case a focus on the end users and budget holders for the tools and services that the end users depend upon would suffice, following up with a second interview, with a more detailed set of technical questions, directed to the ICT specialists.

Interviewers were proposed by each of the EOSC-hub partner organisations contributing to the task. They also agreed to target specific organisations for each of their interviews. The proposed

personnel had a solid understanding on the scope of the exercise, and some had experience in client facing activities or customer/user interaction.

In the interest of comprehensiveness and comparability of the interview results across various organisations, it was agreed that all groups would be asked the same questions, but it was recognised that flexibility would be required in order to accommodate the various differences in perspective of the interviewees. Interviewers were encouraged to take note of the salient answers and information provided by the interviewees, and if possible, also record the entire conversation for further analysis, collating the answers post interview in a structured format in line with the interview order of questions and any additional information provided. A mix of quantitative and qualitative analysis of the information collated was then performed, serving to address the specific main objectives of the demand-side market research exercise. The questions were mapped to the objectives of the demand analysis as described in Table 1.

Table 1: Objectives of the demand analysis and related questions

Objectives	**Questions**
Understand the need for and level of demand of digital services for research in the context of the EOSC	B1, B2, B3, B4, B5, B8, C1, C2, C3, D8, D9, D10
Understand the manner in which such needs and demand are currently satisfied	B6, B7, B8, C1, C2, C3, C4, C5, D1, D11, D13, D16
Understand the challenges presently being faced in respect of analysis workflows, data management and related infrastructure and services	B9, C1, C3, C4, C5, C9, C10, D1, D5. D11, D13, D15, D16, D17, E2
Identify current and preferred delivery models for such services	C1, C2, C3, C4, C6, C7, C9, C10, D1, D10, D12, D15
Identify funding streams and procurement constraints	C6, C8, C9, C10, D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D12, D13, D16, D17
Identify areas of improvements for business models	C4, C5, C6, C7, C9, C10, D2, D3, D6, D7, D8, D9, D10, D11, D15, D16, E

In order to ensure a uniform approach and consistent quality across the interviews, the team of interviewers were thoroughly briefed on the process by an expert in market research, who also offered advice on how to conduct the interviews (including methods to navigate the interviews by discussing the questions in detail and identifying possible points for further in-depth probing or for when skipping questions might be opportune) and how the interview feedback should be captured, structured and collated. The interview questions can be found in Appendix IV.

Appendix II. Demand-side market research: Timeline

The methodology was implemented following the timeline presented in Table 2. The list of organisations that participated in the market research is presented in Appendix V.

Table 2: Timeline of the information gathering and analysis process

Phase	Activity	Period
Phase 1. Establishing baseline and scope		
	1.1 Desk analysis to establishing baseline and scope	Mar-Jun 2018
Phase 2. Testing assumption		
	2.1 Identify interview range and scope, devise interview parameters and draft the questionnaire	July - Sep 2018
	2.2 Train Interviewers	2 Oct 2018
	2.3 Conduct pilot interviews	7-11 Oct 2018
Phase 3. Targeted and structured interviews		
	3.1 Complete 'questionnaire', introduction and invitation to interviews	Oct 2018
	3.2 Complete the list of organisations targeted for interviews	Oct 2018
	3.3 Engage with interview 'candidates'	Oct-Nov 2018
	3.4 Schedule and undertake rest of interviews	Oct 2018 - Jan 2019
	3.5 Analyse interviews results	Feb - Mar 2019
Phase 4. Wider online survey		
	4.1 Prepare self-applied questionnaire reflecting interviews results	Feb 2019

	4.2 Publish / distribute questionnaire	26 Feb-19 Mar 201921
Phase 5. Final analysis and development of recommendations		
	5.1 Insights from market research and recommendations	Apr 2019
	5.2 Presentation of results in a public event	Apr 2019

[21] News item with opening of the online questionnaire: https://www.eosc-hub.eu/news/eosc-hub-survey-demand-procurement-digital-services-research

Appendix III. Demand-side market research: Findings

Detailed analysis of combined Section 1 Respondent Profile data

The analysis here is based on a merged set of data resulting from the combination of the targeted Interviews and the public Survey responses.

In all, sixty-five responses were collected (thirty-two from the focus group interviews and thirty-three from the public survey); however, five of the survey responses were spoiled. From the incomplete state of some responses, it became clear that some respondents carried out a trial response of the questionnaire before returning to log in again and perform a complete response. These incomplete responses were deleted from the pool to preserve data purity, leaving a total of sixty useable responses. The following Section 1 analysis refers to the aggregated responses of both the interviews and the survey. Responses were mainly from the research community (63%) in three categories: Research Leaders, Researchers and Organisational Leadership/Management (which were usually from the research community). It is notable that most respondents identified themselves in the "Leadership" roles (52%). Technical support personnel responded at the level of 14%, while 23% of the respondents identified themselves as "other" which, so far as it is possible to tell, were involved in various technical and administrative roles. One respondent did not specify his/her role. No respondent identified him-/herself belong to categories of "finance" or "procurement". See Figure 8.

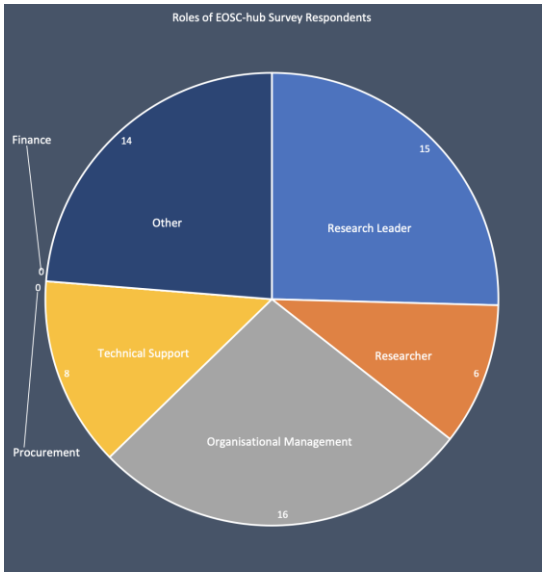


Figure 8: Roles of EOSC-hub respondents

Respondents were not evenly spread geographically; many countries were not represented. 14 countries were covered and their contribution to the overall response was: Netherlands (21%), Germany (19%), UK (14%), France (7%), Spain (7%), Austria (5%), Italy (5%), Sweden (5%), Switzerland (5%), Hungary (3%), Portugal (3%), Croatia (2%), Lithuania (2%) and Norway (2%) See Figure 9 below.

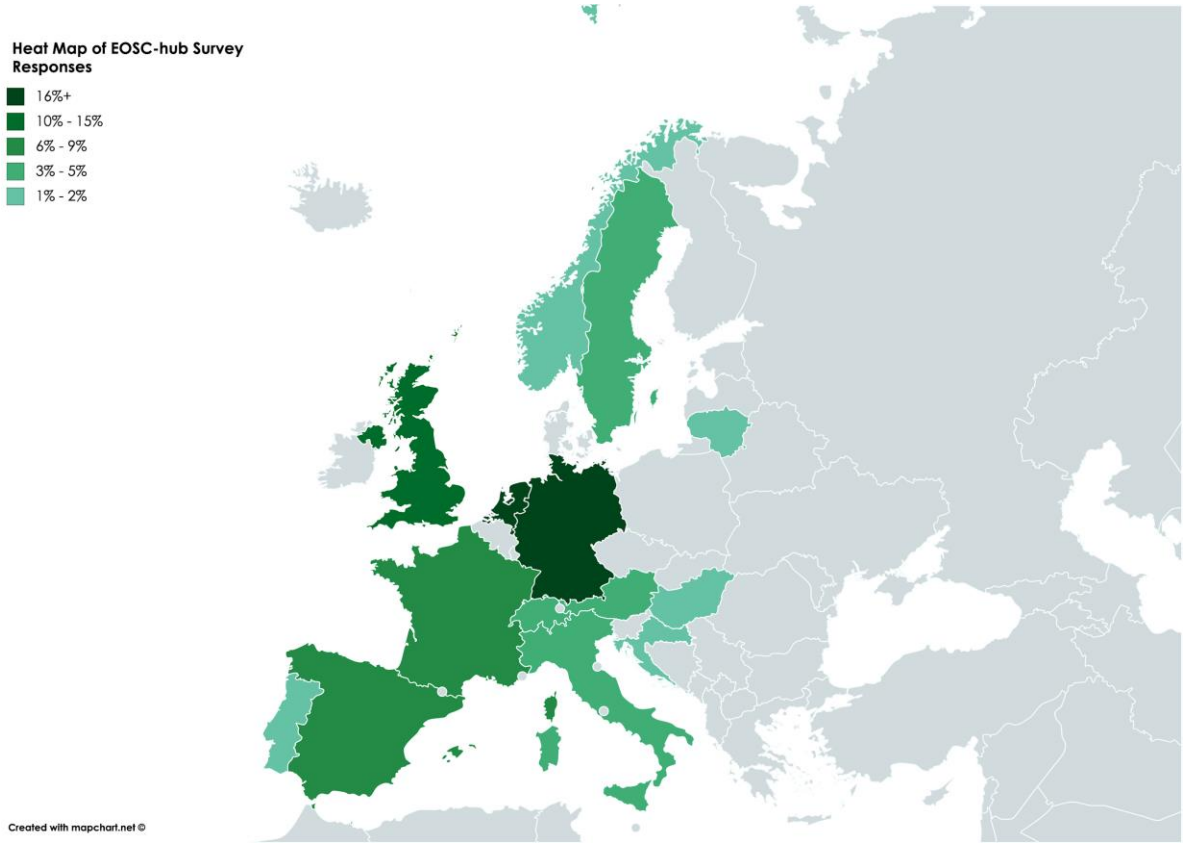


Figure 9: Geographical distribution of the respondents

The responses were also not spread evenly over the three response categories. Those respondents with a scientific focus (either broad or specific) accounted for 93% of responses with only the remaining 7% accounting for those in the management category.

Most respondents were representing single site entities, with only eleven declaring multi-site occupancy and only two of those occupying seven sites. Of those entities that ran e-infrastructure, eleven ran it distributed across more than one site and of those only three had the infrastructure running over five or more sites.

The vast majority of respondents represented large research service provider organisations (56%) or research organisations (26%). An incorrect filing as “large research provider” under “other” has been corrected. The other response categories were: H2020 Entities, Service provider institutes, public sector and other. Collectively these responses amounted to 19% of total responses. The private sector was not targeted in the survey, and consequently, no private sector responses were collected. Figure 10 portrays the distribution of respondents by category.

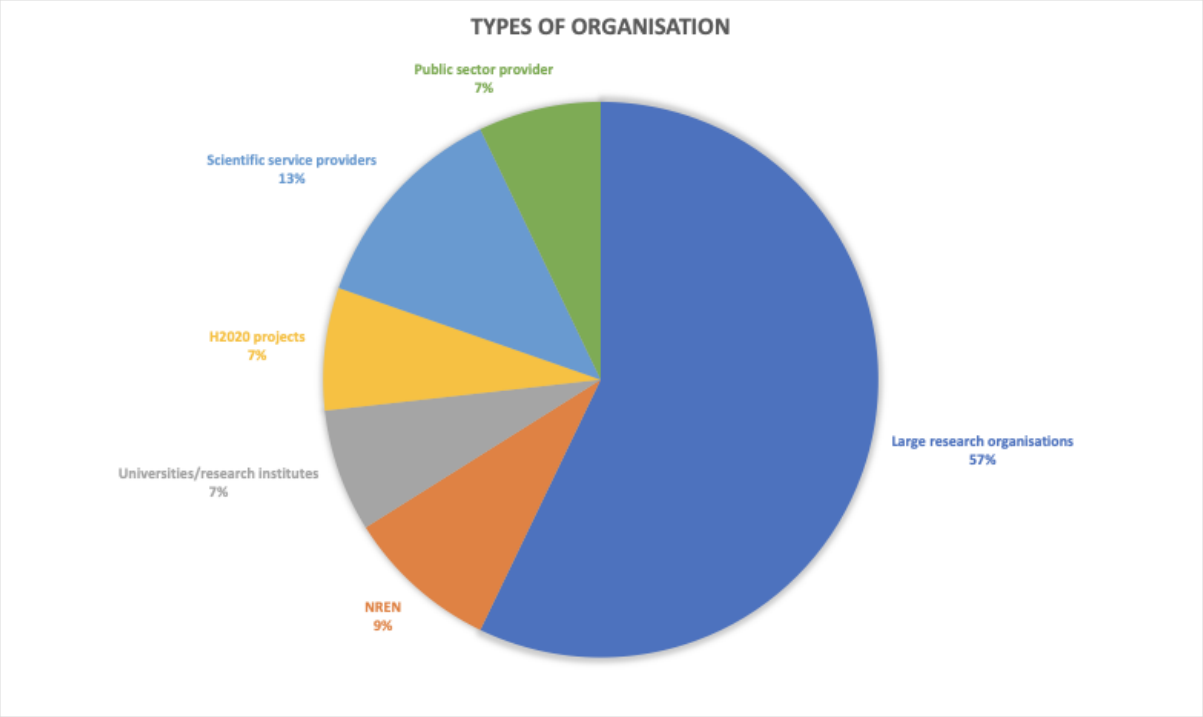


Figure 10: Distribution of respondents by organisation type

The coverage of the Interviews by scientific discipline is defined in Table X. No attempt has been made to interpret domains, only those using the same name have been combined. The structure is, therefore, specific to this study, it does not follow any convention.

Table 3: Organisations by scientific discipline (multiple choice possible)

Physics	**Biological science**	**Life science**
• General physics x 2 • HEP x 3 • Particle physics • Astro-particle physics • Nuclear physics • Nuclear engineering • Fusion Research • Space Science • Space physics • Astrophysics x 3 • Cosmology x 2 • Astronomy x2	• Biology x 2 • Biological data • Bioinformatics x 2 • Structural biology x 2	• General life sciences x2 • Medicine x 2 • Human health • ALS (Amyotrophic Lateral Sclerosis) • Genomics • Environment x2 • Ecology, biodiversity, Environmental science x 2 • Agriculture

● Radio astronomy ● Ionospheric and solar-terrestrial physics ● Photon science ● Plasma physics		
Social science ● Humanities ● Archaeology ● Cultural heritage x 2 ● Libraries and museums ● Arts and humanities - digital archives	**Physical science** ● Earth and Environment research ● Earth-system science ● Atmospheric Physics ● Atmospheric studies ● Climate and water research ● Chemistry x 2 ● Engineering	**Computer Science** ● General IT x 2 ● Data storage and data management services x 3 ● Informatics and data-intensive computing ● Networking infrastructure and Digital solutions ● Data/Information security, trust, privacy x 2 ● Infrastructure and service security, trust, privacy, Identity management ● NREN operations ● HPC
Non-scientific Disciplines ● One response related to services provided generically for all researchers at the University, while two more were related to university undergraduate teaching resources ● One response covered IT Innovation and Audit		

Detailed analysis of Sections 2 to 5 response data

Congruence of Feedback from Interview and Survey Responses in Sections 2 to 5

After this short observation, there follows a detailed analysis of the individual interview and survey responses to Sections 2 to 5. The degree to which the main points of each question were broadly aligned with each other in both the interview and survey responses was striking. There was some divergence lower in the ranking of answers, but these were also very much in the minority. In only five cases it was noted that there was any polarisation in the way the questions were answered. This makes them interesting in their own right and a closer investigation may reveal additional

insight. These questions were the following (NB question numbers refer to the master set configuration):

Questions providing polarised answers

Section 2: Research Users

- Q19 Are user groups producing data?
  - 16% of interviewees claimed not to be producing data.
  - All survey respondents claimed to be producing data.
- Q30 Authorised access to data:
  - Interviewees desire easier means to control access.
  - Survey respondents claim access control is not required.
- Q41 Main user needs of e-infrastructure:
  - Interviewees stated overall that "always on" is most important
  - Survey respondents: stated overall that "always on" is least important.
- Q43 Important supply side service preconditions:
  - Interviewees stated overall that "automated processing" is of highest priority, while "free and easy access" is a low priority (while affordability was a general issue).
  - Survey respondents stated overall that "free and easy access" is of highest priority, while "automated processing" is a low priority.

Section 3: IT Support

- Q62 Future procurement needs:
  - The majority (87%) of interviewees are looking for e-infrastructure service providers for long-term solutions.
  - The majority (67%) of survey respondents are looking for e-infrastructure service providers for short-term solutions.

Detailed analysis of Sections 2 to 5 data taken from the interviews

When discovering services, it might not be immediately clear to the user which organisation ultimately provides the underlying service, because of the way in which the pan-European e-Infrastructures offer services via aggregated frameworks and via national e-Infrastructures. As such, the results of the interviews and questionnaires have been analysed with this in mind.

Section 2: Research Users

User groups ranged from those disciplines frequently using HPC/HTC and working in large groups to a varied range of less common disciplines often working individually, such as: librarians, civil engineers and citizen scientists.

A minority of 16% of respondents claimed to be not producing data, of the rest (84%), 43% claimed to be producing core data. Furthermore, 71% of these data are considered operational. It is found that one of the biggest issues facing scientists is providing a home for their data. Those facilities not producing data were compute facilities that expected users to import data and export results into local storage. Elsewhere in the interview responses, the requirement to move very large data sets was cited as difficult, time consuming and in some cases expensive (even ignoring the costs of local

storage). No clear preference was provided which gave an insight into the motivations for respondents to move towards the use of e-infrastructure services. The most frequent response was "other". Reasons ranged from the important operational considerations to wanting to keep hold of their own hardware (historical justification related to funding and economic factors) through to lack of awareness of e-infrastructures and the perceived difficulty in using them. Of the answers to the "other" questions, scaling was cited slightly more often than "cost reduction" and "ease of management" as motivation to use e-infrastructures. Policy and external compliance issues were minor concerns and internal compliance was of no concern at all.

So far as cloud migration is concerned, the most common type is a publicly funded externally provided solution. The most common provider of such services is revealed to be via the GEANT Cloud Service, which is a catalogue of pan-European cloud framework agreements that gives access to discounted cloud services via service providers and institutions connected to the GEANT Network. However, many users perceive the cloud service to be distinct from the framework agreements. It is also found that the most highly utilised external repository of EU research data is Amazon Web Services. Users like these kinds of service because they are simple. The motivations for changing data culture were very evenly spread.

The vast majority of respondents answered that they were not implementing FAIR principles. Amongst those who were implementing FAIR principles, were the medical scientists who have a long history of needing to discover and reuse data. Justification for not moving towards FAIRification ranged from "cost of effort", "lack of explicit funding" and "lack of tool support" to "can't be bothered" and "no need because we do not reuse data". One respondent commented that: "*very often in recent projects, the EC has demanded that data are preserved while simultaneously failing to provide funding to achieve this requirement"*.

As for data management and processing tool support, the "other" category was the most common response. These answers covered a wide range of reasons why data management and processing are not carried out, often cancelling each other out, e.g. "tools/support provided on local infrastructure but not used" versus "no tool support provided so we can't do it". One response suggested that researchers should be paid to annotate their data so that they are findable. Of the other answers the provision of tool support was the leader in terms of required solution, while the next most popular solution was through the provision of access to existing repositories and registries. One respondent claimed that data reuse was "*a fantasy and that all data are specific to a particular experiment"*; another claimed that "*the wrong data were being FAIRified"* (without further clarification) and another required that the tools used to generate data should also be preserved alongside the data.

Data storage solutions in use throughout Europe cover the complete spectrum: from ad hoc hardware solutions, living under a researcher's own desk, to formally organised remote repositories accessed through expertly curated metadata registries. Aggregation was slightly favoured (by 60% to 40%); of the choices provided, most aggregation was carried out at disciplinary and sub-disciplinary levels.

Data access is almost exclusively seen to be under the control of the researcher generating that data. Some other actors are involved in more sophisticated responses, but they are in the minority

and represent the larger entities which have the resources to develop and implement policies and tools. The responses to the "other" category mainly relate to the desire for access through a controlled-access user-interface, rather than directly to the data where it is stored.

Repositories are commonly used (80%), registries less so (67%). Data analytics services are popular at 70%, while a range of "other" services were revealed to include anything from webserver-based analytics tools to data management, collaboration, search and AWS access.

e-infrastructure is widely used by researchers (53%) and is considered a welcome addition by 35% more. Only 12% consider it to be of no importance. Users' main needs of their e-infrastructure is quite evenly distributed with "always on" just in the lead with a weighted average[22] of 183 and the remaining ease of use/access/configure etc. factors having a similar weighted average in the range of 120-140. So far as expectations are concerned, users of e-infrastructure expect data management and processing issues to be taken care of and the repositories to be trusted. Other factors include the availability of SSO, security and privacy requirements taken care of, access to training (including SLA skills) and the availability of collaboration tools. A gap was identified in cataloguing, when it was noted that a data catalogue as well as a service catalogue is required. It was further noted that any new services should be easy to adopt and not result in changes to existing workflows. Finally, in one specialised case, data-pipeline design and deployment was seen as more important than pure data management and processing.

Considering the important services required by users, a slight majority (approx. 65%) considered automated processing to be the priority, other factors were evenly demanded at between 35%-40%. Other desired attributes were data (and data as content) findability and homogenisation over various national standards, services following the "just-use" principle and accommodate easy workflow integration. Some communities requested data management and processing tools and other specialised tools to allow data domain intersections to be carried out, and on top of all this, some also wanted data analysis as a service. Long term domain agnostic data storage was demanded, and training was also requested by many communities. In one case, it was noted that research leaders would like to access data that illustrates how their teams use the services. All of this must be made available at an affordable price.

Section 3: ICT Support

78.95% of respondents offer data services that are able to support FAIRification. These services include data networks and repositories (#1), analysis services (#2), real-time and sensitive data services (#5), HPC and Cloud services (#4 and #3). Other requested services include data fusion, HTC, storage (especially for very large data sets), security, privacy, specialised FAIR and FAIRification tools (this sits in clear contradiction of the views expressed by the users, in section 2, many of whom are unsupportive of FAIRification and claim they are unaware these services exist).

Most access to e-infrastructures services is by membership (or policy) of some kind (43%). Excellence (19%) is third after open access (24%). Market-driven and combinations of these tie in

[22] This weighted average is an algorithmically determined value, internally generated by the survey tool, that represents the ranking of responses to questions with non-exclusive answer selections.

fourth place at 14%. Other access models include free to a limit then commercial access, subsidised to a limit then commercial access, fair share by time, fair share by utilisation.

Most e-infrastructures are a hybrid of local, national and international resources. The hybrid at 37% far outstrips the leading competitor which is the pure international resource at 11%. Local and national resources (both at 5% each) clearly trail behind. Other structures include AWS, and individual small hardware solutions sitting under a researcher's desk. Some structures are, in fact, leased from commercial providers. Some entities maintain a local resource but contract-in elasticity on an ad hoc basis from either commercial or public sources, depending upon circumstances.

Most respondents (50%) felt their current e-infrastructure was suitable to meet their needs for the foreseeable future. Another 29% were unsure, leaving only 21% actually in the market for additional e-infrastructure services. Of those expressing uncertainty or definite need, 73% felt that Cloud services would suit them; another 20% were unsure, leaving only 7% least likely to use Cloud services.

Many respondents revealed that elements of their own infrastructure services are externally procured (73%), examples being 'cloud services', storage, back-up services and fibre optic network, and that this is for two main reasons: insufficient internal capacity and the need for flexibility and cost efficiency. Funding constraints and accounting procedures do not seem to be much of a problem. One respondent pointed out that, when procuring infrastructure, it is important to note that the specific requirements of low-latency and high bandwidth can be difficult to source from commercial service providers (with suitable price and levels of service) and for this reason such services tend to be fulfilled by specialist public service providers. Flexibility is key. Some respondents reveal here that they are already jointly pooling effort and purchasing resources to obtain better deals, although in order to do that sometimes a compromise has to be made in finding common denominators, which may result in only basic services being jointly procured. There is an expectation for commercial-grade services even from non-commercial suppliers and ideally respondents want to purchase environments that are able to adapt to the new technologies that are developed to be deployed on the infrastructure. It is worth noting that some justification was also provided for not procuring external services at scale: internal skills development, improved in-house support of research community, challenges in audit, licensing restrictions, retention of control, survival during financial squeezes, risk mitigation (against becoming reliant upon services that are taken out of service by the provider for various reasons) and integration flexibility.

As for the challenges that are being addressed through the deployment of e-infrastructure, federation of resources leads (47%), closely followed by specific-compute-at-scale (42%), long-term simple storage (37%) and commodity-compute-at-scale (32%). Behind these four front-runners interdisciplinary and cross-lingual issues lead with 26%, followed by support for distributed tools at 11%, with the three remaining issues tying in last place with 5%. Burst services are requested, as are services that fit well with NREN business models. Workflow-based integration is also a concern, as are data preservation and meeting demand. Some service-provider respondents have had to resort to withdrawing the marketing of their services in order to manage demand. Other respondents act more like sponsors or facilitators and enablers between users and the actual service providers, especially commercial service providers. Notable other requests include: provision of security and privacy services and EU-level data federation and curation. One respondent is exploring the utility

of GPU-based processing. Another recommends that EOSC-Hub includes all stakeholders in design and decision-making, as this approach at the national level worked well for them.

The vast majority of respondents (87%) are looking for e-infrastructure service providers for long-term solutions. There is a wide variety of motivations for doing this. In terms of the current delivery model, IaaS (+58%) was procured heavily in favour of PaaS (-70%) but SaaS was procured in a more balanced manner (+45%). The same outcome was broadly repeated when it came to which delivery models would be chosen in the future, with most respondents looking to acquire these new services within the coming year, if the procurement process was not already underway.

Few respondents seemed to know how many external providers were involved in the services they deliver. Those who could answer provided firm numbers ranging from zero to two and others offered "a few". AWS occurred again here, along with Google. INDIGO DC, DANS/SURF, GEANT and Sciebo were also mentioned; one respondent currently using Google is exploring the use of EGI in future.

Section 4: Procurement

Most of the procured e-infrastructures had a hybrid scope that ranged from a local and national to international (29%). Following this was the hybrid solution covering national and international resources (18%). Thereafter, national solutions were procured (12%), finally pure local and international services were procured, plus a hybrid of local and national services, all at the same rate (6%)

- Of the local infrastructures, most were a mix of owned and leased resources (45%), closely followed by solely owned resources (36%). No respondent claimed to operate on purely leased resources. Other "ownership" models included pay-as-you-go and in-kind contribution.
- Of the external resources most were publicly supplied (54%) with a runner up being a hybrid of public and private supply (38%). Only 8% were supplied with services from private sources alone.
- The basis of supply followed a similar pattern, in that subsidised access was the most prominent (58%) followed by free access (42%). Pure commercial access came next at 33%. An alternative model in use was a basic service offered for free but additional services had to be paid for out of grant income.

The service suppliers can be organised under two main categories: public or private. In no particular order, here are the suppliers provided by interview respondents.

Table 4: Service suppliers by category (public sector vs commercial)

Public	Private[23]
○ NRENs ○ DANS ○ Regional university collaborations ○ EGI ○ National nodes ○ Local resources	○ AWS ○ Google ○ CloudFerro ○ Serco-OVH ○ Orange Cloud ○ Atos ○ T-Systems

Payment models are especially varied. The most popular model is a bundled service invoicing (21%), followed by casual payments on demand (14%). Flat rate invoicing, periodic payments (e.g., subscription services) and ad hoc payments come in joint last with 7%. Alternative practices include: PAYG with minimum consumption, bundled purchase with needs-based allocation, "annual fee", purchasing credits for later use or for onward dissemination to users and pay-per-use. The preference appears to be pay-per-use, without an up-front fee.

Income streams for the responding entities were overwhelmingly from public sources, either as direct allocations (91.3%) of which some entities also take additional income through competitive grants (65.22%). Only 9% of overall income has private origin. Some reported that their parent university receives the grant, not the institute, and the university decides allocations of financial resource, which makes planning at their level very difficult. Certainty over income levels to improve planning is very much desired by the respondents.

Most funding streams come with rules attached (42%). Only 25% of income streams are rule-free. One reported problem is that the process to obtain public funds does not match the pace needed to respond to the user needs. Many entities have to adhere to stringent national rules as well as EU rules. National rules may restrict funding so as to benefit researchers in the same country. All respondents reported having to follow the EU Public Procurement Directive; some respondents, acting as intermediary agents, create agreements with their partner organisations within the framework contracts to facilitate the procurement process. Very few respondents reported that procurement rules prevent them buying services (9%), with the same percentage not knowing. 83% of respondents reported that their income levels allowed them to buy services adequate to their needs, leaving only 17% unable to meet their service needs through existing income, with 36% of these believing that through reorganisation they could deliver the required services within their budget.

The highest rated answer regarding the offer of support with procurement was "don't know" (58%) while 25% felt that such an approach would be a help to them. However, it's not clear how the term 'support' was interpreted, thus contributing to the inconclusive response and follow up with the specific respondents will be required. The conditions that limit suppliers were widespread with no obvious initial target. Answers ranged from 20% to 7% and covered all but two of the options, these

[23] Some of the private sector providers were selected in a tender to provide services in the Copernicus DIAS initiative: https://www.copernicus.eu/en/copernicus-dias-contracts-signed

exclusions being: a requirement for proprietary solutions and the need for home language support. The most common answer type was “other” and the effort that respondents are putting into trying to overcome their procurement problems is demonstrated within this answer.

Many types of procurement procedure are being used. Open (50%), restricted (40%), negotiated (30%), competitive dialogue (20%), and from these responses it is clear that some respondents are using more than one procedure. Again, the GEANT framework features as an exemplar. Most respondents (67%) follow a defined purchasing process. All respondents using negotiations in procurement feel they are important. Negotiations are used mainly to reduce costs and improve performance (83% overall). The procurement lead time ranges between one month and one year, and market consultations are employed by most respondents (75%). Life-cycle costing is considered by most respondents (70%) and some consider other factors such as standards compliance, life-cycle management and also forecast the longevity of service vis-a-vis quality of service. In terms of life-cycle costs, 63% of respondents consider: purchase costs, migration costs and maintenance costs, and 50% consider running costs and exit costs. It should also be noted that the life-cycle costs considered, are cited by those *not* procuring external services as justification for retaining services inhouse, believing that owning their own resources and running their own services is more cost effective. 33% of respondents consider the functional aspects of a service during procurement. One respondent also looked at non-functional aspects of security and privacy during their own procurement process. Many respondents feel that a uniform costing model for annualising storage costs would be useful in the procurement process.

The leading type of contract is the framework contract (31% explicitly declared, 76% implied from data) followed by a multi-annual contract (24% implied from data).

Insufficient data were collected about annual organisational budgets to be assessed here.

There are two reported main types of e-infrastructure cost: HR and provisioning (both 50%). Of these, most of the costs relating to FAIRification are deployed in infrastructure (50%) with 25% deployed in HR, the remaining 25% is deployed in specific metadata standardisation activities.

Fragmentation in the research data landscape is certainly seen as a problem (89%). The problem is addressed through collaboration and data management and processing skills development. Identification of EU service providers is perceived to be difficult. Favoured sources of economy are found mainly in moving towards interdisciplinary services from domain-based services (60%) while joint procurement gets 20% of the vote. An alternative, reported to be in use, is similar to the US model where industry is approached to build a solution for all parties. This is then deployed for the pre-agreed fee (20%). Some are prepared to consider joint procurement, but are reticent, desiring that such services should be basic and address privacy and security issues.

Section 5: EOSC-hub Collaboration

Twenty-one respondents expressed interest in cooperating with EOSC-hub and eleven respondents have procurement activities planned to run in 2019. A variety of constraints and suggestions were expressed in relation to possible collaboration.

Detailed analysis of Sections 2,3 and 5 data taken from the public online survey

Only one respondent completed Section 4. The responses have been removed from the analysis in order to prevent bias.

Section 2: Research Users

User groups ranged from the specific to the general: from those disciplines frequently using HPC/HTC and working in large groups to a wide range of general user types such as: "researcher" and "topic specialists" working in even larger groups.

All respondents claimed to be producing data. 75% of respondents were producing data throughout the life-cycle and 63% claimed to be producing core data. Furthermore, 75% of these data are considered operational. "Ease of management" was cited as the main motivation for moving to the use of e-infrastructure (88%), this was closely followed by "reduced cost of operation/ownership" and "overcoming scaling problems" (75% each). A specific point related to "technical and functional requirements matching" was mentioned in the "Other" category. Internal compliance issues were of minor concern, while external policy and compliance issues were of no concern at all.

So far as cloud migration is concerned, the most common type is a publicly funded externally provided solution, but the provider is not clear. However, as was the case in the focus group interviews, in this sample we find that actually, the main external repository of EU research data are private repositories. It is no longer clear if the main provider is Amazon Web Services. The inconsistencies in the way the question was answered may have been caused by the fact that such storage is paid for by individual researchers, directly from their public grant. Users like these kinds of service because they are simple. Specific problems related to data transport and portability for the HPC and big-data communities were cited as reasons for not exploring Cloud as a solution.

Once again, the motivations for changing data culture were evenly spread, although training and support were seen as slightly less of a problem than the other factors.

The vast majority of respondents answered that they were not implementing FAIR principles. In the vanguard are the medical scientists who have a long history of needing to discover and reuse data, but now they are joined by climate scientists citing the ESGF CORDEX project as justification. A single justification for <u>not</u> moving towards FAIRification was cited as "complete ignorance" which it is claimed is "by far and away" the biggest factor.

In the context of data management and processing tool support, analytical tools are clearly cited as the most common type of institutional data processing support. These are followed equally clearly by data description and formatting tools and controlled vocabularies. Help desks and disciplinary repositories follow on with "other", which revealed that access to some data management and processing tools and services is provided through research communities rather than the host institution.

Data storage solutions in use throughout Europe cover the complete spectrum: from ad hoc hardware solutions, sited under a researcher's own desk to formally organised remote repositories accessed through expertly curated metadata registries. Of these, 75% are aggregated, while 25% are not aggregated. Of the choices provided, most aggregation was carried out at disciplinary and interdisciplinary levels (33% each), while disciplinary data account for 17% of aggregations. It was

pointed out in the “other” category that some scientific domains aggregate according to their temporal and geophysical characteristics, while other aggregation requirements vary depending upon data type.

Data access is generally seen to be under the control of the researcher generating that data. Some other actors are involved in more sophisticated responses, but they are in the minority and represent the larger entities which have the resources to develop and implement policy and tools. The responses to the “other” category point out that open data requires no access control.

Repositories are commonly used (88%), registries less so (63%). Data analytic services are also popular at 50%, while a range of “other” services were revealed to include HPC, various levels of OGC service and generic compute services.

e-infrastructure is widely used by researchers (75%) and is considered a welcome addition by 13% more. Only 13% consider it to be of no importance. Users’ main needs of their e-infrastructure are not evenly distributed. In the lead is “does not require ICT expertise” and “help when needed” with a weighted average of 163 each. Next comes “easy to configure” with a weighted average of 150 and the remaining “always on/easy to access/easy to use factors having a similar weighted average in the range of 100-113. So far as expectations are concerned, users of e-infrastructure significantly expect data management and processing issues to be taken care of and for the repositories to be trusted. Other factors include ease of use and configuration in domain-relevant languages.

Considering the important services required by users, a large majority (100%) considered free and easy access to e-infrastructure services to be the priority. The next highest demand was for easy integration of legacy data (63%). The three other factors declined from 38% (automated processing) to 13% (cross-disciplinary FAIRness) in popularity. Other desired attributes were: Big Data compatibility and metadata integration. There is some interest (50%) in accessing other data services via e-infrastructure but no further details were given.

Section 3: ICT Support

84.62% of respondents offer data services that are able to support FAIRification. These services include data networks and repositories (#1), analysis services (#2), HPC (#3), real-time and sensitive data services (#4) and Cloud services (#5).

Most access to e-infrastructures services is by membership (or policy) of some kind (77%). Excellence (23%) is third after open access (38%). Market-driven and combinations of these tie in fourth place at 15%. In the other category, an access model offering open access which is limited to employees and students is described, but this is a specialisation of the open category and that category score reflects this.

Most e-infrastructures are a hybrid of local, national and international resources. This hybrid at 54% far outstrips the leading competitor which are pure local services at 23%. National and international resources (both at 8% each) clearly trail behind. Other structures include hybrids of local and commercially (or community) leased services which are used to contract-in elasticity on an ad hoc basis.

Most respondents (54%) felt their current e-infrastructure was suitable to meet their needs for the foreseeable future. Another 23% were unsure, leaving another 23% actually in the market for

additional e-infrastructure services. Of those expressing uncertainty or definite need, 33% felt that Cloud services would suit them, while many more (67%) were unsure. However, no respondent was actively hostile to Cloud services.

Many respondents revealed that their own infrastructure services are externally procured (69%) and that this is for two main reasons: insufficient internal capacity and the need for flexibility when procuring additional services. Funding constraints also seem to present a lower level problem at 33%. Accounting procedures are not seen as an issue. However, when outsourcing is considered, some unforeseen pragmatic issues are also brought into the calculation. Some respondents view e-infrastructure as an opportunity to increase the prospect of obtaining EC co-funding, while others note that they are forced to use e-infrastructure because the instruments their experiments depend upon are often located at remote sites.

As for the challenges that are looking for solutions through the deployment of e-infrastructure, federation of resources leads (67%), followed by long-term preservation of data for results replication (50%). Five challenges then tie at 33% (specific-compute-at-scale, long-term simple storage, support for distributed and centralised tools, and sourcing expert support not available locally). The remaining three categories tie for last place at 17% (commodity-compute-at-scale, and one respondent citing long-term data preservation to achieve Green / Gold publishing standards, and interdisciplinary and cross-lingual data identification purposes). An unforeseen challenge, intrinsically addressed by e-infrastructure, was noted in the “other” category and this is in its use as off-site data-centres with redundancy and back-up services able to support an organisation in the event of disaster recovery.

The respondents to the survey hold very different views to the focus group interviewees in terms of the longevity of the services they require. The vast majority of respondents here (67%) are looking for e-infrastructure service providers for short-term solutions. However, when considering the motivation for this preference, the leading responses (50% each) both relate to long-term problems (Long-term SLA support and customisation partnerships). In terms of current delivery model, IaaS (+50%) was procured heavily in favour of PaaS (-100%[24]). SaaS procurement was not favoured strongly (+33%). The rationale for future Cloud services revealed a slightly different outcome. IaaS was still in the lead (+100%), with PaaS coming second at +67%, followed by SaaS at +56%. Most are looking to acquire these new services over a range of lead times, from a few months to a few decades.

The number of external service providers ranges from one to eleven, but this is taken from a sample size of only nine responses. Unsurprisingly, AWS is named as a provider, along with Google GCP and MS Azure. EOSC-Hub was mentioned, as were a number of national public providers (e.g. DANS, UKCloud, CSC, etc.). Some institutes still run their own private services. LOFAR relies on a very specific set of services hosted at named sites, all are critical to the mission. Oddly, LinkedIn was

[24] The negative percentage describes a compounded negative response. Across the three related questions half of respondents **did** procure IaaS, all **did not** procure PaaS and one third **did** procure SaaS. The three questions were structured as Yes/No to each type and combining the responses any other way made less sense than this method. -100% was chosen over 0% because an active decision was made by respondents to vote "no", it does not arise out of the passive result of no one voting for this selection.

mentioned as a service provider and one respondent explicitly stated they did not want to answer this question, without saying why.

Section 4: Procurement

The analysis from this section is removed because the questions were very poorly addressed in terms of the number of responses. One question was answered by no one, one was answered by three respondents, the remainder were answered by only one respondent.

Section 5: EOSC hub Collaboration

Eighteen respondents addressed EOSC hub collaboration. Twelve (67%) respondents expressed interest in cooperating with EOSC hub and eleven have procurement activities planned to run in 2019 and 2020. (It is not clear if these eleven are a subset of the twelve respondents expressing an interest in collaboration, or if they are an intersection of those interested and those not interested).

Of the ten respondents who left their name and identified their organisation, none left contact details. Respondents would make use of EOSC-hub support in the following manner:

- Conditional cases
    - Interest is based on what EOSC hub services can offer
    - Efficiency and time criticality: if it takes a week between asking for resources and being granted access to them (particularly for compute), then existing cloud provider is better
        - This is less important for storage
    - Data security is of paramount importance
        - It must be convincingly demonstrated.
- Specific cases
    - Running workflows
    - Methodology alignment for on-boarding, SLA framework, reporting, access alignment, etc.
    - Evaluation of e-infrastructure services for the implementation of a RI.
        - Support in selection of services and evaluation would be needed
    - Harmonise with eInfraCentral and OCRE
    - Investigate if procurement of cloud resources can be simpler and more cost transparent while retaining the same level of security and trust as on-prem systems permit.
- General cases
    - Application in projects like any other local of regional e-infrastructure.
    - Involvement in the design of EOSC services.
    - Ultimately interested in ways to achieve best value for effort/investment.
    - Development of services.
    - Integration with core services.

Appendix IV. The interview questions

In this section, we provide the list of questions that have been used during the interview phase.

A. Interviewee/s identification

1. Organisation name:

2. Country:

3. EU / non-EU

4. Stakeholder category:

1. Organisations serving large organised research communities
    - EIRO
    - Big lab
    - ESFRI
        - RI cluster / RI projects / RI landmarks
    - International research collaborations
2. Organisations serving small research groups / individuals
    - Universities
    - Research centres
3. National Research and Education Networks (NREN)
4. H2020
    - H2020 project
    - EOSCpilot demonstrators
5. Scientific-service provider institutes
6. Private sector – business organisations
7. Public sector

5. Scientific discipline focus:

- Broad / generic,
- Specific, please enumerate which one/s:

Person 1a name: Person 1a role:

Person 1b name: Person 1b role:

Person 2 name: Person 2 role:

Person 3 name: Person 3 role:

B. User base needs, drivers and constrains

1. Who are the main user groups of research e-infrastructure and services within your organisation? Can you estimate their respective numbers? Do you see yourself serving the same or wider community in the future?

- Individual researchers or groups
- Which scientific disciplines
- Data providers or data users
- Research or technology development users

2. Are your user groups actively managing data on local e-infrastructure?

3. Is the data produced throughout the research lifecycle regarded as a core and valuable output of the research process? Is research in your institution producing operational data[25]?

4. What might drive your organization to make increased use of e-infrastructures, especially cloud-based ones? Either in-house or external.

Would the following factors possibly change your organization's data culture? (i.e. the frequency with which e-infrastructures would be used) (please rank them according to their likely importance):

- Policy (e.g. in line with upcoming update to the Commission Recommendation on access to and preservation of scientific information),
- Regulation (e.g. EC plan to make FAIR data the default option under Horizon 2020),
- Financial support (e.g. EC plan to make FAIR data related costs eligible for reimbursement under H2020 projects),
- Other incentives (e.g. active data management is a criterion in the regular evaluation of your organisation; career rewards for researchers reflect the value of FAIR data),
- Training and support,
- Adequate tools and services.

5. To what extent are the FAIR data principles already pursued by and implemented in your organisation?

- we have been implementing and using FAIR or similar principles for years,
- we are beginning to implement part of or all the principles,
- we are not implementing them due to
    - not subscribing to data sharing or automated workflows,
    - lack resources,

[25] i.e. datasets intended for multiple use and re-use. E.g. data from astronomical surveys, genomic data or large economic datasets as opposed to experimental data specifically generated to solve individual research questions (and therefore less likely to be of more general value)

- already having another satisfactory way to share data.

6. What type of data managing/data processing services are available at your institution?

- Provision of analysis tools running on appropriate computing infrastructure
- Provision of timely and effective support/help desk services to install, run, access, connect, etc. the tools or services
- Provision of tools to facilitate data description and formatting (e.g. to make data FAIR)
- Support from experienced data curators
- Sourcing access to disciplinary data repositories and registries
- Other, specify:

7. How is data currently stored and curated / maintained in your organization? Is data currently aggregated to any degree? At which level of aggregation?

- Generic,
- Cross-disciplinary,
- Disciplinary,
- Sub-disciplinary.

How and by whom can the data be accessed?

8. What e-infrastructure services are currently used in your institution and to what purpose? e.g: repositories, registries, data analytics and other services?

What is the importance of e-infrastructure within your wider research infrastructure?

- widely used by the research community in its daily research work
- welcome addition to physical facilities, but only moderately used (e.g. by certain disciplines)
- not a major one.

9. Which are the main needs of your user base vis-à-vis research e-infrastructure? What preconditions, in your experience, are necessary for their adoption of e-infrastructures as data management tools?

Are the following conditions important?

On the user side:

- Data management skills
- Trust in repositories and registries

On the service side:

- Automated processing services
- Easy integration of legacy data
- Free and easy access to such infrastructure and services

- Cross-disciplinary and cross-geographical FAIRness

10. What kind of data are you user groups interested into but cannot access at the moment? What about data services? Why?

C. Current demand of research e-infrastructure and services (mainly for IT dept.)

1. Do you provide infrastructure and services that enable analysis of data and data storage and retrieval (especially as FAIR data)? If yes, what kind of research data access, e-infrastructures and related services do you currently provide to your user base?

- Use of e-infrastructure, data networks and repositories
- Services for the analysis and reuse of research data (e.g. analytics, fusion, mining, processing, data conversion and annotation (e.g. for FAIR data), etc.
- Services to manage sensitive or real-time data
- Super-computing facilities
- Cloud based, on-demand HPC services
- Others, please specify

2. What is the access policy to your e-infrastructure and data services for users?

- Excellence-driven (access exclusively based on peer review)
- Market-driven (paid access)
- Membership/policy based (access by affiliation)
- Wide (open to everyone, free access, etc.)
- Combination, please specify

3. How is your current research e-infrastructure structured? (e.g. single-site vs. distributed system with national nodes; cloud-based; everything in-house vs. external suppliers).

Is it part of a wider international ecosystem?

Will your current e-infrastructure meet your needs for the foreseeable future?

Could cloud-based infrastructure be used for this?

4. What part of the e-infrastructure services do you procure externally?

5. Why would you outsource e-services?

- insufficient internal capability and/or capacity
- flexibility and cost efficiency
- funding constrains,
    - Accounting pressures e.g. CapEx vs OpEx
- Other, please specify

6. What are your main challenges that you are trying to solve with the services you are procuring or consider procuring?

- Federate distributed sources of data
- Computing at scale based on commodity infrastructure (e.g. conventional Cloud services)
- Computing at scale based on specialised facilities (e.g. HTC, HPC, etc.)
- Long-term preservation of data
    - Simple preservation
    - Includes the requirement to preserve everything to allow the replication of results
    - Green / Gold publishing of results data
- Short/medium term analysis
    - Central
    - Distributed
- Source external expert support
- Interdisciplinary and cross-lingual identification of relevant research data
- Other, please specify

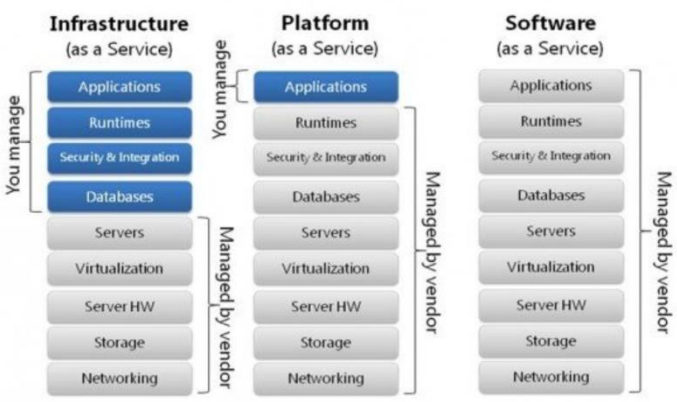
7. Are you looking at service providers and suppliers for short-term or long-term solutions? Why?

- long-term support to sustain an SLA for a defined service offer for many years
- long-term partnership to design/develop and support a customized solution
- short-medium term to support a specific project
- other, please specify:

8. When you procure cloud services, what kind of delivery models do you purchase?

If you would procure cloud services, what kind of delivery models would you choose? Why?

- IaaS Infrastructure as a Service
- PaaS Platform as a Service
- SaaS Software as a Service



9. What is the time schedule to acquire the services / resources and what is the related capacity?

Type of service	Quantity (order of magnitude)	When needed (Year)
Add examples: e.g. IaaS cloud		

10. How many external providers of research e-services do you currently hire? Which of these are most needed and why? (e.g. legacy or proprietary system, volume, cost, etc.)

D. Procurement and finance

1. What type of e-infrastructure and services is your organisation currently using? Why?

- Own, national, international?
- If external, public or private?
- On a free/subsidized or commercial basis?

Who is/are your actual supplier/s of e-infrastructure and services?

2. What are the current payment models that your organisation use for e-infrastructure related services?

3. What are your funding streams?

- public budget allocations
- national and EU funds – on a competitive basis
- private donations
- own revenues
- other, please specify

4. In your organisation, which procurement rules are applicable? (e.g. EC Directive on procurement, other national regulations) Do the funding streams have an impact on the procurement process/rules?

5. Are the current funding streams allowing you to acquire e-services? If not, how could they be organised in order to allow you to do so? (e.g. Would it help funding to be structured in a way that it can be used to buy through something like EOSC hub, i.e. without formal tender?)

6. Are there any other conditions that limit your choice for the suppliers?

- Certifications schemes and standards (e.g. ISO, CoreTrustSeal, other FAIR accreditation, etc.)
- Requirement for specific proprietary solutions (e.g. suppliers with unique IP)
- Customer support 24/7 in own language
- Geography (e.g. based in the EU only)
- Political or environmental issues,
- Proven technology,
- Open source,
- Other, please specify.

7. Which procedure is your organisation using when procuring e-services?

- Open,
- Restricted,
- Negotiated,
- competitive dialogue,
- other, please specify.

8. How important are negotiations with suppliers when purchasing e-services? Do you negotiate rather on the results to achieve, the price or the methodology to achieve the results? What is your average lead time for procurement?

9. Prior to the launch of the procurement process, do you organise market consultations to engage with potential suppliers and to get a clear idea of the state-of-the-art? Which sourcing techniques do you use?

10. Does your organisation consider life-cycle costing when conducting procurement of e-services? Which costs do you take into consideration (exit costs, maintenance costs, migration costs, other please specify)?

11. When purchasing e-services do you use functional specifications rather than technical specification to make sure that the suppliers will offer you an adequate/expected solution?

12. What types of contracts do you already conclude?

- Framework contract or agreement, which type (terms and conditions agreed for the duration of the agreement OR price are determined by a ‘mini-competition between the selected suppliers)
- Single task contract
- Multi-annual,
- Other, please specify.

13. Can you give an indication of your current costs / budget linked to research e-services (e.g. data management, curation and publication)? (e.g. capital investment, annual budget, etc.)

14. What is your organisation's total annual research budget[26]? (Alternatively, ask for an estimate of the percentage of their budget that goes into e-services)

15. What are your current main types of costs related to e-infrastructure and data sharing activities?

- Human resources
- Infrastructure
- Overhead
- Legal costs
- Use costs

Which do you estimate to be the main costs involved in making research data FAIR?

16. Is the existing fragmentation of the research data service landscape a problem for your organisation and if so, how do you manage this challenge? How would you try to achieve economies of scale in the e-services you provide to your user base?

- Shift from the use of a number of disciplinary thematic focused services to the acquisition of access to interdisciplinary e-infrastructures
- Procurement of joint / bundled services
- Joint procurement together with other organisations (e.g. at national, regional or thematic level)
- Other, please specify.

17. If you do/would undertake procurement of data services jointly with others, what challenges to this have you faced / do you envisage?

E. Collaboration with the EOSC-hub project

The EOSC-hub project has a funded activity to develop procurement and purchasing frameworks that can help publicly-funded organisations in optimising their sourcing activities. The goal is to create, in the context of the European Open Science Cloud, a number of procurement pilots. In this context, please tell us.

1. Are you interested in receiving support from the EOSC-hub project procurement experts with expert consultancy, specialised procurement toolkit, etc.?

2. How would you make use of such support services from the EOSC-hub project procurement experts, if they were to be available to you? Do you have a specific procurement e-services activity happening in 2019 that could be benefiting of these?

[26] This is a question cross-referencing the answer to the previous one and benchmarking it against the Final Report of the High Level Expert Group on the EOSC that estimated that on average about 5% of total research expenditure should be spent on properly managing and stewarding data in an integrated fashion.

Appendix V. Organisations who participated in the market research

People representing the following entities provided feedback to the questions developed. Please note that in some cases there are more than one response from some entities. These responses were checked to ensure that no duplication was taking place and allowed to stand because the represented the views of either: different parts of the entity's organisation or were provided by persons occupying different roles within the entity's organisation.

Organisations that participated in the Interviews

1. The Istituto Nazionale di Fisica Nucleare
2. Port d'Informació Científica
3. Kormányzati Informatikai Fejlesztési Ügynökség
4. The German Network for Bioinformatics Infrastructure
5. Centre for Synthetic and Systems Biology, University of Edinburgh
6. Distributed Research utilising Advanced Computing
7. Universidade de Évora
8. University Jaume I
9. Culham Centre for Fusion Energy
10. Hogeschool Inholland
11. Istituto Centrale per il Catalogo Unico delle Biblioteche Italiane
12. Amsterdam Universities Medical Center
13. Vast Lab
14. SURFcumulus
15. Institute of Applied Computer Science, Vilnius Gediminas Technical University
16. Laboratory of Instrumentation and Experimental Particle Physics
17. Laboratoire d'Infochimie Université de Strasbourg
18. Digital Research Infrastructure for the Arts and Humanities
19. European Incoherent Scatter Scientific Association
20. Consortium of European Social Science Data Archives)
21. European Space Agency
22. Biobanking and BioMolecular resources Research Infrastructure)

23. LifeWatch
24. University of Groningen
25. Project MinE
26. The European Synchrotron
27. European Bioinformatics Institute of the European Molecular Biology Laboratory
28. Environmental Research Infrastructures building FAIR services
29. Worldwide e-Infrastructure for NMR and structural biology
30. Research Centre Juelich:
    - IT-Services Department
    - Purchasing division
    - Institute of Energy and Climate Research: Troposphere

Organisations that responded to the online survey

1. French Institute of Bioinformatics (Elixir-FR)
2. Royal Holloway, University of London
3. Netherlands eScience Center
4. Universitat Politecnica de Valencia
5. Kormányzati Informatikai Fejlesztési Ügynökség
6. *Anonymous response*
7. Research Centre Juelich
8. Science and Technology Facilities Council
9. Utrecht University
10. SWITCH
11. Netherlands Institute for Radio Astronomy (LOFAR infrastructure)
12. Naturalis Biodiversity Center
13. Karlsruhe Institute of Technology
14. European Bioinformatics Institute of the European Molecular Biology Laboratory
15. Fontys University of Applied Sciences
16. Umweltbundesamt GmbH
17. SURFsara
18. European Incoherent Scatter Scientific Association

19. Karlsruhe Institute of Technology
20. Deutsches Elektronen-Synchrotron
21. University of Vienna / ACOnet
22. European Incoherent Scatter Scientific Association
23. United Kingdom Atomic Energy Authority
24. SWITCH
25. Technical University Delft
26. European Organization for Nuclear Research
27. InterMine
28. Zagreb National and University Library

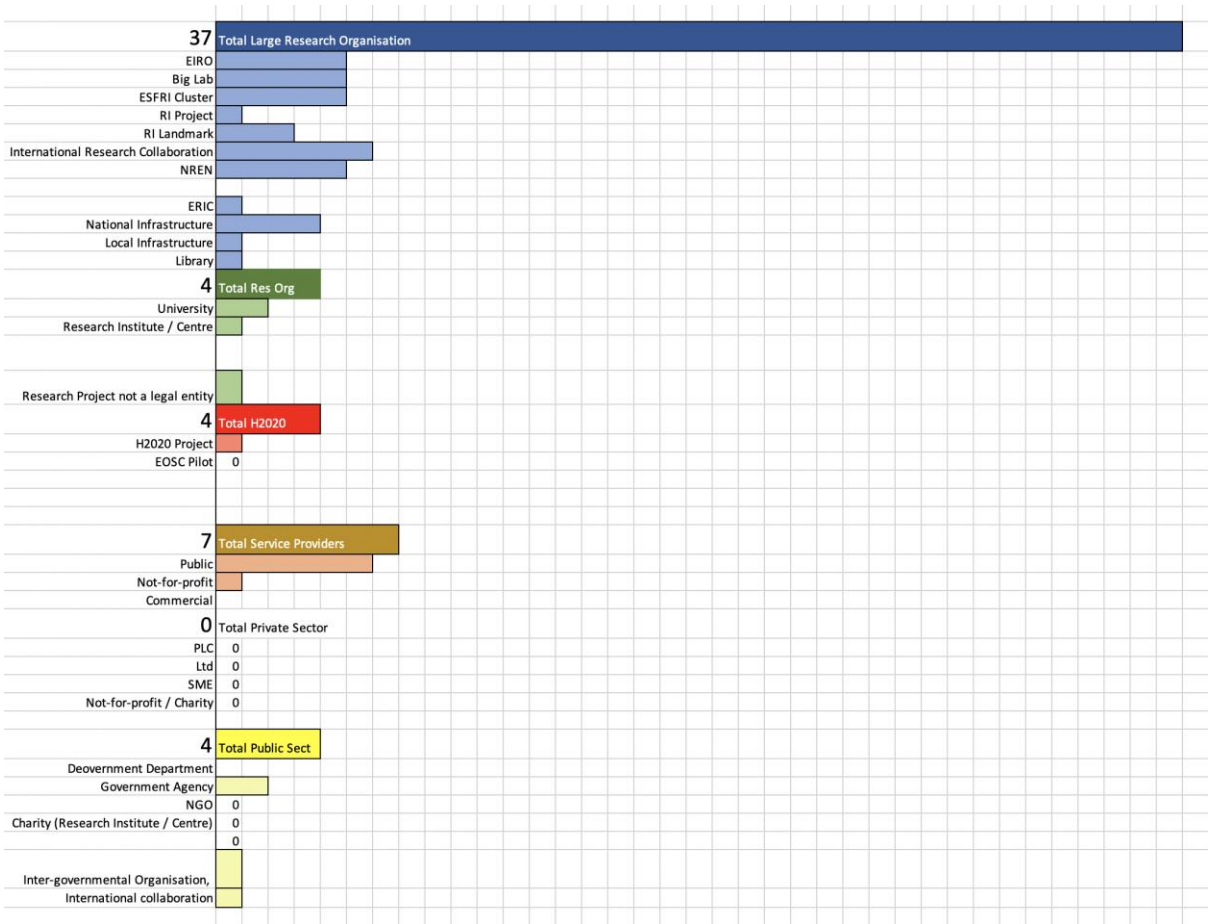


Figure 11: Respondents by stakeholder category

Appendix VI. Lessons Learnt

At the end of this document we would like to reflect on the process and the methodology we adopted, and review improvements if such an exercise were to be repeated. Indeed, we have found this exercise useful and insightful, and with better preparation we think that it would be valuable to repeat every couple of years. However, here is what we learnt:

- Further work needs to be done to reach all the EU member states. In this exercise, there are some notable gaps and we underestimated the effort we would need to really reach everyone (e.g. an idea is to rely on country-based governance boards to reach all countries, for instance the newly established EOSC governance body or existing bodies such as e-IRG delegates).
- Better engagement of higher management of the potential demand-side organisations is needed and these need to have clear knowledge of procurement procedures and strategies
- Terminology used in surveys in the future could be simpler (e.g. 'IT Services' instead of 'e-Infrastructures' or, for example, asking about generic needs rather than FAIR)
- Engagement of more varied roles of interviewees would be needed. For example, in this case we clearly need more procurement and finance people.
- We underestimated the reticence to share of many people and sometimes, therefore, some questions looked as if they were just answered for the sake of answering and not enough depth was given (mostly applies to the online survey)
- We could have been more disciplined and left some questions out since the survey contained too many questions as we were trying to cover all bases.
- Resource availability and expertise from the EOSC-hub partners has been a challenge, mostly due to the underestimation of the effort needed to really engage with the right people from the right organisations.
- For the future it would be good to be connected with other networks to get a better distribution of answers and not just leave this in the small EOSC- hub community.
- We could have considered splitting the deliverable into two parts, publishing the results of the survey first, with the business models analysis and recommendations following later.