Research Infrastructures & NRENs Requirements for Commodity Cloud Services

Version: Draft 0.1
Main authors : João Fernandes (CERN), Bob Jones (CERN), Marion Devouassoux (CERN)
Co-authors:, Jakob Tendel (DFN)
Acknowledgements: David Heyns (GÉANT)

Report produced in the context of OCRE WP2/T3:
"*Requirements Collection for Cloud Commodity Services*"

Table of Contents

Abstract

Leveraging lessons learned on previous activities to access commercial cloud services, CERN is leading the task “requirements collection for commodity cloud services” in the Open Clouds for Research Environments (OCRE) project[1]. In a two-fold activity, CERN built on the work performed during the award-winning HNSciCloud project[2], introducing voucher schemes to research, collaborating with the European Council of Doctoral Candidates and Junior Researchers (Eurodoc)[3] and Marie Curie Alumni Association (MCAA)[4] in order to collect requirements for individual researchers (OCRE Wave 1)[5]. The second activity relates to the definition of the requirements for the establishment of a “Buyers Group” model, derived from Pre-Commercial Procurement exercises, where a lead buyer organisation aggregates demand and financial pre-commitments, acting on behalf of a set of organisations or a community. These activities also build on the results of a demand-side market research to understand the needs and level of demand for digital services for the research community, in the context of the EOSC. The market analysis is documented in the EOSC-hub D12.1 deliverable “*Procurement requirements and demand assessment*”[6]. The combination of these activities contributed to the establishment of the baseline for requirement collection both for Individual Organisations (OCRE Wave 2) and the aggregation model based on Buyers Groups (OCRE Wave 3). This document will be complemented by a quantitative analysis of the requirements presented.

Introduction

In 2016, the European Commission proposed the creation of the European Open Science Cloud (EOSC), with the aim of offering Europe's 1.7 million researchers and 70 million science and technology professionals a platform to store, share and re-use their data across disciplines and borders. The minimum requirements for commercial or non-commercial entities that wish to participate in the EOSC as service providers have been discussed in the EOSCpilot project[7]. The introduction of innovative business models implies in some cases aggregating the committed demand of contracting authorities purchasing from suppliers, which potentially brings economies of scale and process efficiencies.

This model has been used in pre-commercial activities such as HNSciCloud and ARCHIVER[8]. Building on this experience, CERN is participating in the Open Clouds for Research Environments project (OCRE), that aims to accelerate cloud adoption in the European research

[1] https://www.ocre-project.eu/
[2] https://www.hnscicloud.eu/
[3] http://www.eurodoc.net/
[4] https://www.mariecuriealumni.eu/
[5] https://zenodo.org/record/3564668#.Xfd95dZKjUI
[6] http://bit.ly/EOSC-hubD12_1
[7] https://eoscpilot.eu/
[8] http://archiver-project.eu

community, by bringing together commercial cloud providers, Earth Observation organisations, and the research and education community. OCRE will encourage the consumption of commercial cloud services by research organisations as well as by individual researchers using different access mechanisms. CERN is leading the task “Requirements Collection for Commodity Services” under the Work Package “Community Outreach and Requirements”.

After making an overview of the lessons learned from past activities and listing the key elements in the process to allow the formation of buyer groups to in order to establish an aggregated model for purchase of commercial cloud services, with the main advantages associated, this paper provides an analysis of the technical and organisational requirements these services need to fulfill in order to be of significant added-value for research activities and, more generally, for the education community. The next steps and implications for the EOSC are summarized in the last section.

Lessons Learned from Current and Past Experiences

Buyer Groups in HNSciCloud & ARCHIVER

In general terms, procurement law allows buying organisations to establish commercial and contractual models that reduce the need for repetitive tendering for similar goods and services. These may take the form of framework agreements, for example, where one contracting authority, operating as a lead buyer establishes framework agreements for use by a number of other contracting authorities as in the case of Pre-Commercial Procurements (PCP) instruments.
HNSciCloud was a European Pre-Commercial Procurement (PCP) project aiming to establish a hybrid cloud platform combining commercial services with existing publicly funded on-premise resources, to support the deployment of high performance computing and big-data capabilities for scientific research.
The HNSciCloud Buyers Group was formed by ten research organisations (CNRS, CERN, EMBL, ESRF, DESY, IFAE, INFN, KIT, SURFsara, and STFC) from across Europe, purchasing services on behalf of seven ESFRI research infrastructures:

- Euro-BioImaging: European Research Infrastructure for Imaging Technologies in Biological and Biomedical Sciences
- CTA: Cherenkov Telescope Array
- BBMRI: Biobanking and BioMolecular resources Research Infrastructure
- ELIXIR: A distributed infrastructure for life-science information
- ESRF: Extremely Brilliant Source
- European-XFEL: European X-Ray Free-Electron Laser Facility
- HL-LHC: High-Luminosity Large Hadron Collider

In a competitive R&D process specific to PCP, two consortia were awarded contracts for the pilot phase:

- The consortium led by T-Systems, with IaaS capacity provided also by T-Systems
- The consortium led by the RHEA group, with IaaS capacity provided by Exoscale (A1 Telekom Austria Group).

Within this phase, the two consortia developed a hybrid IaaS cloud platform linking together commercial cloud service providers and several research organisations' in-house IT resources via the GÉANT network.
CERN was the lead procurer and the total procured IaaS capacity needed for the R&D activity was made available in a combination of delivery models: reserved capacity and on-demand services in the form of vouchers to be distributed to end researchers from the communities represented by the research organisations participating in HNSciCloud.
The ARCHIVER project is following the same Pre-Commercial Procurement approach for cloud-based long term archiving and data preservation services. While HNSciCloud focused exclusively on IaaS, ARCHIVER is focusing one level up the stack (SaaS), performing R&D on innovative services on top of commodity IaaS. The ARCHIVER Buyers Group is formed by four research organisations (CERN, DESY, EMBL-EBI and PIC) and twelve preferred partner organisations, representing a diverse range of scientific domains and use cases.
The model is similar to HNSciCloud where these organisations support the deployment of advanced research use cases from four ESFRI infrastructures: HL-LHC, ELIXIR, CTA and European-XFEL. Tender launch will happen in Q1 2020.
The aforementioned examples demonstrate that within an established framework, organisations representing a group of buyers (e.g. for a certain scientific discipline and/or with similar needs) or even a central purchasing body, can act as Lead Buyer and main procurer, becoming the contracting authority, taking responsibility both for coordination of the contracting authorities and execution of the procurement procedure.

There is a considerable number of advantages in aggregating demand through a Buyers Group:

- The procurement process is done by one organisation, the lead buyer on behalf of several organisations in different scientific domains and European countries, allowing the procurement to become faster, simpler and more cost-efficient.
- Aggregation increases demand as some organisations are unlikely to engage in isolated procurement activities.
- Higher volumes steaming from the demand aggregation generate economies of scales leading to volume discounts and preferential terms.
- All members benefit from past experience and acquired expertise in selecting digital services for research, assessing services against key requirements, such as data sovereignty, data protection and security.
- In the context of H2020 procurement projects, the establishment of Buyer Groups allows its members to benefit from procurement funds without being partners of the projects

(partners are the only entities able to claim costs for co-funded procurements), making the approach scalable to reach a larger number of organisations and foster adoption.
- Using a small number of providers (more than one), selected via a mini-competition, allows service continuity and avoids vendor lock-in.
- Buyer Groups act as catalysts for adoption since other research organisations can follow the footsteps of becoming a lead buyer for their own research community after witnessing the model executed with practical successful examples.

In order to enable a joint-procurement, careful planning is needed concerning the management of the procurement exercise itself in order to achieve a genuine commitment from the buying organisations. This will in turn provide the market the necessary comfort that their effort in responding to the tender will not be wasted. This is a key element in order to achieve preferential rates and terms for those participating, as suppliers need to really see value in taking part.
To this end, the ARCHIVER project[9] has established an original Joint Procurement Agreement (JPA) that defines the roles of each Buyer Group member and establishes its governance as well as the levels of pre-commitment of funds for procurement.
In addition, the main procurer/lead buyer will need to provide 'added-value' services to support the initiation and execution of a joint procurement, in order to ease the process of planning and implementing the tendering activity.
In the context of the recently funded H2020 cluster projects, specifically concerning the INFRAEOSC-04 call, the projects PANOSC[10], EOSCLife[11] and ESCAPE[12] have envisioned a procurement activity aiming at better investigating how sourcing of commercial services could be improved. As an example in this context, joint procurement is believed to be viable although the detailed concrete setup yet needs to be defined. In all these cases, cooperation with OCRE as a pilot procurement vehicle for EOSC is expected for these projects.

Taking into account all of the above, a **Buyers Group can be defined** by the following six points:

- Each Buyer Group member organisation commits procurement budget.
- Financial commitment is critical even in the presence of a co-funding mechanism.
- Each Buyers Group is represented legally by a Lead Buyer/Main Procurer.
- Each Lead Buyer organisation must be in a position of handling the procurement process, internal accounting, invoicing and payments.
- A Joint Procurement Agreement (JPA) is signed by every member of the Buyers Group.
- The JPA includes the financial commitments and management aspects of the tender process.

[9] http://archiver-project.eu/
[10] https://www.panosc.eu/
[11] https://www.eosc-life.eu/
[12] https://projectescape.eu/

GEANT IaaS Framework

In the GÉANT GN4-2 project[13], cloud, identity management and network requirements were gathered amongst the European National Research and Education Networks (NRENs) and their member education and research institutions. These were applied in a call-for competition with 36 participating countries for public cloud IaaS solutions, leveraging the 2014/24/EU procurement directive. This directive allows multiple organisations to run a tender through a single contracting authority (in this case, GÉANT).

The IaaS tender was conducted in 2016, and resulted in framework agreements with 20 providers, which lasts until the end of 2020: Amazon (AWS) and Microsoft (Azure) through a number of resellers, Cloudsigma, Dimension Data, Interoute, ITsoft, KPN, Lattelecom, Telecom Italia Sparkle, T-Systems and Vancis. All framework agreements are procurement-compliant in the EU and EEA, provide legal compliance and allow for institutional-supplier call-off contracts, in which an institution acts as buyer and (at its discretion) asks a supplier selected in the IaaS tender, to sell its services under the prices, terms, and conditions specified in the framework agreements. These call-offs can take place under local (national) law of the institution.

Furthermore, the services can be acquired through invoice-based billing (no credit card needed), and the suppliers offer aggregated user discounts. These same discounts apply to all eligible 10,000 institutions, regardless of size.

All suppliers are required to support SAML2 to enable single sign-on, some connect to eduGAIN. Concerning network connectivity, some providers have established connectivity through direct private peerings with the GÉANT network in order to ensure a potentially significant reduction in the data ingress/egress charges.

EOSC-hub

CERN and GÉANT are partners in the EOSC-hub project[14] Work Package 12. The EOSC-hub has a 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. WP12 contributes to the design of future business models and procurement frameworks for acquiring digital services from both publicly funded and commercial providers. In this view, the EOSC-hub project published the results of a demand-side market research to understand the need for and level of demand for digital services for research in the context of the EOSC[15]. It explores the manner in which such need and demand are currently met and challenges faced in respect of analysis workflows, data management, and related infrastructure and services. It identifies current and preferred delivery models for such services as well as funding streams and procurement strategies and proposes areas of improvement for business models. The report argues that the formation of central purchasing bodies must be centrally coordinated. While the concept of aggregated procurement as a whole implies that the community would collectively

[13] https://www.geant.org/Projects/GEANT_Project_GN4
[14] https://www.eosc-hub.eu/
[15] http://bit.ly/EOSC-hubD12_1

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 coordination could prevent this from happening and could instead ensure that central purchasing Bodies focus on 'thematic' or research domain demand aggregation, whilst representing a 'single voice to market' in its engagement with the supply market.
The recommendations made by the EOSC-hub project are complementing the experience in accessing commercial cloud resources gathered by initiatives HNSciCloud and ARCHIVER projects.

Research Organisations Requirements Analysis

A market research conducting an analysis of current challenges in accessing digital services by the research community is provided in the EOSC-hub in D12.1[16].
In short, researchers expect access to be free at the point of use and that commercial cloud services support their research data management plans, as a key aspect of the research lifecycle.
Researchers supported by their institutions are mostly interested in accessing storage and compute services. Apart from object storage capacity, popular storage services required by the research community include cloud synchronization and sharing services (e.g Dropbox, Nextcloud, Owncloud, etc.) as an emergent way of sharing data but also accessing and interacting with existing research data repositories[17].
Regarding compute services, they are often orchestrated through software container technology [18]. Container technology has emerged as the *de facto* standard for commercial cloud services, allowing for example, the easy deployment of a research analysis with all the necessary associated components such as the operating system distribution, associated libraries, databases, data-taking conditions, etc.[19].
Modern data science combines container technology with Machine Learning (ML) and Deep Learning (DL) algorithms, becoming themselves ubiquitous in scientific research, replacing at increasing speed the use of traditional statistical techniques. These algorithms are becoming a key enabler for full research reproducibility. Concrete examples come from different scientific fields, also allowing cross-fertilization between them (e. g. High Energy Physics and Earth Observation[20]). These algorithms require very often the availability of hundreds of units of accelerator hardware such as GPUs[21] and FPGAs[22], support for fast interconnectivity across[23]

[16] http://bit.ly/EOSC-hubD12_1
[17] https://doi.org/10.1016/j.future.2017.09.019
[18] https://www.nature.com/articles/d41586-019-03366-x
[19] http://ep-news.web.cern.ch/content/re-discovering-higgs-boson-cloud
[20] https://2019.isc-program.com/presentation/?id=proj139&sess=sess208
[21] https://indi.to/WvTyw
[22] https://indi.to/ZpWJZ
[23] https://indi.to/SZFNy

and seamless integration with ML foundational platforms[24] running on top of the accelerator hardware infrastructure, as part of the service offer.

Moving beyond IaaS to SaaS

Cloud services are traditionally classified by the generally accepted terms IaaS, PaaS and SaaS. However, as cloud services mature and evolve, such terminology doesn't provide enough precision to differentiate between cloud service offers[25].
Examples as the one mentioned above about data science based on Machine Learning demonstrate that advanced cloud services cut across the IaaS/SaaS stacks offering HWaaS accelerator architectures combined with software layers via containerised environments.
Another example is the recent and growing requirement for engineering software tools and services, such as MATLAB, Mathematica, Comsol, SolidWorks, Ansys, AutoCAD, etc. in the cloud. Accessing cloud-based engineering and design tools, would allow researchers and engineers to leverage the highly scalable compute power of cloud services lends itself to engineering simulation, making it possible to run resource intensive simulations that otherwise couldn't be run on-premise due to licensing and/or compute limitations. Beyond the ability to handle larger and more complex problems, simulation in the cloud also means engineers can run several studies simultaneously as an efficient and rapid way of obtaining comparable results. In addition, the paper "*A survey of Cloud-Based Design and Engineering Analysis Software Tools*"[26] argues that "In a broader sense, Cloud-Based Design and Engineering Analysis (CBDEA) can be used to build highly customized and agile information and communication technology (ICT) infrastructures (Infrastructure-as-a-Service) that integrate application software, digital models, sensors, machines, and big data analytics applications for cyber-enabled product development systems".

Cloud Service Validation & Cloud Vouchers

Research organisations require to test cloud services in order to use them effectively. To this end, a technical validation test-suite has been created in OCRE, in order to package and easily deploy tests derived from scientific use cases[27]. The test-suite includes tests from multiple domains providing a scalable and uniform way of deploying validation tests and therefore assessing transparently and efficiently commercial cloud services.

The analysis of the requirements and usage of cloud credits (vouchers) by individual researchers (Long-Tail-of-Science) is detailed on the report "*Long-Tail-of-Science's*

[24] https://www.tensorflow.org/overview
[25] https://www.uni-ulm.de/fileadmin/website_uni_ulm/iui.inst.200/files/staff/K%C3%A4chele/kaechele13beyondiaasandpaas.pdf
[26] https://www.researchgate.net/publication/306392346_A_Survey_of_Cloud-Based_Design_and_Engineering_Analysis_Software_Tools
[27] https://ocre-testsuite.readthedocs.io/en/latest/

Requirements for Commodity Cloud Services in Europe"[28]. Cloud credit/vouchers are considered as well a fundamental requirement for research organisations to be included as part of procured capacity. Specifically, cloud vouchers are considered a very efficient way to encourage the uptake of commercial cloud services, offering a number of advantages:

- Can easily be distributed to researchers represented by a research organisation.
- Can be used to validate cloud resources before procuring them at scale.
- Offer the possibility to rapidly scale cutting edge technology for emerging R&D (immediate examples are Machine Learning deployments using expensive hardware accelerators, research reproducibility test-beds or cloud-based quantum computing emulators).
- Allows research organisations to effectively track consumption across their communities, a critical requirement for lead buyers or central purchasing bodies in a Buyers Group.

Other critical requirements for cloud services validation include network peering connectivity and Federated AAI services.
Network connections across the GÉANT network for example, ensure minimisation of network charges (IP transit) for research and education institutions, which is essential, as traffic charges (e.g. egress charges from service providers) for the provision of data intensive services are a barrier to the adoption of cloud-based services by the research community.
The wide-area network plays a crucial role in connecting users to cloud-based services that are not co-located. This is particularly true when large data volumes need to be accessed on-demand, where network connectivity must be adequate to support the technical requirements. It would be an advantage if network traffic could be exchanged between different cloud service provider locations over the GÉANT network, for questions of performance and cost, as long as cloud service providers can ensure that all data movement between their locations across the GÉANT network is only on the scope of scientific and education activities.

In terms of federated Authentication, Authorization and Identity (AAI) schemes, research organisations require cloud service providers to support inter-federation services, such as eduGAIN and community AAIs based on the AARC Blueprint Architecture (BPA)[29], supporting protocols such as SAML 2.0[30] and OIDC[31].
To make available a testing environment through a central supported AAI proxy service, so cloud service providers can test the interface of their services with research community federated Identity Providers (IdPs), to virtually reduce to zero the customization effort on the contractor's side, would be a key advantage. Such testing environment would be made available through a proxy service hosted by GÉANT. The objective is that commercial cloud service providers are able to connect to Identity Providers in eduGAIN and/or research community AAI services that will be available in the future through the EOSC and make the

[28] https://zenodo.org/record/3564668#.XhSKcGxKjUI
[29] https://aarc-project.eu/architecture/
[30] https://developers.onelogin.com/saml
[31] https://openid.net/connect/

necessary checks such as user attribute matching for example, in order to ensure that a full authentication workflow is successfully completed, achieving uniformly and rapidly a successful integration.

Implementation of Exit Plans & Cloud-Backups

Successful cloud deployment models for research organisations data imply the existence of verified data exit plans. Lack of implementation of exit strategies remains one of the key factors hampering the broad adoption of commercial cloud services by research organisations, raising concerns of vendor lock-in and inability to migrate across services. The SWIPO code of conduct [32] for cloud providers is an important step in the European data economy in reducing these risks. Research organisations require cloud services that allow the implementation of an exit and/or migration strategy between service providers or across on-premise/public cloud-based environments, supporting the use of open source software, vendor independent standards and interfaces to port data and the provision of generic open APIs, in order to be able to integrate the purchased services either in innovative workflows, or existing IT infrastructure. Another approach for the implementation of an exit plan is the support of data escrow services[33]. Escrow allows the implementation of an exit strategy between service providers or across on-premise/public cloud-based environments. Users escrow data, retrieving it directly from the escrow provider. In this context, public research organisations act as data stewards with well-defined data management plans.
A concrete example of the implementation of a successful exit strategy can be found in ARCHIVER deliverable D1.1 "*Initial Data Management Plan*"[34].

NRENs Requirement Analysis

National Research and Education Networks (NRENs) are providing a large range of cloud services to research and education organisations across within their own country. The level of usage is heterogeneous both across countries and users as the level of experience in using cloud varies from "first-timer" to "expert". A few basic common characteristics of dealing with university cloud consumption is the presence of a central IT office that usually has its roots in the scientific computing lab, from a couple decades ago, nowadays grown to be responsible for IT service provision all across the campus. On the other hand some university faculties and departments are often quite autonomous in satisfying their cutting edge research computing needs. Taking this into account, we can group university cloud requirements as follows:

1) "old-school" central IT office:

They tend to focus on enterprise workloads (administration applications, student management, e-learning environments, backup, etc.) and legacy compute services (CPU, storage, VMWare,

[32] https://www.swipo.eu/media/SWIPO_press_release.pdf
[33] https://tools.ietf.org/html/draft-ietf-regext-data-escrow-02
[34] https://zenodo.org/record/3600421#.XhSYqmxKjUI

etc.). Universities, as public sector institutions operate under national directives/conventions covering:

- public procurement
- data protection legislation -
- integration of employee/student identity management -
- integration of new solutions into existing workflows -
- low level of expertise outsourcing i.e. inexperience engaging with professional services

2) “cloud-savy”, autonomous faculties, departments and its members:

They normally embrace a role of cloud-competency centres requiring multi-cloud solutions with cross-cloud-management capabilities. These centres are often interested in higher discounts due to greater consumption and are willing to engage in “smart” cloud configuration options to make savings (pre-commit, spot market, etc.). “Cloud-savvy” users are open/interested to go cloud-first and discontinue on-premise resources, which facilitates the adoption of advanced cloud service usage by institutional users.

The cloud services consumption through GÉANT IaaS available since 2016 has provided some insight in some issues, that can become a set of lessons learned that should be taken into account in the OCRE tender:

- The GEANT Iaas Tender IaaS has eased the procurement administrative burden and provided a chance to tweak service delivery conditions.
- Many cloud-inexperienced institutions had blocking issues earlier in the cloud adoption chain. As a consequence, they could not benefit from the tender advantages during the lifetime of the framework.
- There were perception issues around the concrete advantages of cloud services over on-premise in terms of its relative cost, which resulted in a general lack of motivation to adopt.
- The level of demand for cloud services from institutional users remained largely undefined.
- Concerns with compliance from cloud providers with EU legislation such as general data protection (GDPR) were often hampering adoption.
- In general terms, institutional funding guidelines are very specific and limited to hardware purchases and not adapted to broader cloud services adoption.
- There were general difficulties in many of the potential organisations willing to consume cloud, in comparing CAPEX vs. OPEX due to the way IT resources expenditure is calculated.

The GÉANT IaaS tender is found by the NREN community useful for cloud-aware institutions as it helped these organisations to procure more easily and at lower cost. However, it did little to

help first-time users to overcome initial issues in adopting cloud, as it was not found at the point of offering cloud services considered “turn-key” enough for many of the first-timers.

Next Steps and Implications for the EOSC

The OCRE project foresees the launch of a pan-European tender to deliver commodity cloud services for the research communities in Europe. The requirements for commercial cloud services identified in this report for research organisations and national research and education networks (NRENs) were proposed to the OCRE management team as an input for the preparation of this pan-European tender.

The exploration of innovative business models (e. g. buyers group, cloud vouchers and combinations of both) could lead to the frameworks resulting from the OCRE procurement exercises either being published/promoted or accessed via either the future EOSC Portal Catalogue, or via a dedicated area of the EOSC Portal.

In practical terms, framework agreements designed to leverage economies of scale through aggregation can offer preferential rates and terms to the end researcher and most of all provide her/him access to cloud services that in other models would be very difficult to achieve. The field experience gathered in exploring these mechanisms in OCRE can provide valuable feedback if a future similar functionality can be provided through the EOSC Portal in order to ensure that end users can benefit from the improved rates and contractual terms, as uptake and aggregate spending increases.

Apart exploring the aforementioned procurement and transaction models to maximise commercial cloud services adoption, OCRE can also provide a valuable contribution to the future rules of participation in the EOSC for commercial cloud providers. A set of technical validation rules can be derived from the extensive field deployment of test suites that gather scientific use cases from multiple domains. In addition, field testing of the technical and organisational measures necessary to ensure conformity with the current European legislation (e.g. GDPR and Free Flow of Data) can provide the necessary safeguards to end researchers in using high quality digital services guaranteeing digital sovereignty and ensuring legal certainty in how researchers produce, use and reuse scientific data.