ECMWF Copernicus Procurement

Invitation to Tender



Copernicus Climate Change Service

Volume II

Provision of Cloud Hosting Services for Climate Data Store (CDS)

ITT Ref: C3S_23b

ISSUED BY: ECMWF Administration Department Procurement Section

Date: 10 May 2017

Version: Final



Implemented by



Table of Contents

Table of Contents						
1	Ge	neral	eral Specifications3			
	1.1	Def	initions	3		
	1.2	Def	initions of Cloud Computing Environments	5		
	1.3 Introduction					
	1.3.1		CDS Toolbox Software	5		
	1.3	.2	Assumptions	3		
	1.3	.3	Project timeline and Contract duration	Э		
2	2 Requirements		Э			
2.1 Format of Tenders			Э			
	2.1	.1	Summary:	Э		
2.2		.2	Experience and References	Э		
	2.1	.3	Scope of requirements and questionnaire (Volume III A)	Э		

1 General Specifications

1.1 Definitions

General definitions can be found in Volume I. Definitions specific for this ITT (applicable to Volume II Volume III A and Volume IIIB) are defined below.

Name	Definitions
API	Application Programming Interface
Application	A program for the end user
Broker	A middleware software component which will schedule and forward requests from the web portal to remote repositories
Cloud - On-premises Private Cloud	Cloud computing environment dedicated to ECMWF where the Servers/Resources are physically installed on ECMWF premises.
Cloud - Public Cloud	Cloud computing environment in which ECMWF virtualised services will draw their resources from a shared pool of servers across public networks.
Cloud - Hybrid Cloud	Cloud computing environment which uses a mix of on-premises private cloud and public cloud services with orchestration between platforms.
CDS	Climate Data Store
СМР	Cloud Management Platform
CSP	Cloud Service Provider
Data	The raw data collected
Data repository	A generic name for a system that holds data and/or products. This can be a database, a collection of files, etc.
Data Provider	An organization that make its data and products available through the Climate Data Store
Metadata	Ancillary information about the data (e.g. What was done, when it was done and conditions)
Pre-Operational phase	Phase 2: Testing and building of operational procedures inside the components
Product	A derived, valued added piece of information, usually generated from raw data
Product Catalogue	Holds a list of available products
Proof of Concept phase	Phase 1: Building and testing of the overall architecture
QoS	Quality of Service

SBA	Societal Benefit Areas
SLA	Service Level Agreement
Tool	Means software that performs computations on some input data or products and produces some output data or products
Toolbox	For the purpose of this ITT, the toolbox refers to the database referencing the list of available tools, workflows and applications
URL	Uniform Resource Locator
Users	Users of C3S infrastructure
Web portal	The web page to the CDS portal that enables the users to view information, access / perform tasks on the Product Catalogue, toolbox and CMS
Workflow	A series of invocations of software tools whereby the output of the preceding tool becomes the input for the one which follows it until the required processing chain is completed

1.2 Definitions of Cloud Computing Environments

The definitions in accordance with the understanding and requirements of ECMWF for the types of Cloud computing environments that are referred to in this document are as follows: **On-premises Private Cloud:** Cloud computing environment dedicated to ECMWF where the Servers/Resources are physically installed on ECMWF premises.

Public Cloud: Cloud computing environment in which ECMWF virtualised services will draw their resources from a shared pool of servers across public networks.

Hybrid Cloud: Cloud computing environment which uses a mix of On-premises, Private Cloud and Public Cloud services with orchestration between platforms.

1.3 Introduction

The European Commission (EU) has entrusted ECMWF with the implementation of the Copernicus Climate Change Service (C3S) for the period of 2014 to 2020. Its principal objective is to enhance the knowledge base to support climate change adaptation and mitigation policies.

The Climate Data Store (CDS) constitutes the core building block of the C3S and will be at the heart of the C3S infrastructure. The system is designed to provide a consolidated view and point of access to climate data that is distributed over multiple data suppliers. All C3S data will be catalogued in the CDS, and augmented with quality information that allows the data users to discover, appraise and select appropriate data. The CDS will also provide a set of tools (toolbox), workflows and applications that will allow users to perform processing, computation, transformation and visualisation on top of catalogued data. Users will be presented with options to either download raw data or to build workflows based on concatenation of one or more data sources and parameterised tools, submit these to the system, directly visualise data or even to create ad-hoc applications through available functionalities. The web portal and backend systems manage all interfaces and brokers between internal components and the distributed data providers through so called adaptors. The user experiences a single portal environment with consistent style, interface and controls.

Data requested on the web portal can be delivered raw to the user's home site, or it can be moved into the CDS compute area for processing.

In order to assure the highest quality of service, the Copernicus Climate Change Service aims to benefit from the latest state-of-the-art technologies, hardware and software, setting up a reference framework infrastructure able to accommodate evolving needs and requirements coming from a wide diversity of users and sectors, from experts and scientists to policy makers, while keeping the integrity and robustness of the system.

Aligned with the European Cloud Computing Strategy adopted by the European Commission, the Copernicus strategy foresees an incremental and fast adoption of cloud based infrastructures. In line with these aims, and taking into consideration the huge volumes of data needed to be transferred into the system, it has been decided to host the CDS on a fully managed On-premises Private Cloud with an additional Public Cloud service to allow for growth-the Hybrid Cloud solution.

The purpose of this call for tender is the procurement of a fully managed Cloud Hosting Service, complying with the requirements, motivations and assumptions stated in this document. ECMWF also seek additional/optional support services.

1.3.1 CDS Toolbox Software

The CDS software is built following the basis of a Service Oriented Architecture in which each functional component is engineered as a self-contained entity that exposes a well-defined interface through which its services may be invoked.

This architecture represents a flexible distributed solution which scales horizontally and where each service can be deployed on its own Virtual Machine (VM).



Fig.1: CDS Toolbox Infrastructure

The CDS services are briefly summarized here:

- Web Portal: The CDS-Toolbox user interface is a single consolidated front-end bringing together the Drupal CMS and the data portal user interface and presented as an enterprise web application. A fully integrated data portal user interface is implemented by utilizing the Web APIs provided by each service supplying information for display and functional interfaces for requests.
- **Catalogue**: The Catalogue provides the inventory of all datasets available across all data providers. The Catalogue functions are supported by a dedicated search engine and index (Solr) that provide fast faceted search capabilities.
- **Toolbox:** The Toolbox provides an inventory of all tools available in the CDS Toolbox. For each tool the inventory describes the parameterization of the inputs, the format of the outputs and the means of invoking.
- **Broker/Scheduler:** The Broker/Scheduler is a single component that sits at the heart of the system workflow. All user requests for data retrieval and processing are mediated and managed by the Broker/Scheduler. It is responsible for directing the user requests to the appropriate endpoints, either to data providers via the corresponding adaptor, or execution of tools in the CDS Compute area. The Broker handles the outputs of these operations and drives the workflow by feeding the outputs to the next step in the flow. Finally the Broker returns the URL of the resultant dataset to the user. The Scheduler maintains queues that implement the Quality of Service rules and user priority rules that control the order and simultaneity of requests within the CDS.
- **Compute:** Provides a cluster of nodes for execution of processing tasks using tools from the toolbox. The compute nodes are 'fronted' by a compute service that abstracts the details and management of the individual nodes. Compute functions may also be supported by some data providers, in which case the Broker directs the request via the appropriate adaptor.

- Adaptors: The interface between the Broker/Scheduler and the External Data Providers. They provide an interface translation that allows the CDS to request data retrievals, data sub-setting and processing requests. An adaptor is implemented as a service that conforms to a common interface.
- **Data Services:** provides a so called "Local Data-store" within the CDS for permanent or temporary storage of data, intermediate products and processing results. CDS components push, pull and manage data in the local store. Data retrieved from external providers is pushed into the Local Data-store by adaptors. This data may be input to further processing steps, or can be provided to users via the appropriate download services.
- User Services maintains the User Profile Database that captures the details of each user, including their privileges and priorities regarding access to data/processing resources.
- **CDS Web API.** An Application Programming Interface that provides a language agnostic programmatic interface to the CDS functions, including Catalogue, Toolbox and Broker/Scheduler.
- Monitoring & Metrics: Each component in the CDS is responsible for maintaining relevant status and metrics regarding its operations to be reported on demand. Each component makes this information available via its service Web API. The Monitoring & Metrics service routinely interrogates each component to gather information. In addition to the monitoring of system status, each component logs messages regarding their activities, and C3S staff have the capability to access and interrogate these logs.

1.3.2 Assumptions

The following assumptions apply to some of the requirements included on this tender:

- CDS is an innovative and unique platform currently under development. Operational variables in terms of data volumes, compute workloads, user accesses and future system requirements is at this stage based on estimations. It is expected that this information will be further adjusted based on monitoring and statistics once the system enters on its operational phase.
- The Cloud infrastructure will host multiple CDS environments (Production, Development, Testing, Integration ...)
- The CDS infrastructure will provide a one-stop shop, 24/7/365 operationally available and unique point of entry for a wide community of users to several heterogeneous, large and distributed data sources (from Megabytes to Petabytes). User perception of system performance and availability is cornerstone.
- CDS is a distributed platform providing access to external and internal data repositories.

1.3.3 Project timeline and Contract duration

The supplier shall deliver, deploy, and configure a hybrid Cloud with at least the minimum requirements by the end of Q3 2017 or earlier.

The contract shall have an initial duration to the end of 2018 with optional yearly renewals thereafter until the end of 2021.

2 Requirements

2.1 Format of Tenders

Responses to the requirements section of this ITT should compose of the following:

2.1.1 Summary:

Tenderers should, at the beginning of their proposal, include an executive summary of the response to the ITT showing an understanding of our requirements, and how their proposal would meet our needs. Following requirements should be covered in this respect:

Strategy and Capabilities

Tenderers should be able to deploy a Hybrid Cloud Hosting Model as described in the scope of this tender (Section 1.2) and provide advanced and expert support, solutions and recommendations to ECMWF for the Cloud Hosting requirements.

Tenderers should provide information about the Hybrid Cloud solution they propose, highlighting how it meets our requirements in terms of scalability, reliability, cost, performance and suitability to the specificities of our CDS (data, products, users, etc.) including their recommendations.

Timeline and Delivery

Tenderers should confirm that for the On-Premises Private Cloud, they are able to deploy a fully managed platform at ECMWF as soon as possible and, in any case, by end of Q3 2017 at the latest.

2.1.2 Experience and References

The tenderers should have a proven track record providing enterprises with cloud-based infrastructure services. Tenderers shall provide details of at least 2 successful projects similar to our requirements highlighting their active involvement/contribution to them. Names and contact details for the organisations involved should be included. ECMWF reserve the right to contact these individuals for clarification and feedback purposes at any point during the ITT process.

2.1.3 Scope of requirements and questionnaire (Volume III A)

Tenderers shall respond to each functional and technical requirement included in Volume III A Scope of Requirements and Questionnaire with a clear YES/NO answer for compliance, together with additional information giving background reasoning for the response.

Tenderers may also include additional documentation, or links to specific pages/documents, but these should be limited and only included where they enhance their proposal. These documents must not be in place of full and clear responses to the requirements described above, and if unclear as to their purpose, irrelevant or unduly lengthy, may be ignored by ECMWF.