RES Data Management Services
Call for Proposals 2024

Opening date:	**16 November 2023**(*)
Closing date:	**23 January 2024, 11:00 CET**
Communication of allocation decision:	End of March 2024
Start of allocation:	**01/04/2024**
Type of Access:(†)	3 to 5 Year Project Access

Data Management Resources available

RES Data Node	File/object system	2024			Available CPU hours	Available VMs
		Storage	Objects	Tapes (backup)		
BSC	GPFS/swift, HSM	6045	-	-	Yes	Yes
CESGA	Lustre	0	-	200		
COMPUTAEX	GPFS/Lustre	165	-	-		
CSUC	NFS, SMB/CIF, S3	100	-	60		
IAC	VFS Lustre	140	-	-		
PIC	dCache	500	-	1000		
SCAYLE	S3	175	-	-		
UMA-SCBI	S3	0	1800	-	Yes	
UZ-BIFI	ceph	250	-	-	Yes	Yes

Introduction

RES calls for proposals for Data Management Services open each year in November, once the terms of the call and the resources available have been approved by RES Council, and remain open for a minimum of 8 weeks.

Scientific groups in need of resources for the exploitation of their scientific or research data (storage, virtual machines, residual computing capacity) are invited to apply to this call. Other data services, e.g.: long-term storage are out of the scope of this call.

Projects Details and Execution

The Principal Investigator is the investigator responsible for the project, and must indicate his/her affiliation in the intranet system.

Exploitation: Each project may request data and computing services to exploit the stored data. In the DMP, it will be possible to indicate the node that is considered the most suitable, according to the services offered. This selection must be duly justified and will be considered at the time of assignment. Projects without an exploitation plan will not be accepted.

Monitoring: The projects will have to undergo a periodic technical evaluation process (monitoring) by the access committee, which will decide on the allocation of resources for the project. The nodes will provide a resource usage report for each project for technical review on a semi-annual basis. The principal investigator will have to submit an update of the project DMP on an annual basis. The scientific evaluation of the project will be every two years.

Completion: The Principal Investigator of the project will be responsible for specifying the destination of the data at the end of the project (transfer and/or destruction modalities). As a general rule, projects will have a maximum period of one month from the evaluation results' communication to leave the node. The project's principal investigator may negotiate the conditions of completion of the project with the node, according to the data management policy of the host node.

Eligibility criteria

Projects that meet the minimum criteria and for which the access committee has positively assessed the management viability, availability of resources, and scientific relevance of the project will be admitted.

Storage: This call is open to projects requiring between 200 and 1000 TB. The access committee may consider accepting projects outside this range when this is justified by the need to have services exclusively offered by RES or because it is of special scientific interest.

Duration: The duration of a data projects can be from 3 to 5 years. The duration of the projects will have to be specified at the time of application. Projects will be subject to annual evaluations. Those projects not passing successfully two consecutive evaluations may have their duration shortened.

Application

Applications to this call are submitted through RES intranet forms (https://res-intranet.bsc.es/), following the submission forms therein. Other formats or references to external information will not be accepted. All information must be submitted in English.

Applications consist of a **Project Scientific Relevance** and a **Project Technical Description** (Data Management Plan, DMP). The Technical Description (DMP), to be updated annually, includes:

- Responsible of the data project information (cv, team, publications, team)

- Scientific goal for and motivation for storage in a large data infrastructure

- Technical management plan Datasets description: volume and type of data

    - Datasets description: volume and type of data

    - Use of standard, publicly accessible formats

    - Complete and accessible description (supporting documentation for the project, metadata, data model) so that the data is understandable to other groups

    - Permissions on access to data (public, private, embargoed, or restricted, with the definition of the corresponding license)

    - Justification of the computing resources to deploy platforms and services necessary to access the data (some nodes can offer data services)

    - Project duration and estimated useful life of data, beyond the project (includes project completion plan)

    - Minimum planning of 3 years

Please note that if you wish to continue work after the end of a project, a new proposal (i.e. a continuation proposal) needs to be submitted in addition to a final report of the previous activity. This is also necessary when the scope or requirements of a running project are significantly modified during its execution.

Definition of the pre-reservation system: a project can be submitted to the call with a DMP before having data, conditional on obtaining another data generation project (HPC or another). The reservation will be reviewed in advance and reassessed according to the calendar of the call for data projects, and the expected period of obtaining the data.

RES data nodes

The RES Data Nodes available for the call are:

RES Data Node	File/object system	Location
BSC-CNS	GPFS/Swift, HSM	Barcelona
BIFI-UZ	CEPH	Zaragoza
COMPUTAEX	GPFS/Lustre	Cáceres
CESGA	Lustre	Santiago de Compostela
CSUC	NFS, SMB/CIF, S3	Barcelona
IAC	VFS Lustre	La Palma
PIC	dCache	Barcelona
SCAYLE	S3	León
SCBI-UMA	S3	Málaga

BSC-CNS – Barcelona Supercomputing Center- Centro Nacional de Supercomputación, data and computing node, hosted in Barcelona

- Total storage capacity of the center: 120 PB
- Available support: HSM (disk and tape)
- Data access protocol: Nfs, gpfs, Swift
- Connection and data transfer: ssh, scp, sftp, rsync, bbcp, ftps, gridftp, sshftp.
- Data services: Backup service available. Virtual Machines to access the data and deploy services are offered by the center personnel. EUDAT data services can be provided: B2DROP, B2SAFE and PIDs.
- Computing services: Cloud cluster (OpenStack) and HPC cluster Nord3. RES computing system MareNostrum4 accessible via RES computing calls.

BIFI-UZ – Institute of Biocomputing and Physics of Complex Systems of the University of Zaragoza, data and computing node, hosted in Zaragoza

- Total storage capacity of the center: 0.9 PB
- Available support: disk
- Data access protocol: NFS/SMB (30%), S3 (70%)
- Connection and data transfer: CEPH
- Data services: Backup service available. Virtual Machines to access the data and deploy services are offered by the center personnel. Support in data mining, big data and machine learning tools.
- Computing services: Cloud cluster and RES HPC cluster CAESARAUGUSTA.

COMPUTAEX - Centro Extremeño de iNvestigación, Innovación Tecnológica y Supercomputación, Fundación COMPUTAEX, data and computing node, hosted in Cáceres

- Total storage capacity of the center: 0.86 PB
- Available support: disk
- Data access protocol: GPFS, Lustre
- Connection and data transfer: ssh, rsync
- Data services: Technical support for data management of data of disciplines like Chemistry, Physics, Health, etc., connected to public databases (Eurostat, AEMET, …)
- Computing services: RES computing cluster Lusitania

CESGA - Fundación Pública Gallega Centro Tecnológico de Supercomputación de Galicia, data and computing node, hosted in Santiago de Compostela

- Total storage capacity of the center: 3.35 PB
- Available support: disk and tape
- Data access protocol: Lustre
- Connection and data transfer: gridFTP, Globus, Aspera, scp, sftp, rsync.
- Data services: Galaxy Portal for Genomic data analysis available. Weather data service Thredds. Support for clustered databases like Galera.
- Computing services: RES and PRACE computing cluster FinisTerrae-II. Cloud computing and service (EOSC). Hadoop platform supporting Spark, Jupyter notebooks, Hive, Impala.

CSUC - Consorci de Serveis Universitari de Catalunya, data and computing node, hosted in Barcelona

- Total storage capacity of the center: 1.6 PB
- Available support: disk
- Data access protocol: NFS, SMB/CIFSm FTP, S3
- Connection and data transfer: S3 and FTP
- Data services: Virtual Machines to access the data and deploy services are offered by the center personnel. A GridFTP service can be provided under request. Optional version control service for the S3 protocol. Storage snapshots available under request.
- Computing services: RES computing clusters PirineusII and Canigó

IAC - Instituto de Astrofísica de Canarias, data and computing node, hosted in La Palma

- Total storage capacity of the center: 1.34 PB
- Available support: disk
- Data access protocol: Lustre VFS
- Connection and data transfer: ssh, GridFTP
- Data services: GitLab is available for software version control, exclusive for the data node projects.
- Computing services: RES computing cluster LaPalma

PIC – Port d'Informació Científica, data node, hosted in Barcelona

- Total storage capacity of the center: 55 PB
- Available support: disk and tape
- Data access protocol: webdav/https, gridftp, xrootd, NFS

- Connection and data transfer: GridFTP, FTS, Rucio

- Data services: The storage service permit the definition of the number of replicas per file, as a data backup service.
- Computing services: Virtual Machines to access the data and deploy services are offered by the center personnel. Batch computing for non-interactive jobs; JupyterLab notebooks service for interactive data analysis.

SCAYLE – Centro de Supercomputación de Castilla y León, data and computing node, hosted in León
- Total storage capacity of the center: 1.5 PB
- Available support: disk
- Data access protocol: S3
- Connection and data transfer: S3 compatible tools: Rclone, S3sync
- Data services: Backup system available for all data, not as a second copy for the user
- Computing services: RES computing cluster Caléndula

SCBI-UMA – Servicio de Supercomputación y Bioinformática of the University of Málaga, data and computing node, hosted in Málaga.
- Total storage capacity of the center: 5 PB
- Available support: disk
- Data access protocol: S3
- Connection and data transfer: ssh, sftp
- Data services: The system redundancy needs no backup
- Computing services: RES computing cluster Picasso