

# Barcelona Supercomputing Center – Centro Nacional de Supercomputación

## Resolución Comité de Acceso

### Asignación de recursos de datos para la Red Española de Supercomputación (RES)

## Convocatoria de datos 2021

Barcelona, abril de 2021

## 1. Introducción

El Comité de Acceso del Barcelona Supercomputing Center – Centro Nacional de Supercomputación es un órgano asesor del Director que informará las solicitudes de acceso al Centro de los investigadores y grupos de investigación que lo soliciten. El Comité propondrá al Director, en base a la calidad científica y técnica de las propuestas recibidas una lista razonada y priorizada de las solicitudes. Corresponde al Director la decisión sobre los accesos autorizados.

El acceso es para los recursos de la Red Española de Supercomputación (RES). La asignación entre los diferentes sistemas se hace con motivos de necesidad de las actividades y de eficiencia. El protocolo de acceso aprobado por la Comisión Ejecutiva del BSC está publicado en la página de web del BSC: <http://www.bsc.es/RES>

## 2. Análisis

En esta primera convocatoria, la RES ha asignado recursos para servicios de datos para el periodo 2021-2023, concretamente 5,78 PB para 2021, 7,63 PB para 2022 y 9,09 PB para 2023. Adicionalmente se han asignado 1,1 PB a proyectos con duración hasta 2025, así como máquinas virtuales y 390.000 horas de CPU para contribuir a la explotación de los datos albergados. Estas capacidades de almacenaje se han distribuido entre las diferentes infraestructuras de datos disponibles en los nodos de datos de la RES: Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC), Instituto Astrofísico de Canarias (IAC), Universidad de Málaga (UMA), Universidad de Zaragoza (UZ), Consorci de Serveis Universitaris de Catalunya (CSUC), CénitS-COMPUTAEX (CENITS), Fundación del Centro de Supercomputación de Castilla y León (SCAYLE), Fundación Pública Galega Centro Tecnológico de Supercomputación de Galicia (CESGA) y Port d'Informació Científica (PIC).

Todas las actividades han sido evaluadas por los paneles de expertos, clasificando las solicitudes según si eran excelentes, muy buenas y buenas. Adicionalmente, se han tomado en consideración los criterios de evaluación descritos en la Sección [Comentarios sobre la evaluación](#).

En esta primera convocatoria se han podido asignar la totalidad de propuestas elegibles recibidas, es decir, todas las propuestas que han superado la verificación técnico-administrativa y que han superado el nivel de calidad y excelencia mínimo indicado por el comité de acceso. La asignación de capacidad de almacenaje se hace para la duración completa del proyecto, con asignación completa desde el principio o incrementos periódicos, pero con revisiones anuales sobre el uso de la capacidad y del Data Management Plan (DMP) asociado, que pueden conllevar a revisar o incluso anular la asignación.

### 3. Comentarios sobre la evaluación

El Comité de Acceso del BSC-CNS ha seguido los siguientes criterios para la evaluación de las actividades:

- 1) Reglas generales
  - a. La relevancia del proyecto científico en que se enmarca la actividad propuesta (20 %)
  - b. La justificación de la actividad propuesta a realizar en la RES para la consecución del proyecto científico global (30 %)
  - c. La calidad científica del grupo solicitante (10%)
  - d. La experiencia y capacitación en servicios de gestión de datos (10 %)
  - e. La necesidad real de los recursos de la RES para realizar la actividad (20 %)
  - f. La adecuación técnica del proyecto a la arquitectura de los recursos de la RES (10 %)
- 2) Participación de grupos españoles en las actividades solicitadas
- 3) Actividad específica dentro de un proyecto de investigación. El acceso a los recursos de la RES corresponde a actividades específicas dentro de un proyecto de investigación, y no corresponden a agrupaciones de diferentes actividades de investigadores de comunidades virtuales.

Siendo esta la primera convocatoria de este tipo, no se han considerado antecedentes de uso, resultados y buenas prácticas, tal como es preceptivo en la evaluación de actividades de cómputo.

### 4. Consideraciones adicionales

#### 4.1. Actividades industriales

Cualquier actividad industrial está sujeta a las mismas condiciones de calidad que las actividades de investigación pública. Todos los usuarios con actividades industriales, y con acceso a las máquinas de la RES deben pagar por el acceso a los recursos siempre que las actividades no sean incluyan en el concepto de Open R&D. El precio se calcula para cada una de las actividades que lo indiquen, teniendo en cuenta los recursos solicitados (humanos y técnicos) y el interés científico/económico de la actividad.

## 4.2. Política de uso de tiempo de CPU y otros servicios adicionales

En la actual resolución, se ha realizado asignación no sólo de espacio de almacenamiento, sino también tiempo de CPU y otros servicios adicionales. Se ha tenido en cuenta el tiempo de CPU solicitado, así como el disponible y la eficiencia en la utilización de los recursos. Para aquellas actividades que lo han solicitado, se ha especificado la asignación de tiempo de CPU y de otros servicios adicionales, notablemente máquinas virtuales para la explotación de los datos almacenados.

## 5. Listados y asignaciones

A continuación, se incluye la lista de las actividades que tendrán acceso a los diferentes nodos de la RES.

Lider	Título	Centro	Almacenamiento	Servicios adicionales
Adam Hospital	BioExcel-CV19 MD database and server	BSC	- 2021: 500 TB en HSM (copia simple en cinta) - 2022: 500 TB en HSM (copia simple en cinta) - 2023: 500 TB en HSM (copia simple en cinta)	
Adrià López Baucells	Implementation of Passive Acoustic Monitoring using machine learning algorithms for species ID in the Bat Monitoring Program: a new acoustic repository for wildlife research.	CSUC	- 2021: 90 TB - 2022: 125 TB - 2023: 200 TB	Hasta dos máquinas virtuales
Alfonso Valencia	Federated EGA	BSC, IAC	- 2021: 300 TB en HSM de BSC (copia simple en cinta) y 300 TB en IAC (backup) - 2022: 300 TB en HSM de BSC (copia simple en cinta) y 300 TB en IAC (backup) - 2023: 300 TB en HSM de BSC (copia simple en cinta) y 300 TB en IAC (backup)	Hasta dos máquinas virtuales
Alfonso Valencia	Characterization of liquid phase nuclear compartments using Hi-C and multiscale modelling	BSC	- 2021: 24 TB en HSM (copia simple en cinta) - 2022: 160 TB en HSM (copia simple en cinta) - 2023: 307 TB en HSM (copia simple en cinta)	
Andres Pacheco Pages	Experimental data storage of the ATLAS experiment from publications of the IFAE group	PIC	- 2021: 200 TB en disco y 200 TB en cinta - 2022: 200 TB en disco y 200 TB en cinta - 2023: 200 TB en disco y 200 TB en cinta	100k horas de CPU por año y servicios WLCG

Carles Bo	ioChem-BD	BSC	- 2021: 22 TB en HSM (copia simple en cinta) - 2022: 66 TB en HSM (copia simple en cinta) - 2023: 220 TB en HSM (copia simple en cinta)	Hasta dos máquinas virtuales
Christian Neissner	PAU survey: High-precision photometric redshift from narrow band observations in the visual range	PIC	- 2021: 75 TB en disco y 25 TB en cinta - 2022: 100 TB en disco y 100 TB en cinta - 2023: 150 TB en disco y 150 TB en cinta	50k horas de CPU por año
David Posada	Uncovering cancer evolution through high-throughput DNA sequencing	CESGA	- 2021: 150 TB - 2022: 225 TB - 2023: 300 TB	20k horas de CPU por año
Francisco Prada	Dissemination of the Uchuu cosmological simulation	CESGA	- 2021: 109 TB - 2022: 210 TB - 2023: 300 TB	Jupyter Lab Notebook en cluster Hadoop cluster, 2 VMs
Gustavo Yepes Alonso	The database of the 300th Galaxy clusters simulation project	BSC	- 2021: 260 TB en HSM (copia simple en cinta) - 2022: 300 TB en HSM (copia simple en cinta) - 2023: 300 TB en HSM (copia simple en cinta)	200k horas de CPU por año, en Nord
Javier Castañeda Pons	Gaia: Telemetry and Main Database backup and Catalogue releases	BSC	- 2021: 800 TB en HSM (copia simple en cinta) - 2022: 1200 TB en HSM (copia simple en cinta) - 2023: 1600 TB en HSM (copia simple en cinta)	Hasta dos máquinas virtuales
Javier Rico	Deploying the MAGIC Data Legacy	PIC	- 2021: 200 TB en disco y 200 TB en cinta - 2022: 200 TB en disco y 200 TB en cinta - 2023: 200 TB en disco y 200 TB en cinta	
José Flix Molina	A Spanish data cache service for the CMS experiment	PIC	- 2021: 100 TB en disco - 2022: 200 TB en disco - 2023: 200 TB en disco	Xcache
Jose MC Tubio	Exploring the genomics of cancer	CESGA	- 2021: 900 TB - 2022: 900 TB - 2023: 900 TB	20k horas de CPU por año
Juan Antonio Añel Cabanelas	HERCULES: High vertical Resolution Climate Simulations Dataset	CESGA	- 2021: 100 TB - 2022: 200 TB - 2023: 400 TB	

Kim Serradell	Dissemination of Spanish CMIP6 outputs through an ESGF Data node	BSC	- 2021: 1000 TB en HSM (copia simple en cinta) - 2022: 1500 TB en HSM (copia simple en cinta) - 2023: 1500 TB en HSM (copia simple en cinta)	Hasta tres máquinas virtuales
Pablo Loza-Álvarez	SLN intelligent datasets management	PIC	- 2021: 64 TB en disco y 64 TB en cinta - 2022: 122 TB en disco y 122 TB en cinta - 2023: 180 TB en disco y 180 TB en cinta - fin de proyecto: 354 TB en disco y 354 TB en cinta	
Toni Gabaldón	Integrative Omics Of Fungal Pathogens	BSC, SCAYLE	- 2021: 50 TB en HSM de BSC (copia simple en cinta) y 50 TB en SCAYLE (backup) - 2022: 100 TB en HSM de BSC (copia simple en cinta) y 100 TB en SCAYLE (backup) - 2023: 150 TB en HSM de BSC (copia simple en cinta) y 150 TB en SCAYLE (backup) - fin de proyecto: 250 TB en HSM de BSC (copia simple en cinta) y 250 TB en SCAYLE (backup)	Hasta dos máquinas virtuales

## 6. Sigüientes pasos

El equipo de soporte de cada institución contactará con el investigador principal para proceder a dar de alto a los usuarios y los servicios asociados. En breve, se procederá a la automatización de estos procesos, con servicios similares a las convocatorias de computo.