



Engineering and Physical Sciences
Research Council

Pilot call: Access to the Research Data Facility (RDF) for UK researchers

Call type: Invitation for proposals

Closing date for applications to EPSRC: 12 February 2016, 16:00

Closing date for the RDF technical assessment that needs to be attached to your application: 29 January 2016, 16:00

Related themes and Research Councils: Research infrastructure, BBSRC, ESRC, MRC, STFC, AHRC, DiRAC.

Summary

EPSRC recognises the need for researchers to be able to store active computational data and to be able to use this data for further scientific benefit. This is a call for applications to access storage on the national Research Data Facility (RDF). EPSRC has made the strategic decision to open up the RDF to researchers outside of the remits of NERC and EPSRC to ensure maximum usage and impact from the facility. This call is for applicants within the remit of RCUK but performing research **outside of the remit of EPSRC and NERC**. The total amount of storage available for this pilot call is 1000TB. Applicants will need to submit a Technical Assessment along with an application form and this will be assessed by a peer review Resource Allocation Panel.

Background

The national RDF is a data storage facility funded by EPSRC and NERC, and is housed at the University of Edinburgh's Advanced Compute Facility. It is for Researchers in the UK involved in computational research. It is intended to be used for storage of active research data associated with computational resource. The RDF also contains a data analytic cluster which enables users to run analyses on data hosted on the service.

Please see further details here: <http://www.archer.ac.uk/documentation/rdf-guide/cluster.php>

The aim of this pilot call is to provide access to our national Research Data Facility for researchers within the remit of EPSRC.

EPSRC envisage that the RDF will be used to:

- Enable further analysis of computational research and experimental data.
- Allow storage so that further publication of data can be achieved.
- Allow storage of data that is valuable to a project community e.g. a reference data set.

The RDF is **not** for:

- Storage of sensitive and/or confidential data sets, e.g. those involving human subjects, medical data or personally identifiable information.
- Data with copyright and intellectual property issues associated.
- A permanent store for data or an archive.

Resource available

Applications to this call can request RDF resource for a maximum duration of 3 years. There is a limit to the total amount of TB available against this call. EPSRC aims to allocate around **1000 TB** at this Resource Allocation Panel. Users can request storage in a combination of one or more of the following units: **1, 5, 10, 20, 30, 40 and 50 TB**.

Any resource allocated is for immediate use and the start date of the data storage must be within three months of the panel date (see Call Schedule). Please ensure that the data is in a format that can be readily transferred to the RDF.

At the end of the allocation period users should ensure they transfer their data off the RDF, as the data cannot be held after this time and users will be at risk of their data being overwritten.

It is therefore imperative that applicants are well prepared and only request an allocation they can realistically use in the allocated period.

Applicants can also apply for time on the Data Analytics Cluster. For further details of the software available, see the Technical Details section.

Eligibility

Only individuals eligible to hold a full RCUK grant can apply for access to the RDF. Students wishing to access the RDF through this route must apply through an eligible Principle Investigator.

For information on the eligibility of organisations and individuals to receive RCUK funding, see the RCUK Funding Guide:

<http://www.rcuk.ac.uk/funding/eligibilityforrcs/>

Some communities already have the opportunity to access the RDF and so the following individuals should apply for access through these routes:

1. NERC researchers - Please apply through one of the NERC consortia
<http://www.nerc.ac.uk/research/sites/facilities/hpc/>
2. EPSRC ARCHER RAP and Leadership Applicants - Please apply for RDF resource on your application to these calls. Please see the EPSRC website for details on future calls:
<https://www.epsrc.ac.uk/research/facilities/hpc/access/routes/>
3. EPSRC Applicants - Please apply through the following call:
<https://www.epsrc.ac.uk/funding/calls/rdfaccessopen/>

4. Members of EPSRC HEC Consortia- Please contact the respective consortium directly:
- Materials Chemistry Consortium (<http://www.ucl.ac.uk/klmc/mcc/>)
 - UK Turbulence Consortium (<http://www.turbulence.ac.uk/>)
 - Biomolecular Simulation Consortium (<http://www.hecbiosim.ac.uk/>)
 - Plasma Physics Consortium (<http://www.ccpp.ac.uk/hec/index.html>)
 - Mesoscale Engineering Consortium (<http://gow.epsrc.ac.uk/NGBOViewGrant.aspx?GrantRef=EP/L00030X/1>)
 - UK Car-Parrinello Consortium (<http://www.ukcp.ac.uk/>)
 - UK Turbulent Reacting Flows Consortium (<http://www.ukctrf.co.uk/>)

Technical Details of the RDF

The RDF consists of 23PB disk, with an additional 48PB of backup tape capacity. As well as the storage, the RDF is equipped with Data Transfer Nodes (DTNs) and a Data Analytic Cluster (DAC):

Data Transfer Nodes (DTNs)

Each DTN is equipped with:

- One Intel Westmere 8-core processor, 11 GB memory.

Data Analytic Cluster (DAC)

- 1 login node: Two Intel Ivy Bridge 10-core processors, 128 GB memory
- 12 standard compute nodes each with two Intel Ivy Bridge 10-core processors, 128 GB memory
- 2 high-memory compute nodes each with four Intel Westmere 8-core processors, 2 TB memory
- HyperThreads are enabled on all nodes meaning that standard compute nodes each have 40 CPUs available and the high-memory compute nodes each have 64 CPUs available.
- All DAC nodes have high-bandwidth; direct Infiniband connections to the UK-RDF disks.

For further technical details on the RDF see:

<http://www.archer.ac.uk/documentation/rdf-guide/>

How to apply

Applicants must provide the following two documents:

- **Technical Assessment-** This is important to ensure that the level of resources requested have been appropriately scoped and that all technical requirements have been considered prior to submission.

- **Application Form-** When completing your form you should take into account the assessment criteria given below (see Assessment Criteria)

Applications submitted on an incorrect form will not be considered.

Instructions for applicants on how to apply:

1. Obtain an RDF Technical Assessment form from the RDF service provider here: <http://www.rdf.ac.uk/access/ta/>
2. Applicants should download the application form from the RDF service provider: <http://www.rdf.ac.uk/access/direct/RDF-Direct-Access-Other-AppForm.docx>
3. Submit the completed technical assessment (TA) form, which will be assessed, along with a draft of your case for support to: support@rdf.ac.uk prior to **29 January 2016 (16:00)**.
4. Please make sure the subject header of your submission email states that this is an RDF Storage "submission".
5. The TA will be assessed by the RDF support team and the completed document will be returned to the applicant.
6. If the assessor of the TA has concerns about the project, it may take extra time for these to be addressed and the assessor will contact the applicant in this instance. This may take up to ten days from receipt of the completed form. EPSRC and the service provider cannot be held responsible for applications that miss the final deadline if the applicant has not met the deadline specified above for submission of the technical assessment.
7. Provided the technical assessment endorses the proposal, completed applications should be sent together with the completed technical assessment directly via the form on the EPSRC call page (<https://www.epsrc.ac.uk/funding/calls/rdfaccess/>) before the call deadline of **12 February 2016**.

Guidance on writing an application

1. Case for Support: Brief rationale and objectives for using the RDF

In the case for support section please write a brief description on each of the bullet points outlined below (max. 300 words total):

- Briefly describe the research project or science associated with the data you plan to store on the RDF.
- Why is the RDF the most appropriate resource for this work, rather than other local resources e.g. University resources?
- Provide details of any plans you have to utilise the Data Analytics Cluster.
- What will the novel outputs and impact of your data be? If you are using the RDF to process the data further, e.g. for analysis, what benefits will this achieve?

2. Data Management Plan

Please write a brief description on each of the bullet points outlined below (max. 300 words total):

- Please describe how your data will be managed and curated. For example: How will you ensure all project members adhere to the data management plan? Who has ultimate responsibility for ensuring the plan is implemented? Explain how you will use best practice in data management to ensure there is documentation and meta data associated with your research data.
- Please provide URL links to any existing data policies which you will follow. Indicate where appropriate how you will meet the EPSRC policy on data sharing and access. Justify any restrictions on sharing.
- Please describe what will happen to your data after the resource allocation has ended and show you have an awareness of longevity of data. Will the project associated with the data have ended? Will the data be transferred to another local resource? Will you no longer need the data?

Information on the EPSRC data management policy can be found here:

<https://www.epsrc.ac.uk/about/standards/researchdata/>

See Annex A in the linked document for the RCUK policy on data management.

<http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/RCUKCommonPrinciplesonDataPolicy.pdf>

Assessment

Assessment process

Applications to this call are not subject to postal peer review and will be reviewed and prioritised directly by a generalist peer review panel to allocate resource the Resource Allocation Panel (RAP).

Assessment criteria

The following assessment criteria will be used by the panel to rank the applications:

- **Impact of data:** Will the data have a novel impact? E.g. will it produce further publications? Is the RDF the best or only UK system on which this data can be stored?
- **Technical suitability for the RDF** (as defined in the Technical Assessment): What is the technical suitability of the proposed data for the RDF facility? Are the type, format and volume of data to be stored appropriate for the RDF? Is the level of requested resource (no. of TBs) appropriate and well justified? Is there a clear plan as to how the researcher/s will move data on/off the RDF? Do the applicants have the technical expertise to carry out the work?
- **Management of data:** Is there evidence that the data will be managed effectively? Does the applicant follow any existing data policies? Is there a clear plan for documentation and meta data to accompany the data? Where appropriate do the applicants meet the EPSRC policy on data

sharing and access? Is there a clear exit plan in place for the end of the allocation period?

Key dates

Activity	Date
Technical assessment deadline	29 January 2016
Closing date for proposal submissions	12 February 2016
Resource Allocation Panel	March 2016
Earliest storage transfer start date	March-April
Latest storage transfer start date	April-May

Contacts

For further information about this call, please contact:
Katherine.freeman@epsrc.ac.uk

Change log

Name	Date	Version	Change
Katherine Freeman	27.11.2015		