Platforms EOI: Platform for shared research reference data sets and scientific software tools.

Project title

Platform for shared research reference data sets and scientific software tools.

Field of Research code(s)

- 02 PHYSICAL SCIENCES
- 06 BIOLOGICAL SCIENCES
- 08 INFORMATION AND COMPUTING SCIENCES

EOI Lead Name

Simon Gladman

EOI lead Research Group

Melbourne BioInformatics

EOI lead Organisation

The University of Melbourne

EOI lead Email

Collaborator details

<table>
<thead>
<tr>
<th>Name</th>
<th>Research Group</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sean Crosby</td>
<td>Research Platform Services</td>
<td>The University of Melbourne</td>
</tr>
</tbody>
</table>

Project description

This project will deploy a platform to provide services consisting of fast local caches to significant international research reference data sets and scientific software repositories as well as a service to distribute authoritative local data sets and software tools. The project will provide a common platform extendable to all research computing communities in Australia by building upon the existing efforts from both the national physics and biosciences research computing communities. The project will also focus on engaging with a wide range of research domains to promote and assist in the uptake of the platform and also work with computing infrastructure providers to support the technology stack.

The project is unique in that it focuses not only on effectively bringing international resources to local repositories but also on distributing Australian capabilities to a global audience.

It will be accessible from all computational infrastructure within Australia including HPC/HTC systems, research or local cloud facilities, as well as commercial cloud solutions. It will be compatible with containerised environments, virtual machines and physical servers.

Existing technology

Adopt

The technology that the proposed platform will adopt is the CernVM File System (CernVM-FS), a read-only file
system designed to deliver scientific software onto virtual machine and physical worker nodes in a fast, reliable and scalable way. Specifically, we will:
1. Deploy CernVM-FS Stratum 1 Replica Services
2. Deploy an Australian CernVM-FS Repository Gateway and Publisher Service
3. Deploy an Australian CernVM-FS Server Meta Information

### Anticipated requirements

<table>
<thead>
<tr>
<th>Annual funding</th>
<th>$0 - $99,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed length</td>
<td>3 years</td>
</tr>
</tbody>
</table>

### Other information

Other information you wish to provide

It will provide a reference model for other research domains that need to maintain versioned software stacks or versioned data sets for reproducible science methodologies. It will promote and encourage the uptake of the project's unified and centrally supported platform to reduce the complexity and effort of maintaining scientific software stacks within individual domains.

It will reduce the software maintenance overhead of system administrators in Australia by standardising scientific software distribution approaches and by not having to maintain institutional copies of large reference data.

### Terms

I agree to the terms

Yes