Geophysical Research Abstracts
Vol. 18, EGU2016-11164, 2016
EGU General Assembly 2016

Virtual Labs (Science Gateways) as platforms for Free and Open Source Science

David Lescinsky (1), Nicholas Car (1), Ryan Fraser (2), Carsten Friedrich (3), Carina Kemp (1), and Geoffrey Squire (3)

(1) Geoscience Australia, Canberra, ACT, Australia, (2) CSIRO, Kensington, WA, Australia, (3) CSIRO - Data61, Acton, ACT, Australia

The Free and Open Source Software (FOSS) movement promotes community engagement in software development, as well as provides access to a range of sophisticated technologies that would be prohibitively expensive if obtained commercially. However, as geoinformatics and eResearch tools and services become more dispersed, it becomes more complicated to identify and interface between the many required components. Virtual Laboratories (VLs, also known as Science Gateways) simplify the management and coordination of these components by providing a platform linking many, if not all, of the steps in particular scientific processes. These enable scientists to focus on their science, rather than the underlying supporting technologies.

We describe a modular, open source, VL infrastructure that can be reconfigured to create VLs for a wide range of disciplines. Development of this infrastructure has been led by CSIRO in collaboration with Geoscience Australia and the National Computational Infrastructure (NCI) with support from the National eResearch Collaboration Tools and Resources (NeCTAR) and the Australian National Data Service (ANDS). Initially, the infrastructure was developed to support the Virtual Geophysical Laboratory (VGL), and has subsequently been repurposed to create the Virtual Hazards Impact and Risk Laboratory (VHIRL) and the reconfigured Australian National Virtual Geophysics Laboratory (ANVGL). During each step of development, new capabilities and services have been added and/or enhanced. We plan on continuing to follow this model using a shared, community code base.

The VL platform facilitates transparent and reproducible science by providing access to both the data and methodologies used during scientific investigations. This is further enhanced by the ability to set up and run investigations using computational resources accessed through the VL. Data is accessed using registries pointing to catalogues within public data repositories (notably including the NCI National Environmental Research Data Interoperability Platform), or by uploading data directly from user supplied addresses or files. Similarly, scientific software is accessed through registries pointing to software repositories (e.g., GitHub). Runs are configured by using or modifying default templates designed by subject matter experts. After the appropriate computational resources are identified by the user, Virtual Machines (VMs) are spun up and jobs are submitted to service providers (currently the NeCTAR public cloud or Amazon Web Services). Following completion of the jobs the results can be reviewed and downloaded if desired.

By providing a unified platform for science, the VL infrastructure enables sophisticated provenance capture and management. The source of input data (including both collection and queries), user information, software information (version and configuration details) and output information are all captured and managed as a VL resource which can be linked to output data sets. This provenance resource provides a mechanism for publication and citation for Free and Open Source Science.