Exploiting Data Intensive Applications on High Performance Computers to Unlock Australia's Landsat Archive

Matthew Purss (1), Adam Lewis (1), Roger Edberg (1), Alex Ip (1), Joshua Sixsmith (1), Glenn Frankish (2), Tai Chan (3), Ben Evans (4), and Lachlan Hurst (5)

(1) Geoscience Australia, Canberra, Australia (Matthew.Purss@ga.gov.au), (2) Lockheed Martin Australia, Melbourne, Australia (glenn.frankish@lmco.com), (3) Cooperative Research Centre for Spatial Information, Melbourne, Australia (tai.chan@dse.vic.gov.au), (4) Australian National University - National Computational Infrastructure, Canberra, Australia (ben.evans@anu.edu.au), (5) Victorian Partnership for Advanced Computing, Melbourne, Australia (lachlan@vpac.org)

Australia's Earth Observation Program has downlinked and archived satellite data acquired under the NASA Landsat mission for the Australian Government since the establishment of the Australian Landsat Station in 1979. Geoscience Australia maintains this archive and produces image products to aid the delivery of government policy objectives. Due to the labor intensive nature of processing of this data there have been few national-scale datasets created to date. To compile any Earth Observation product the historical approach has been to select the required subset of data and process "scene by scene" on an as-needed basis. As data volumes have increased over time, and the demand for the processed data has also grown, it has become increasingly difficult to rapidly produce these products and achieve satisfactory policy outcomes using these historic processing methods. The result is that we have been "drowning in a sea of uncalibrated data" and scientists, policy makers and the public have not been able to realize the full potential of the Australian Landsat Archive and its value is therefore significantly diminished.

To overcome this critical issue, the Australian Space Research Program has funded the "Unlocking the Landsat Archive" (ULA) Project from April 2011 to June 2013 to improve the access and utilization of Australia's archive of Landsat data. The ULA Project is a public-private consortium led by Lockheed Martin Australia (LMA) and involving Geoscience Australia (GA), the Victorian Partnership for Advanced Computing (VPAC), the National Computational Infrastructure (NCI) at the Australian National University (ANU) and the Cooperative Research Centre for Spatial Information (CRC-SI). The outputs from the ULA project will become a fundamental component of Australia's eResearch infrastructure, with the Australian Landsat Archive hosted on the NCI and made openly available under a creative commons license. NCI provides access to researchers through significant HPC supercomputers, cloud infrastructure and data resources along with a large catalogue of software tools that make it possible to fully explore the potential of this data.

Under the ULA Project, Geoscience Australia has developed a data-intensive processing workflow on the NCI. This system has allowed us to successfully process 11 years of the Australian Landsat Archive (from 2000 to 2010 inclusive) to standardized well-calibrated and sensor independent data products at a rate that allows for both bulk data processing of the archive and near-realtime processing of newly acquired satellite data. These products are available as Optical Surface Reflectance 25m (OSR25) and other derived products, such as Fractional Cover.