

Web-based Processing of Remote Sensing Data in a Workflows Environment**

Fearn Peter^{a*}, Bellgard Matthew^b, Broomhall Mark^a, Chedzey Helen^a, Garcia Rodrigo^a, Hunter Adam^b, King Edward^c, Lynch Mervyn^a, Schibeci David^b, Smith Glenelg^d

^aRemote Sensing and Satellite Research Group, Curtin University, Perth, Australia.
(p.fearn, m.broomhall, h.chedzey, r.garcia, m.lynch)[@curtin.edu.au](mailto:curtin.edu.au)

^bCentre for Comparative Genomics, Murdoch University, Perth, Australia. (mbellgard, ahunter)[@ccg.murdoch.edu.au](mailto:ccg.murdoch.edu.au)

^cCSIRO Marine and Atmospheric Research, Canberra, Australia. edward.king@csiro.au

^dCSIRO Marine and Atmospheric Research, Hobart, Australia. glenelg.smith@csiro.au

Abstract – Since its inception, Earth observation from space has been hampered by the difficulty of managing and processing the enormous volumes of data that satellite sensors are capable of generating. Modern sensors are able to produce more channels of spectral data more frequently at greater spatial resolution than ever before, and the range of potential geophysical products is continually increasing. Generation of products, access to, analysis and visualisation of these large data sets is often perceived as highly technical and an often daunting set of procedures. This paper introduces a project aimed at providing to non-technical users access to a suite of tools via a web-based graphical user interface to a Workflows system (RS-YABI), linked to a high performance computing system and a national-scale remote sensing data archive.

Keywords: Environment, Services, Facilities, Internet/Web, Workflow, Remote Sensing

1. INTRODUCTION

The AusCover component of the National Collaborative Research Infrastructure Strategy (NCRIS) Terrestrial Ecosystems Research Network (TERN) capability is focussed on delivery of remote sensing products for terrestrial ecosystems research. Simple access to these data will be enabled via tools developed by the National eResearch Architecture Taskforce (NeAT) funded project, “Workflow Services to enable a Large-Scale Temporal-Spatial Ecosystem Digital Information Service”. This NeAT project proposes to:

- Provide a standard Workflow system that will enable TERN’s AusCover data providers and users to easily specify and run custom data processing workflows that can access data from AusCover (and other data sources) and process it on remote computing resources, including the Australian Research Collaboration Service (ARCS) Grid infrastructure.
- Enable end users to easily utilise the massive computing infrastructure associated with the TERN/AusCover data management project, including tools for converting data to standard formats, running standard data processing applications on remote sensing data, and integrating TERN products with other spatial / temporal data sets (from outside the TERN project).
- Enable AusCover data providers to more easily process raw satellite data to generate standard derived products, as well as extracting standard metadata to enable improved search capability.
- Demonstrate the functionality of this generic approach with some exemplar workflows which will involve chained processing services across large numbers of data files, e.g. time-series of large spatial data arrays, and be used to generate data products for use by the researcher community.

2. REMOTE SENSING DATA

There is currently a network of satellite data reception facilities operating across Australia. Locations include Perth, Melbourne, Hobart, Townsville, Darwin and Alice Springs. Data from numerous satellites and sensors are down-linked every day and archived, but of primary interest to this project are data from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors onboard the Aqua and Terra satellites. At present, a number of national programs are endeavouring to coordinate and improve the linkages between the data reception facilities and manage the rapidly growing archives of data. By agreeing to combine data from all the reception facilities into a central repository accessible to all, or, by managing a distributed archive system, duplication of large amounts of data can be avoided, and developments in data management and data processing can be shared by all users.

There are a number of remote sensing data management and distribution programmes being undertaken at present, such as NCRIS TERN which is focussed on terrestrial and atmospheric data, and NCRIS IMOS focussed on marine data. The primary data source for both of these programmes is the national MODIS archive. One of the core aims of each of these programmes is to make remote sensing data searchable and available to end users. There are other associated activities within each programme, such as gathering and providing radiometric and environmental data sets to allow validation and quality control of the remote sensing products, but at present the underlying data processing systems are not accessible to end users, and as such these end users are not able to set up their own processing chains. Reasons end users might want to design their own processing chains include: re-run an algorithm with different input parameters, generate a data set at a different spatial resolution or geometric projection, convert data into different file formats, generate new geophysical products.

The Perth-based supercomputing facility, iVEC’s infrastructure is currently employed in this project and includes a 512 core cluster with access across a 10 gigabit per second network to a petabyte storage system which hosts a national archive of MODIS data.

* Corresponding author

** This work was supported by National eResearch Architecture Taskforce (NeAT) funding

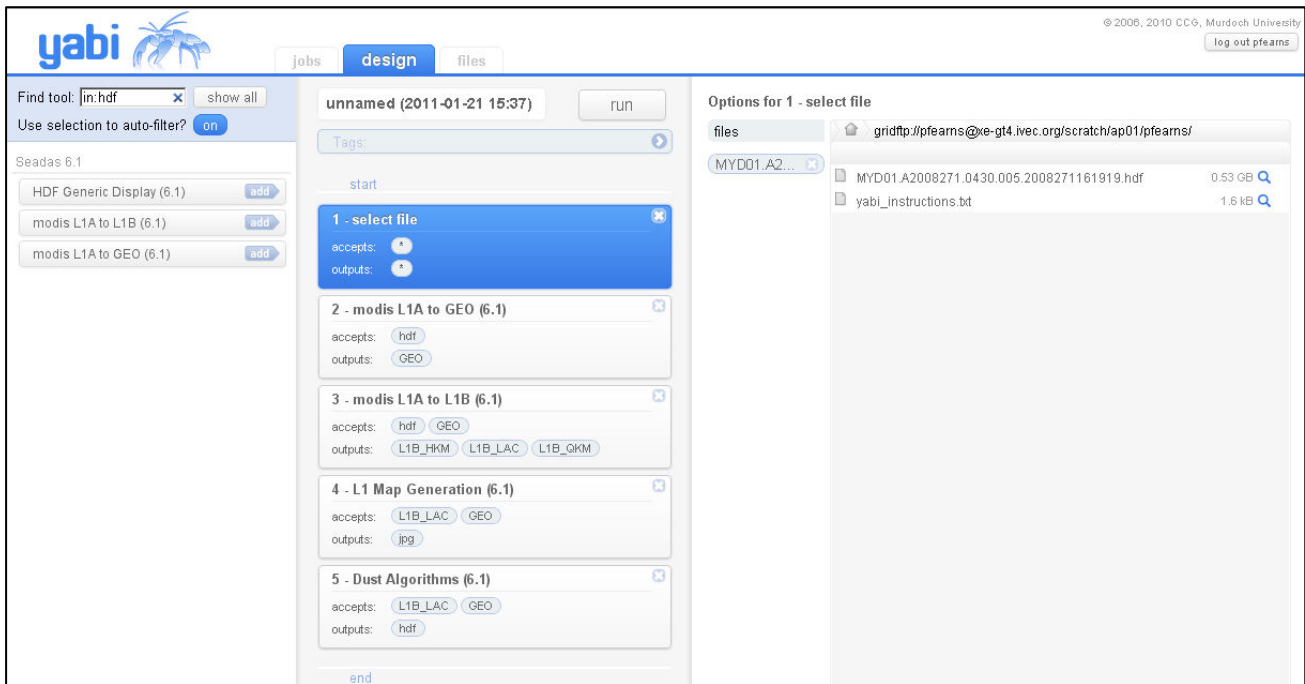


Figure 1 Screen shot of a workflow set up in the YABI GUI.

3. RS-YABI

The core to the system being developed is a Workflows “engine”, YABI, which has been developed by the Centre for Comparative Genomics at Murdoch University. The Workflow system was originally designed to undertake processing of large volumes of genetics and data, but has now been demonstrated to be easily adapted to manage large processing chains in a remote-sensing context, thus we refer to the new system as RS (Remote Sensing)-YABI.

Figure 1 shows a screen shot of a simple Workflow within the RS-YABI GUI. There are three tabs at the top of the screen labelled jobs, design and files. The “jobs” tab provides access to the results of previously run jobs. Data from intermediate steps may be accessed, and saved jobs may be altered and re-used. The “design” tab, as shown in Fig. 1, presents the user with a list of modules on the left hand side, where each module may be selected in turn to build a Workflow in the middle panel of the screen. The Workflow in Fig. 1 begins with a file selection, followed by generation of GEO and L1B MODIS files, a mapping procedure, and finally application of the dust storm algorithm module. The “files” tab provides a simple user interface for access to data sets on the network.

Another aspect to the management of large sets of remote sensing data is the search and access process. This project is utilising the tools developed by WASTAC (WASTAC 2010) to provide consortium members access to the MODIS data archive hosted at iVEC (WASTAC, 2011). Figure 2 shows a screen shot of the data search and delivery tool developed by WASTAC. This interface provides search and access to AVHRR and MODIS data held at iVEC in the WASTAC archive. Users may

select date ranges, sensor and satellite, geographical region, output data formats and some quality control parameters such as percent cloud cover.

4. DATA PROCESSING AND APPLICATIONS

The current version of RS-YABI is being developed with the aim of delivering data and computing resources to a selected number of end users, in particular those with interests in bushfire smoke plumes, dust storm clouds, fire and near-shore ocean temperature. Papers by Chedzey et al.(this volume) and Broomhall et al.(this volume) expand on the processing and delivery of bushfire smoke and dust storm products respectively. All products derived from MODIS data, land, ocean or atmospheric, depend on a number of common processing tools. The remote sensing data processing applications embedded within RS-YABI include components from numerous freely available remote sensing data tools. Some of these are listed in Table 1.

Table 1 Components of the remote sensing data processing software listed here are embedded within RS-YABI

SeaDAS	http://seadas.gsfc.nasa.gov/
IMAPP	http://cimss.ssec.wisc.edu/imapp/
IPOPP	http://directreadout.sci.gsfc.nasa.gov/links/rsd_eo_sdb/PDF/IPOPP_1.6b_Users_Guide.pdf
ms2gt	http://nsidc.org/data/modis/ms2gt/
mrtswath	https://lpdaac.usgs.gov/lpdaac/tools/modis_reprojection_tool_swath
DRL	http://directreadout.sci.gsfc.nasa.gov/

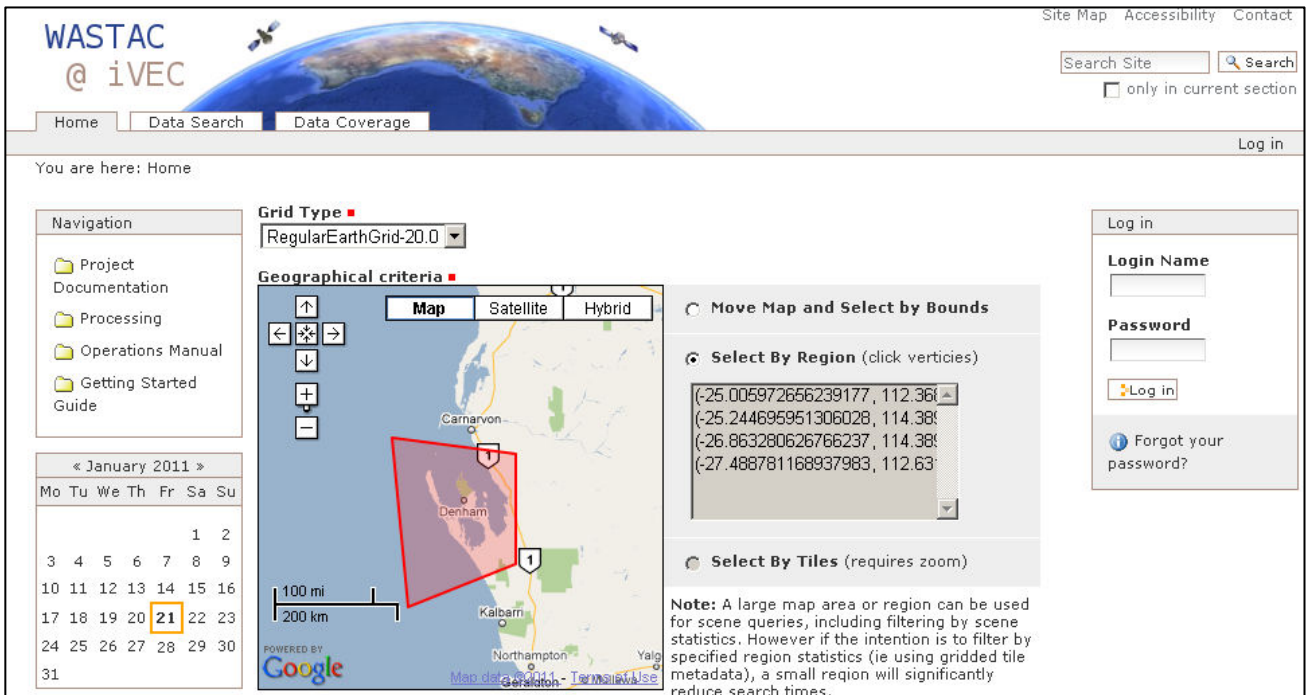


Figure 2 Screen shot of the WASTAC data search and delivery tool

Figure 3 shows an outline of the RS YABI Workflow structure. Each box in Fig. 3 represents one of the selectable modules located to the left of the RS YABI interface (Fig. 1). The paths shown in Fig 3 are shown as typical examples of different Workflows. In practical terms, the Workflow may begin and end at any module in the workflow structure (once the input file or directory has been selected).

The intent with this project is to enable users access to the large remote sensing data base, to allow them flexibility in designing their own processing pathways and to provide some tools to

allow quick and simple analysis of data. To this end we have included modules to produce time series plots of data, individual images (in numerous image formats), as well as “movie loops” of imaged products. Plans are being developed to increase the functionality to include outputs of simple statistical information and simple time series analysis of data.

5. CONCLUSIONS

We have demonstrated the application of an advanced GUI workflow system designed for bioinformatics processing which has been adapted to process remote sensing data. A number of

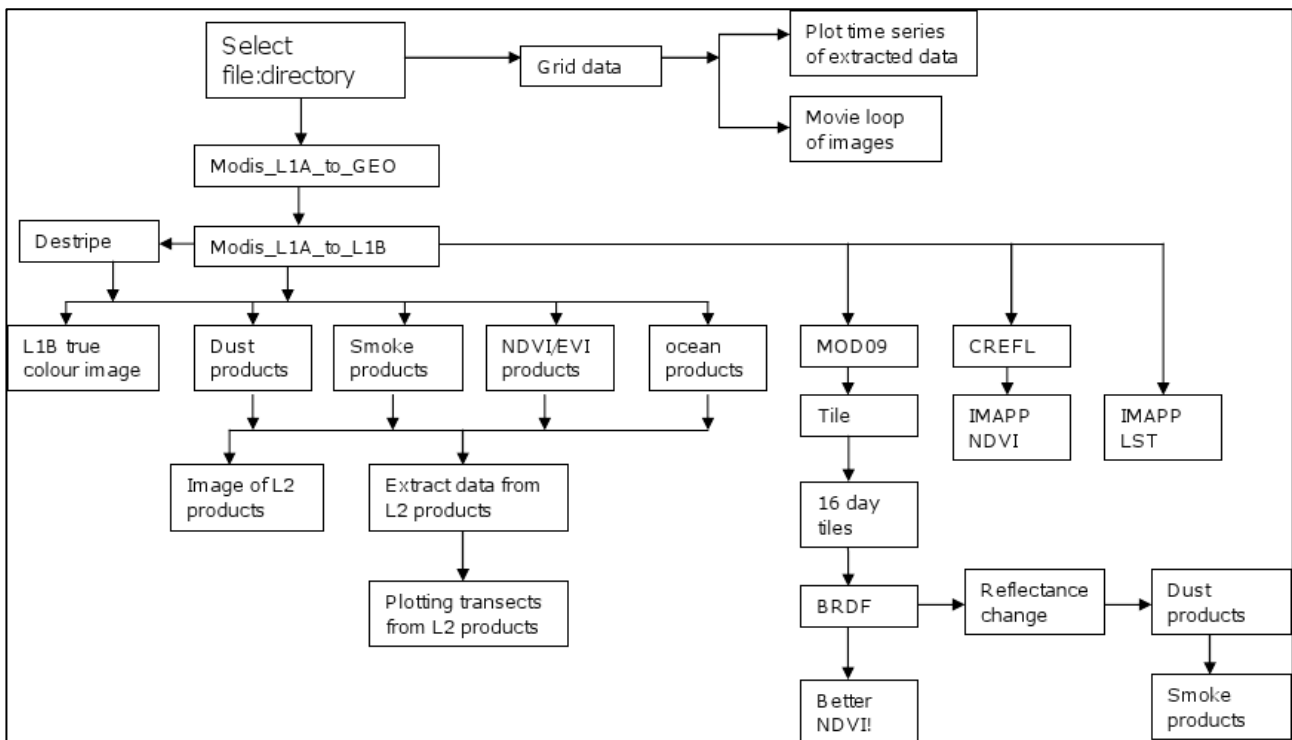


Figure 3 Overview of the RS YABI workflow structure

modules have been integrated into the Workflow to provide users with limited technical knowledge access to the massive computing potential of an HPC system as well as the national archive of raw satellite data attached to the system.

The development and design of the RS-YABI has initially been directed by the experience and expertise of the core team assigned to the project, but the vision for further development is that close interaction with potential end-users will provide input into the development process. We believe this feedback process will help guide the RS YABI design process to deliver a tool that works in a way that the end users understand, and is therefore potentially more useful to end users than a tool designed by remote software engineers.

6. REFERENCES

M. Broomhall, H. Chedzey, R. Garcia, M. Lynch, P. Fearn, E. King, Z. Wang, G. Smith, D. Schibeci, 2011, Ensemble Dust Detection Techniques Utilising a Web-Based Workflow Environment Linked to a High Performance Computing System

Chedzey H., Fearn P., Broomhall M., Garcia R., Lynch M., King E., Smith G., Schibeci D., 2011, Processing Smoke Plume Products from The Moderate Resolution Imaging Spectroradiometer (MODIS) within a Workflows Environment

WASTAC, R. 2010, <http://www.wastac.wa.gov.au>

WASTAC, 2011, <http://wastac.ivec.org/>

Feldman, 2011, <http://seadas.gsfc.nasa.gov/>

7. ACKNOWLEDGEMENTS

This project was funded primarily by the National eResearch Architecture Taskforce, but support, both financial and in-kind, has also been provided by the NCRIS TERN AusCover Project, WASTAC, iVEC, Curtin University, Murdoch University, CSIRO and ARCS.