crc•si Spatially Enabling Australia and New Zealand

P2.23 | Open Spatial Analytics

Project Leader Associate Prof Chris Bellman, School of Science, RMIT University

Research Team Prof Matt Duckham, Dr Alan Both, Abdur Forkan, Associate Prof Gen Tian, Yaguang Tao, RMIT University; Hamish Anderson, Geoscience Australia; Prof Ilkay Altintas, San Diego Supercomputer Centre.

Project Participants RMIT University
Geoscience Australia
San Diego Supercomputer Centre

Objective This project will build a small demonstrator tool to show the capabilities and benefits of adopting open scientific workflow models for spatial analytics.

Outcomes

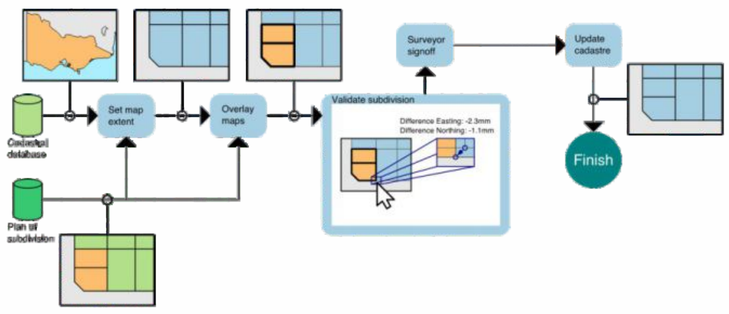
- A visual language for asking questions about spatial data instead of performing operations
- A graphical workbench for flexible and intuitive construction of spatial analytics procedures
- An example workflow based on Geoscience Australia's procedure for constructing national elevation datasets.

Overview

Creating value from spatial data requires expertise and intelligence. Often, this expertise is buried in manuals or locked in the heads of the expert analysts.

This project adapts scientific workflow systems to solve the problem. The resulting tools:

- Capture expert knowledge;
- Increase the warrantability of spatial data products produced; and
- Increase opportunities for collaboration on the spatial analytics.

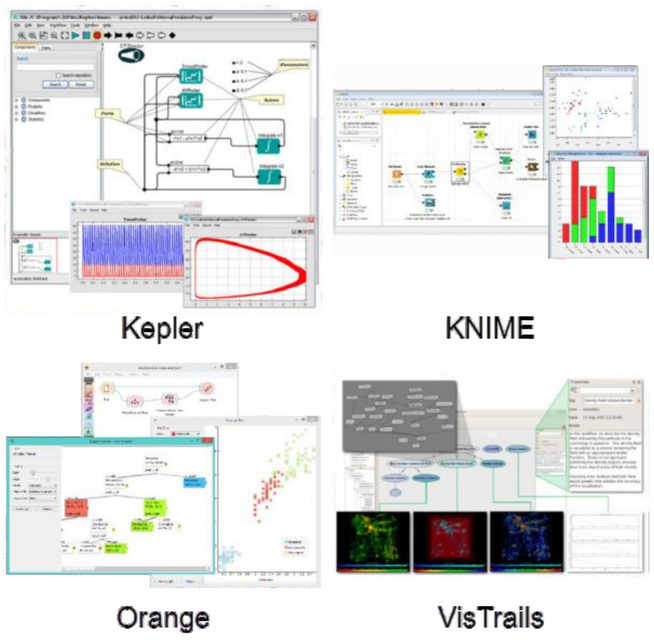


Workflow mockup showing the process of updating a cadastre with a new subdivision

Scientific workflows

Scientific workflows are interactive and intuitive graphics for building analytics. In short, scientific workflows provide a visual language and self-documenting analytics.

Example workflow systems are shown below:



A selection of currently available scientific workflow systems

We investigated 20 systems, establishing a set of criteria before any particular platform was chosen.

These criteria were:

- Open source and cross-platform;
- Developed in a popular programming language;
- Ease of developing a geospatial extension;
- Cloud computing enabled; and
- An intuitive interface with real-time progress indicators.

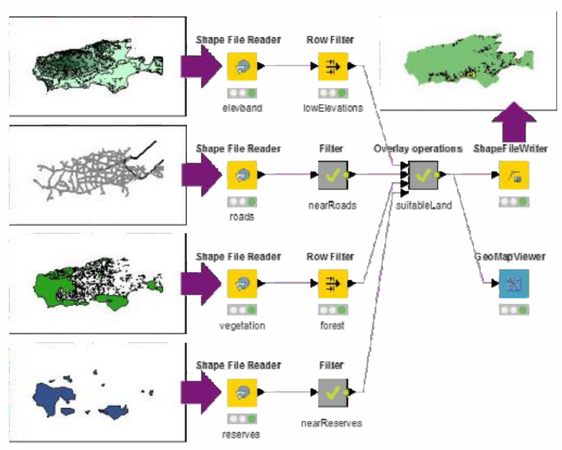
Proprietary systems, such as FME and ESRI Model Builder were also investigated to determine the capabilities of commercially available solutions.

After a thorough investigation of potential systems, we found that KNIME best meet our project's aims.

Integration

A key benefit of using scientific workflow systems is access to their wide array of modules, from machine learning algorithms, to integration with the statistical computing language R.

Presently, these systems have very limited geospatial capabilities. To address this, we integrated the open source geospatial library Geotools as a set of modules within KNIME. The figure below shows a example workflow making use of some of these modules:



Example KNIME workflow making use of geospatial operations to find suitable farming land in Kangaroo Island

Integrating Geotools allows us to combine geospatial analysis with other forms of analytics as well as allowing us to develop a visual language for asking questions about spatial data instead of just performing spatial operations.

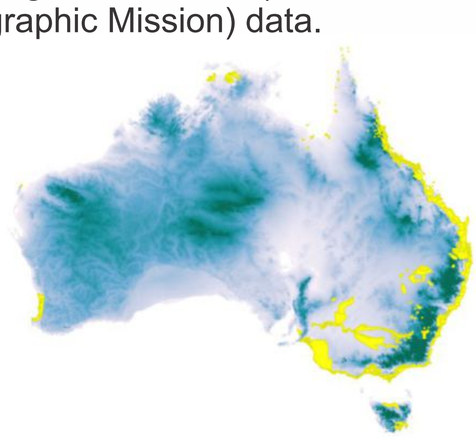
In addition to Geotools, we have integrated a number of other tools including:

- The open source spatial database PostGIS;
- Docker, for encapsulating proprietary and legacy software;
- WMS and WFS (Web Mapping and Feature Services); and
- The provenance management system PROMS.

Case study

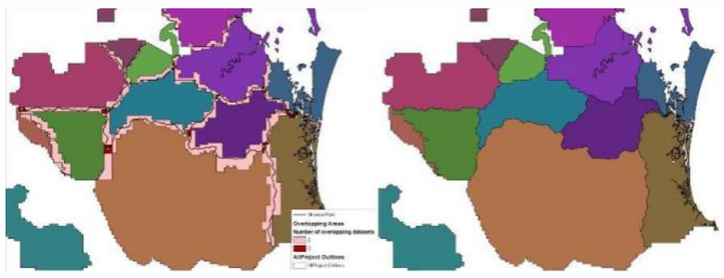
To demonstrate the benefits of workflows, we are implementing Geoscience Australia's procedure for constructing the National DEM (digital elevation model). This task involves:

- Processing over 100TB of high precision LiDAR point cloud data from over 60 different surveys;
- Stitching that data into a seamless mosaic; and
- Blending it with SRTM (Shuttle Radar Topographic Mission) data.



Extents of LiDAR surveys overlayed on SRTM data

This task is currently processed on a single workstation, making use of a collection of proprietary and open source software, some of which is automated through scripts and some which must be processed manually.



Process of stitching LiDAR surveys into a single DEM
Source: Geoscience Australia

We are completely automating the process, using the NCI supercomputer, can quickly regenerate Australia's whenever new LiDAR data is acquired.