Enhance Reusability and Reproducibility using NCI's Provenance Capturing System

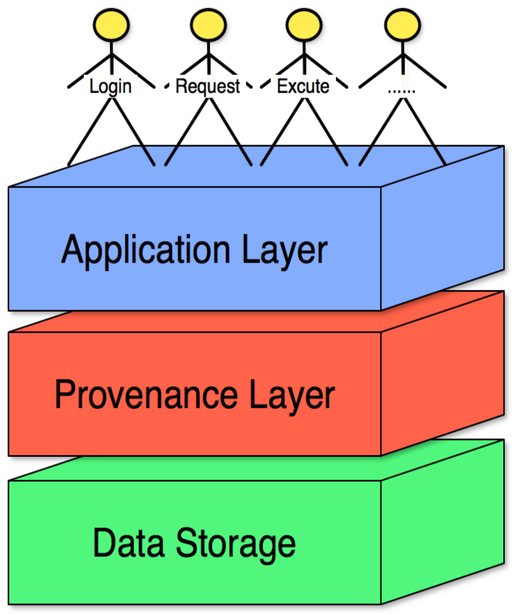
Ben Evans[1], Nick Car[2], Lesley Wyborn[1], Jingbo Wang[1], Wei Si[1]
[1] National Computational Infrastructure, Acton, ACT; [2] Geoscience Australia, Canberra, ACT

The Need for a Provenance Workflow Service

Data publication and citation have attracted considerable attention due to increasing demand to reproduce and validate scientific research outputs. Properly designed, a Provenance Workflow Service has the ability to encode transparency and accountability by providing the capability to record the key dependencies and decisions of any part of a scientific workflow, thus ensuring repeatability and trust.

NCI is using the PROvenance Management System (PROMS)[1] which provides both toolkits in a selection of programming languages for producing provenance reports compliant with PROV, the World Wide Web Consortium's provenance representation standard[2], and a storage system. The PROMS storage system can be queried for individual reports and elements within those reports which then link back to information in the reporting system and the data storage mechanisms it uses.

We see a well-designed comprehensive data management portal together with a provenance workflow service as an integral way of capturing the evolution of the data throughout the full data life cycle, including phases of data downloading, pre-processing and processing, re-processing and ensemble analysis. The NCI PROV server is currently in development and is available from http://proms-dev.nci.org.au/ .



Provenance Architecture

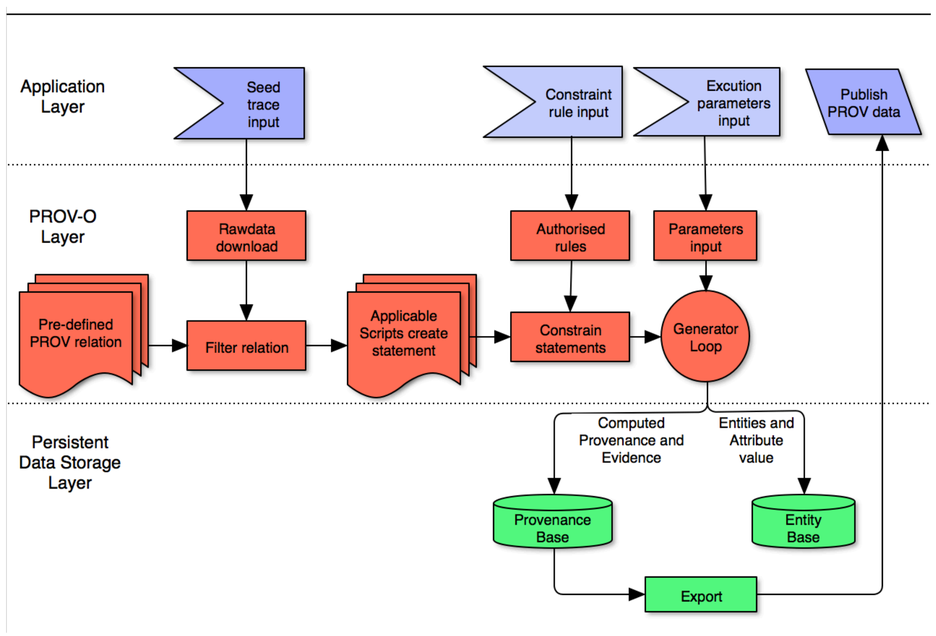
Application Layer: User interaction

Provenance Layer: based on PROV-O, harvest data by PROV relations, and execution of rules and parameters inputs

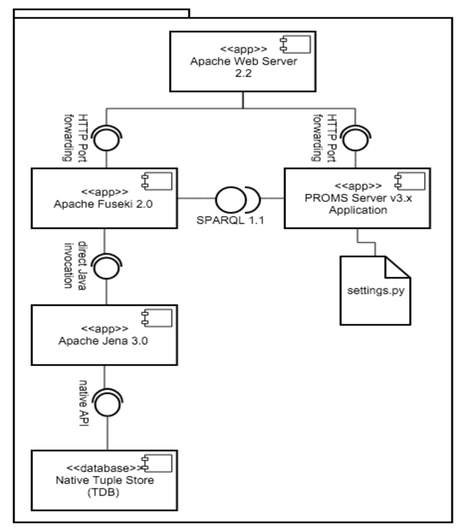
Data Storage Layer: Graph DB and export the provenance data or generate reports

References:
[1]Car, N.J. (2013) A method and example system for managing provenance information in a heterogeneous process environment – a provenance architecture containing the Provenance Management System (PROMS). In J Piantadosi, R.S. Anderssen, and J. Boland, editors, MODSIM2013,20th International Congress on Modelling and Simulation, December 2013, Adelaide, Australia. Modelling and Simulation Society of Australia and New Zealand
[2]Lebo, T., Sahoo, S., & McGuinness, D. (2013). PROV-O: The PROV Ontology. http://www.w3.org/TR/prov-o/. Accessed 1 December, 2015.

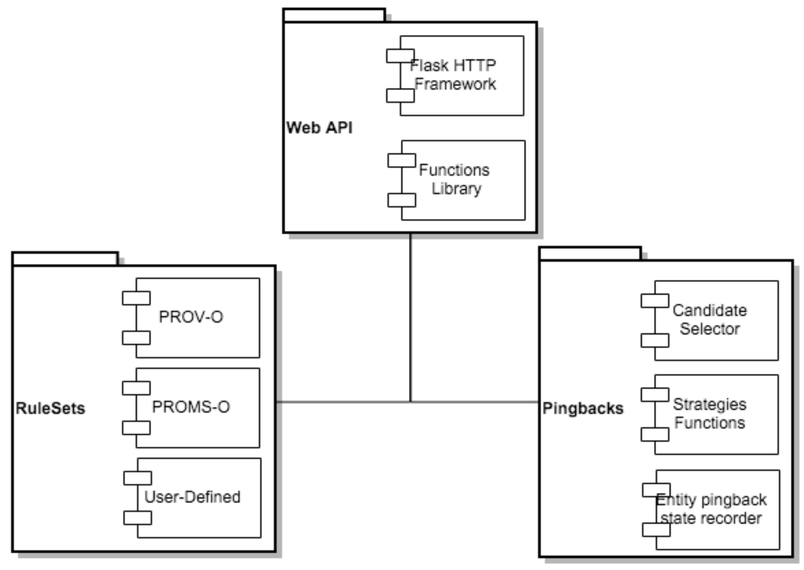
NCI's Provenance Architecture



PROMS Server Architecture

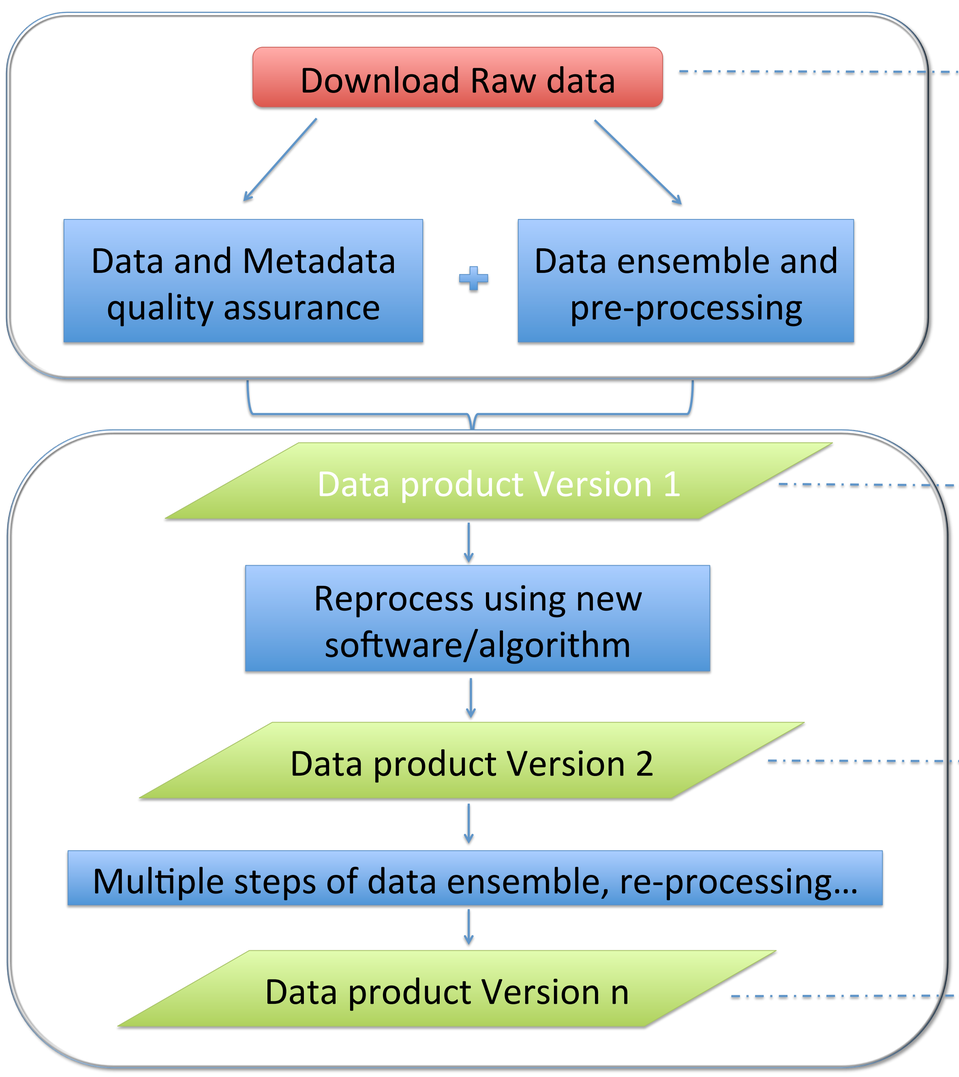


Top-Level Installation

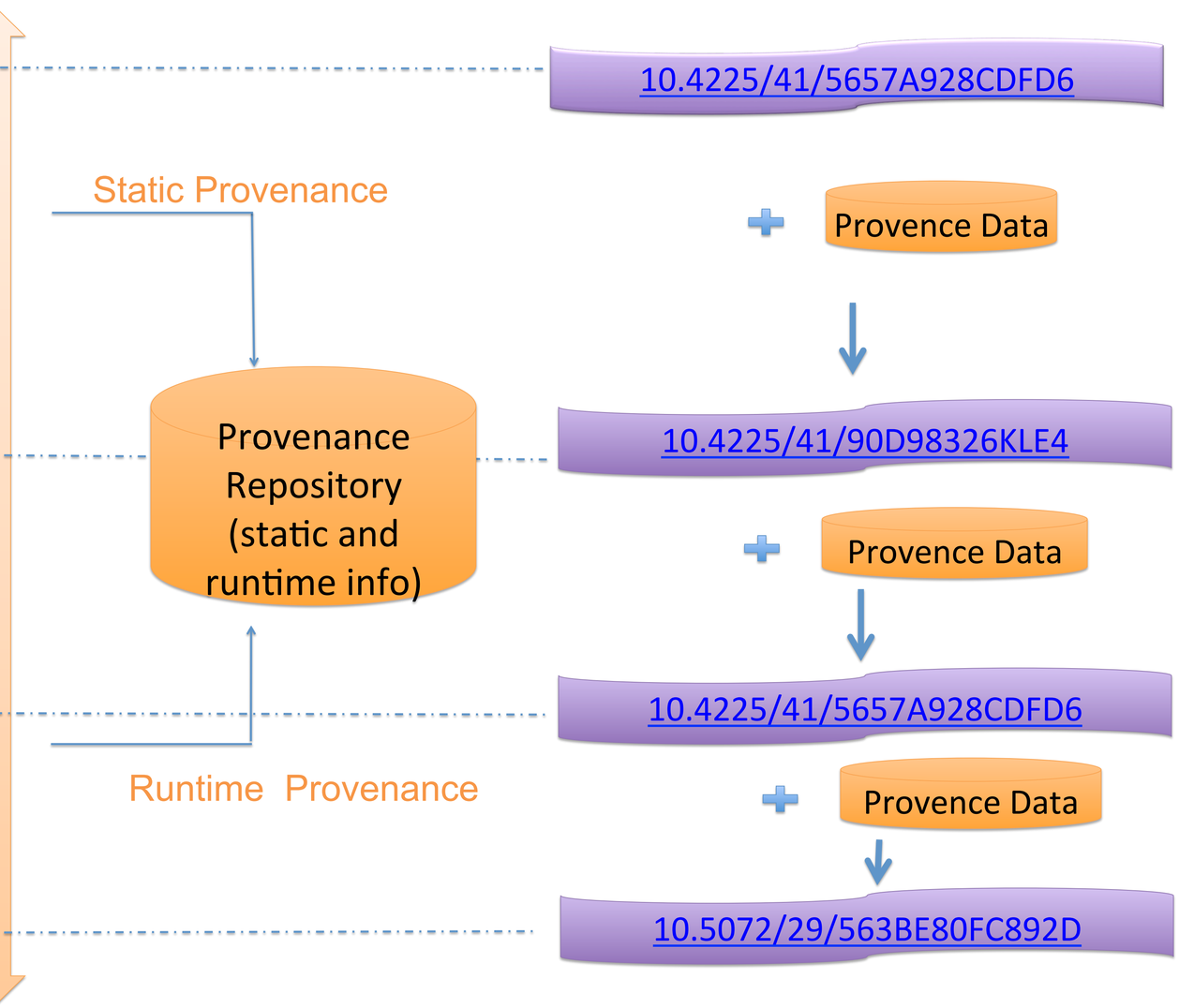


Application Components

Conceptual Climate Modeling Scientific Workflow



Large Scale Dynamic Citation Management



The Provenance Service captures information at each step within the end-to-end workflow, and stores it within the Provenance Repository