

Implementing an organisation-wide approach to provenance management for Geoscience Australia

Fitch, P. ^a, Car, N.J. ^{a*} and Fyfe, S.^b

^a CSIRO Land & Water, Canberra ACT & Brisbane QLD, Australia

^b Geoscience Australia, Symonston, ACT, Australia

Email: nicholas.car@csiro.au

Abstract: Geoscience Australia (GA) is Australia's federal government custodian of national geographic and geological data. It is a science organisation that produces a diverse set of scientific information products such as reports, regular datasets, forecasts and papers. It aims to ensure scientific product quality and transparency according to its *Science Principles*. To do this, GA has embarked on a whole-of-organisation assessment of its product's provenance. Following from that and over the next few years, GA intends to build the systems and skills necessary to capture, manage and use provenance information for many of its products in best-practice and standardised ways.

In this presentation we describe the work undertaken to produce a *Provenance information implementation plan* which is part of GA's product provenance assessment. The work needed to confirm that a consistent approach to the capture and use of provenance information across GA is possible and sensible and, if so, provide GA with a workplan with sufficient detail to guide the first phases of systems and procedures implementation. This work built on a previous provenance scoping study commissioned by GA that recommended it engage with its leadership to articulate the motivation(s) for investment in the area of provenance and the use of a particular provenance information interchange standard, the World Wide Web Consortium's PROV.

This work started with a workshop at GA in November, 2014, open to interested staff in order to understand the similarities in provenance situations and needs across several GA project groups. Significant interest in and the need for managing provenance information within GA was seen clearly as was the potential benefit in applying a systematic and coordinated approach to dealing with it across the organisation. Following the workshop, interviews were conducted with a range of process owners and IT staff who are likely to be stakeholders in a potential future provenance management regime. From the interview questions, we were able to not only identify the specific requirements of provenance information, but also to identify and map into PROV the workflows used to generate several key, very different, products. This showed that describing a heterogeneous set of GA's product generation processes using the single provenance data model previously recommended for use was possible.

After interviews, a framework was developed to 'provenance enable' any GA product by modelling its generation processes and determining the work required to extend them in order to create and transmit provenance information to enterprise provenance architecture. Templates for subsystem components' evaluation regarding their ability to interact with the enterprise provenance architecture were created and descriptions of a small set of enterprise provenance tools and methods that GA might implement were also provided. The tools chosen were either well known publicly available metadata and storage tools or specific provenance management tools from some of the authors' previous work.

Finally, a detailed project plan was created so GA might have an estimate of specific costings and timings associated with initially some pathfinder products' and then latterly all products' provenance enablement.

Keywords: *Provenance, PROV, Geoscience Australia, organisation-wide approach*