**Facilitating stewardship of scientific data through standards based workflows**

Irina Bastrakova1, Carina Kemp1, Anna Kathryn Potter1

**IN53C-1572**

There are main suites of standards that can be used to define the fundamental scientific methodology of data, methods and results. These are firstly Metadata standards to enable discovery of the data (ISO 19115), secondly the Sensor Web Enablement (SWE) suite of standards that include the O&M and SensorML standards and thirdly Ontology that provide vocabularies to define the scientific concepts and relationships between these concepts. All three types of standards have to be utilised by the practicing scientist to ensure that those who ultimately have to steward the data stewards to ensure that the data can be preserved curated and reused and repurposed. Additional benefits of this approach include transparency of scientific processes from the data acquisition to creation of scientific concepts and models, and provision of context to inform data use.

Collecting and recording metadata is the first step in scientific data flow. The primary role of metadata is to provide details of geographic extent, availability and high-level description of data suitable for its initial discovery through common search engines.

The SWE suite provides standardised patterns to describe observations and measurements taken for these data, capture detailed information about observation or analytical methods, used instruments and define quality determinations. This information standardises browsing capability over discrete data types. The standardised patterns of the SWE standards simplify aggregation of observation and measurement data enabling scientists to transfer disintegrated data to scientific concepts.

The first two steps provide a necessary basis for the reasoning about concepts of ‘pure’ science, building relationship between concepts of different domains (linked-data), and identifying domain classification and vocabularies.

Geoscience Australia is re-examining its marine data flows, including metadata requirements and business processes, to achieve a clearer link between scientific data acquisition and analysis requirements and effective interoperable data management and delivery. This includes participating in national and international dialogue on development of standards, embedding data management activities in business processes, and developing scientific staff as effective data stewards.

Similar approach is applied to the geophysical data. By ensuring the geophysical datasets at GA strictly follow metadata and industry standards we are able to implement a provenance based workflow where the data is easily discoverable, geophysical processing can be applied to it and results can be stored. The provenance based workflow enables metadata records for the results to be produced automatically from the input dataset metadata.