Australian Government
Geoscience Australia

2013 AGU Fall Meeting IN53C-1572

Facilitating stewardship of scientific data through standards based workflows

Irina Bastrakova[1], Carina Kemp[1], Anna Potter[1]

[1]Geoscience Australia

Introduction

Three main types of standards can be used to define the fundamental scientific workflow from data, methods and results:

- Firstly, metadata standards to enable discovery of the data (ISO 19115);
- Secondly, the Sensor Web Enablement (SWE) suite of standards that includes the Observations and Measurements (O&M) and SensorML standards; and
- Thirdly, various ontologies that provide vocabularies to define the scientific concepts and relationships between these concepts.

Ultimately, all three standard types are applied by the practicing scientist in line with their data stewardship responsibilities:

- To ensure that any input data can be preserved and curated for the longer term, and
- To enable reuse and repurposing of the data by others beyond what the original data was collected for.

Additional benefits of applying such a standards-based approach include transparency of scientific processes from the data acquisition to creation of scientific concepts and models, and provision of context to inform data reuse and repurposing.

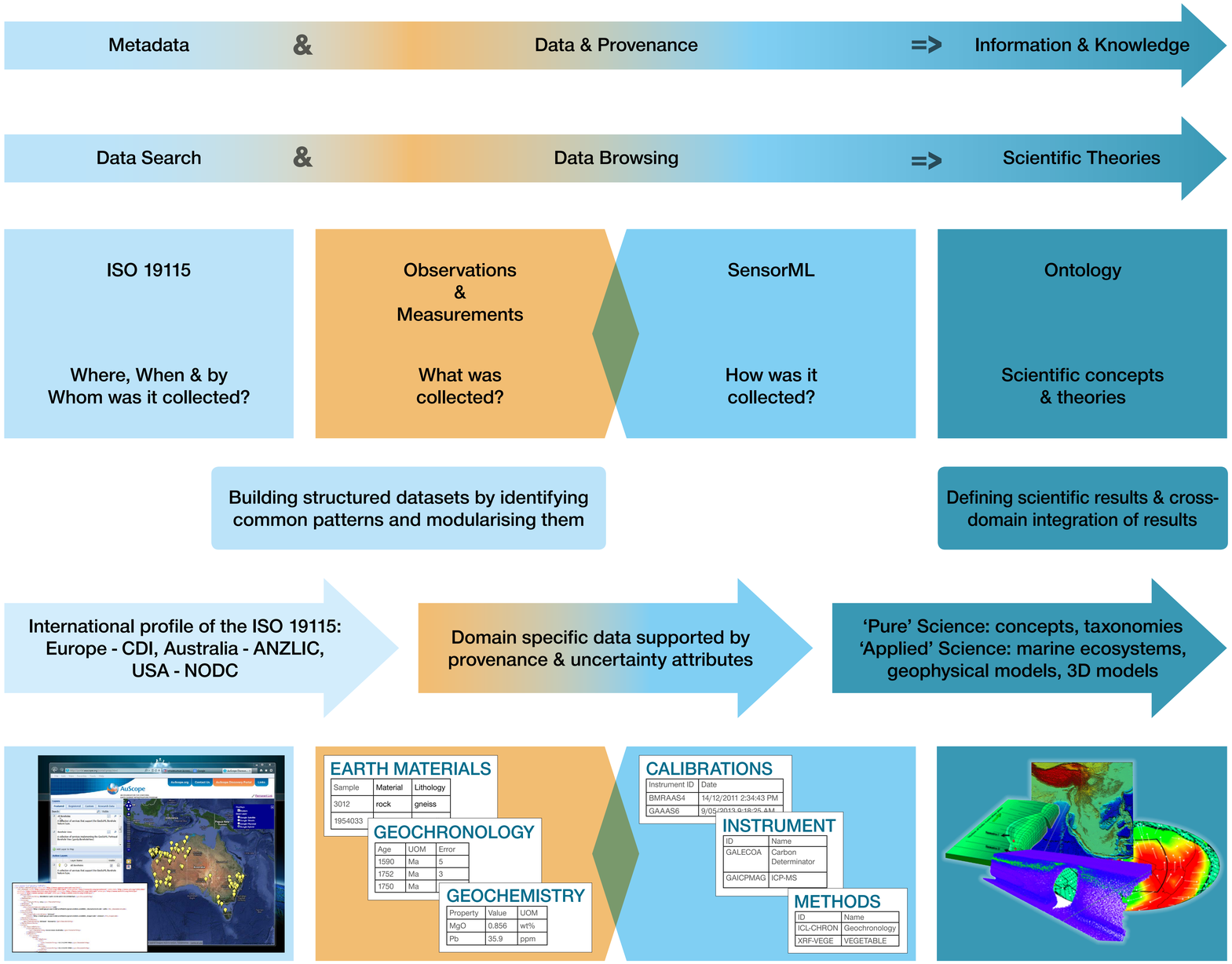
Collecting and recording metadata that enable discovery of data is the first step in the scientific workflow. The primary role of such metadata is to provide details of geographic extent, availability and a high-level description of data suitable for its initial discovery through common search engines. The ISO 19115 standard is commonly used for this step.

Once data are discovered, the search can be further refined using the second type of standards, the SWE suite. SWE provides standardised patterns to describe the observations and measurements taken for data, to capture detailed information about observation or analytical methods and the instruments used, and to define quality and uncertainties. Such information enables standardised browsing over discrete data types across multiple scientific domains. The standardised patterns of the SWE simplify aggregation of various observational data sets and to support research across diverse scientific concepts.

Ontologies, the third type of standards, provide a necessary basis for the reasoning about concepts of ‘pure’ science, and enable linking between concepts from different scientific domains using across domain-specific classifications and vocabularies (linked-data).

Geoscience Australia (GA) 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 data management and delivery. This includes participating in national and international dialogues on development of standards, embedding data management activities in business processes, and developing scientific staff as effective data stewards.

A similar approach has being applied to the geophysical data at GA. By ensuring the geophysical datasets strictly follow metadata and industry standards, a provenance based workflow has been implemented. This has made data easily discoverable and geophysical processing tools can be applied to the data. The provenance based workflow also enables metadata records for the results to be produced automatically from the input dataset metadata.



For Further Information: Irina Bastrakova
Email: irina.bastrakova@ga.gov.au
Ph: +61 2 6249 9201 **Web:** www.ga.gov.au