# Keep Calm and Follow the Standard

The Art of Using and Adapting Standards to locate data

**Margie Smith**

Science Data Governance & Policy
*Informatics | Science Data*
*Digital Science and Information*

Geoscienc

# The presentation

- Why use a metadata standard

- The Geoscience Australia's data catalogue - eCat

- Extending the standard

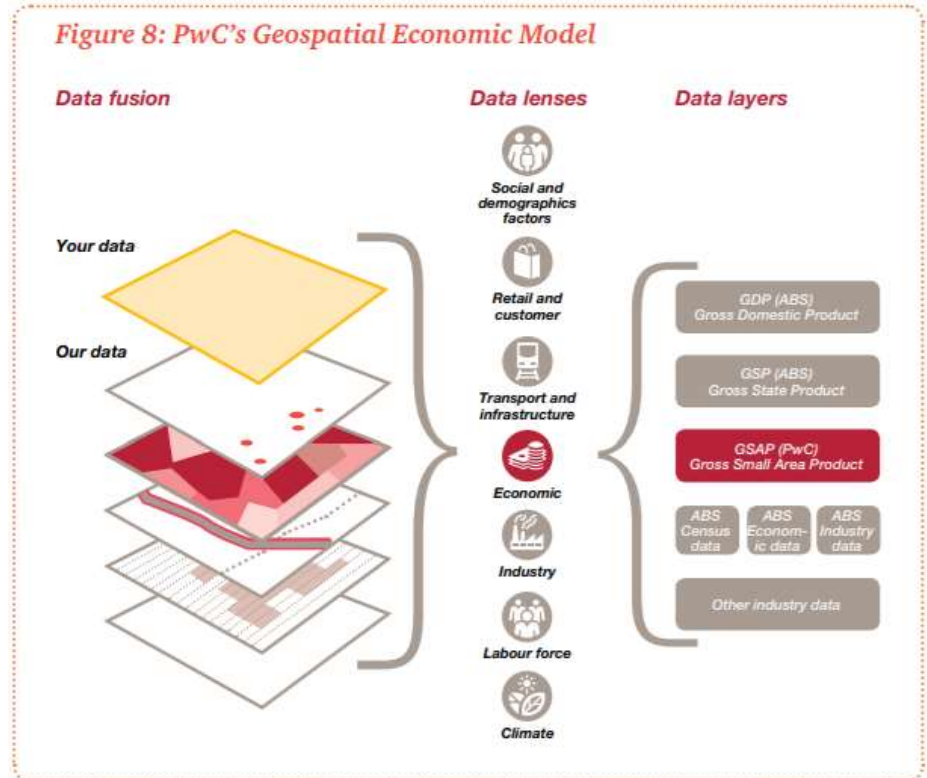- Making our catalogue available to other domain standards

# Why use a standard in business



Data-driven innovation was worth **$67billion** to the Australian economy in 2013.
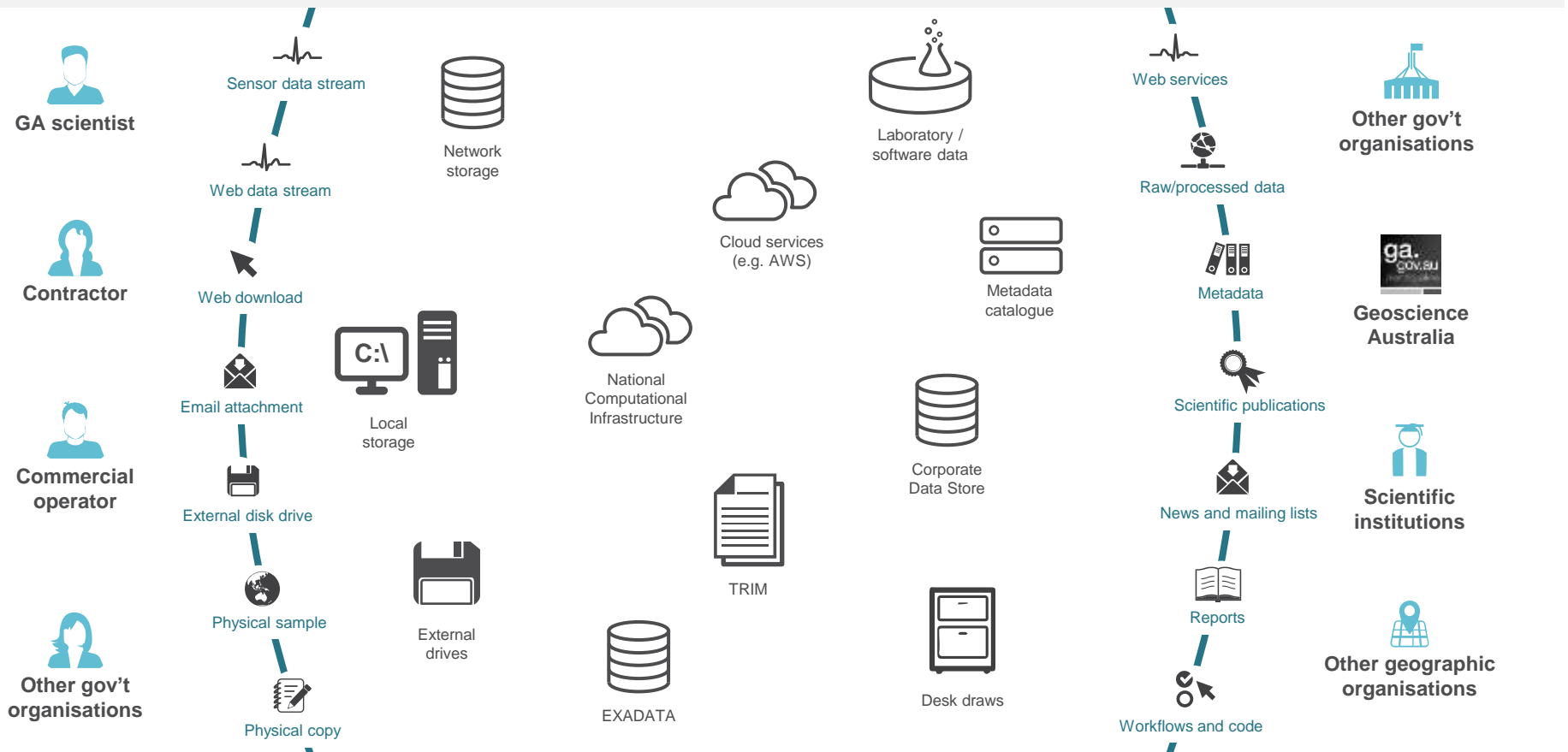


Figure 8: PwC's Geospatial Economic Model

# Why GA cares about standards

- Finding data from projects in the past

- Organisational science principles and operational pillars

- Archive legislation – data are records

- Australian Government Open Data Policy – CC-By by default
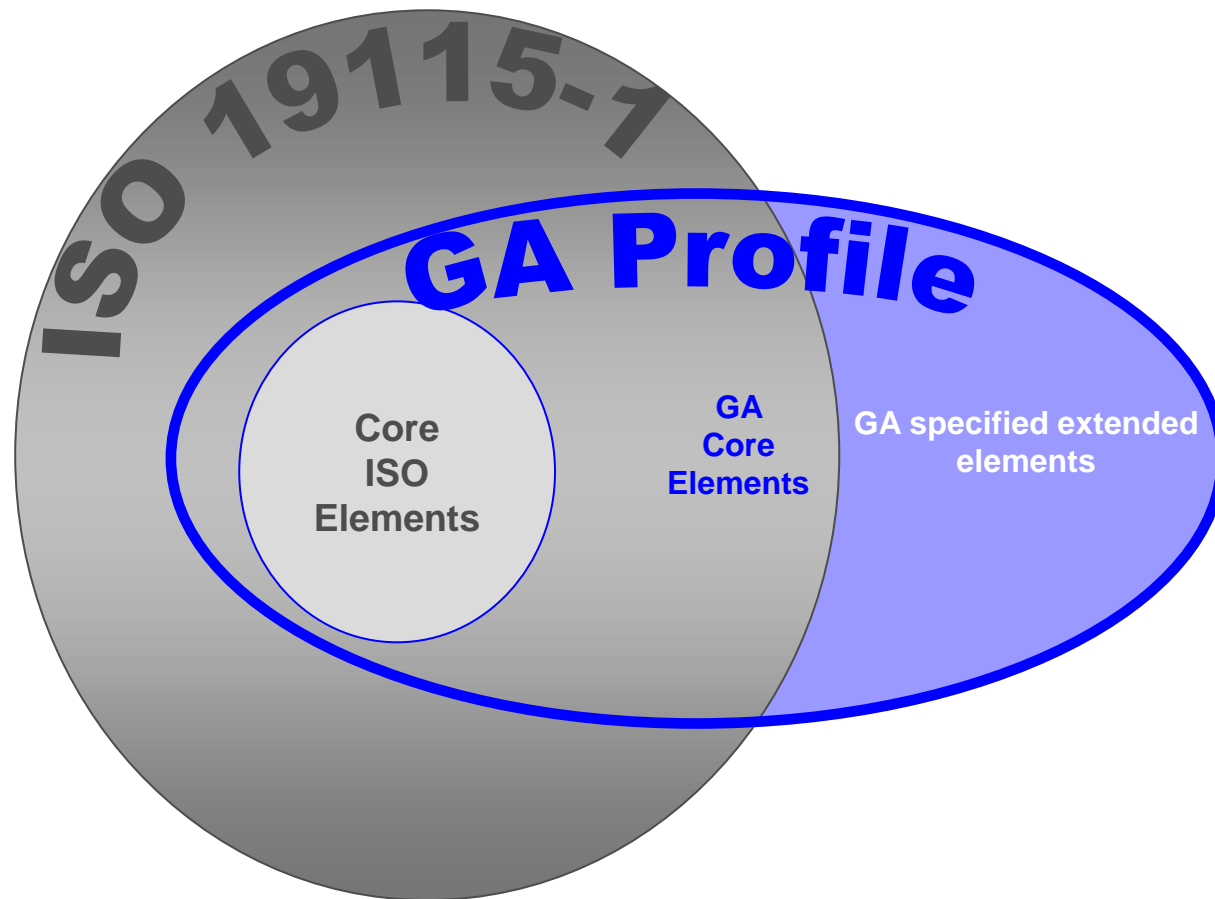
# Finding the data

# The GA Catalogue - eCat

Our enterprise catalogue is based on an international geographic metadata standard – the ISO 19115-1:2014.

- Provide appropriate information to characterise geographic data properly

- Facilitate the organisation and management of geographic information

- Enable users to efficiently apply geographic data by knowing the data basic characteristics

- Enable users to locate, access, evaluate, deliver and integrate geographic data
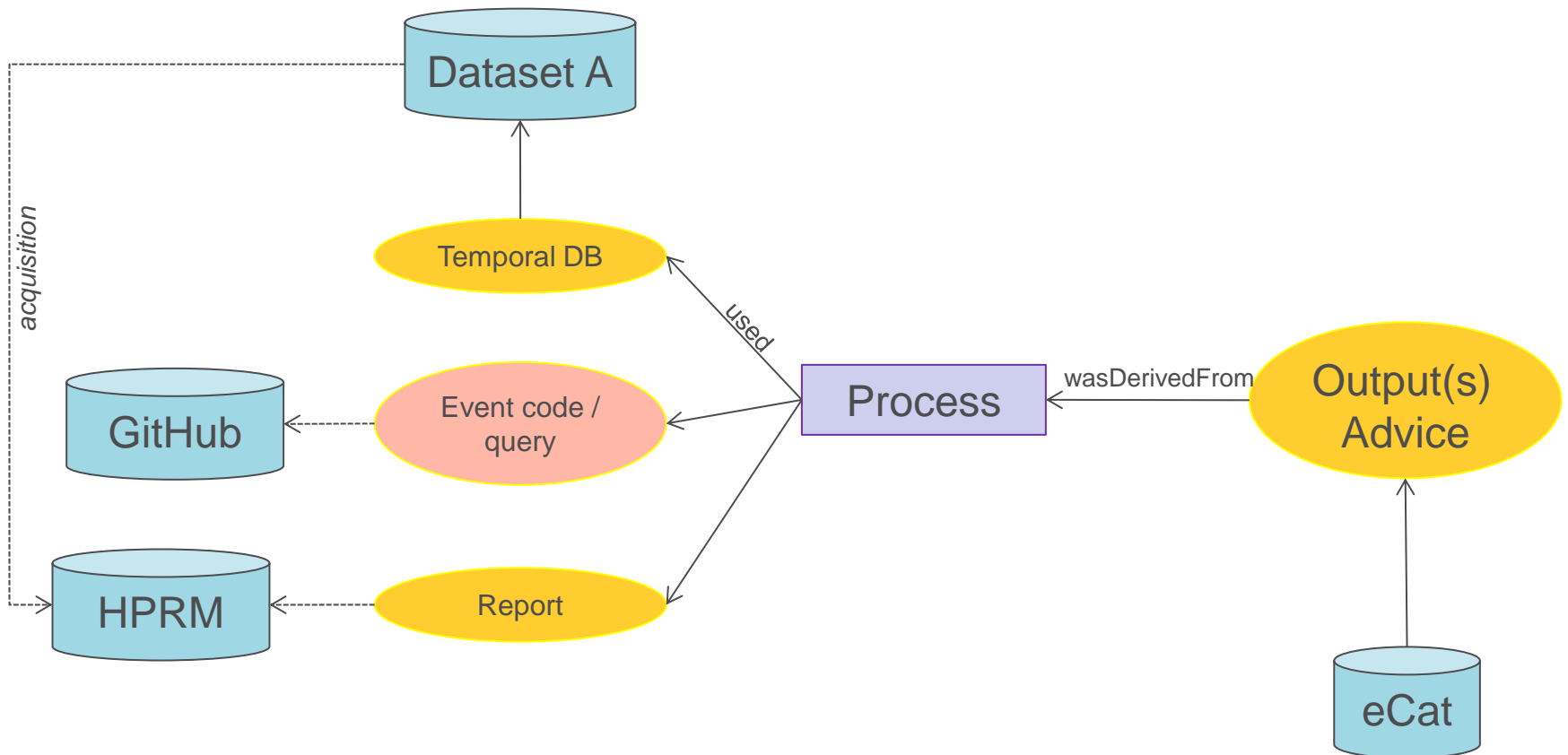
# Extending the standard

# Embedding provenance
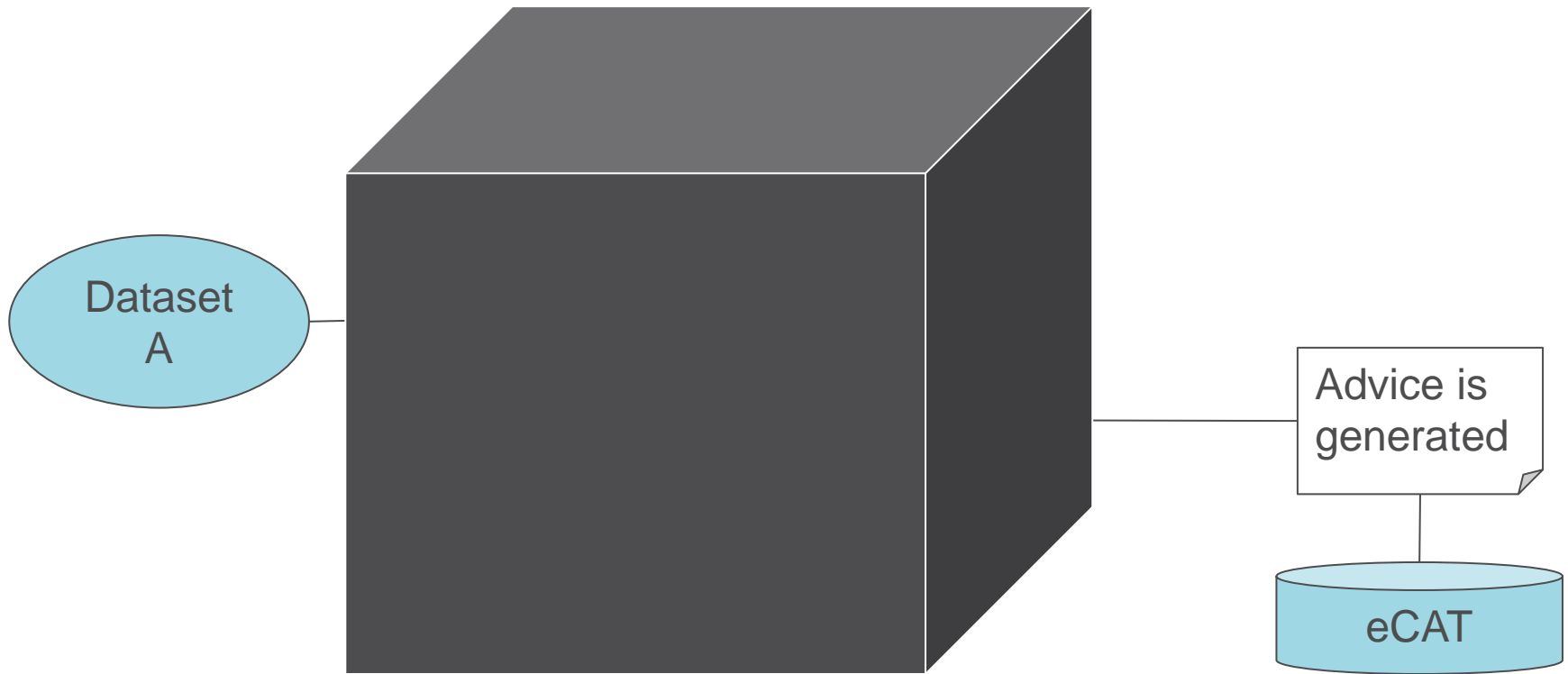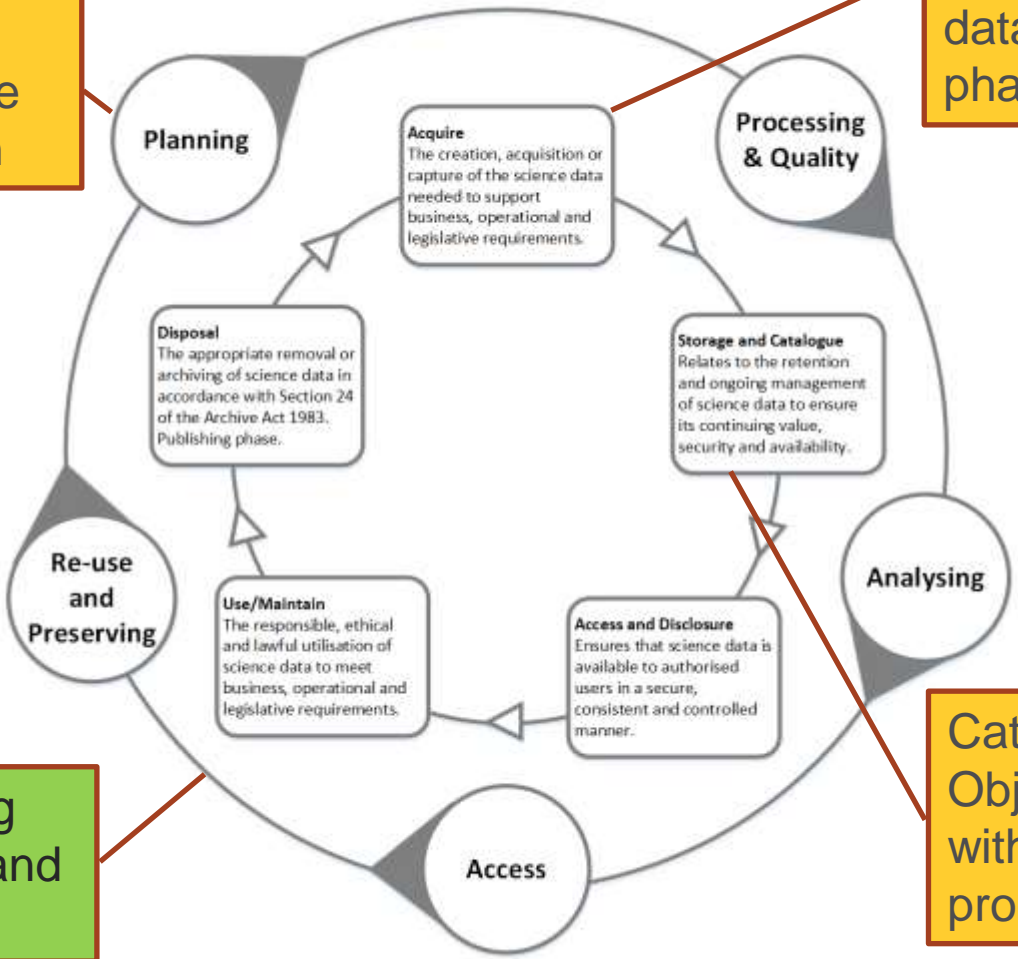
# Metadata elements for data re-use

© Commonwealth of Australia
(Geoscience Australia) 2018

# Supporting the 'Data-First' approach



**Science Data Lifecycle**

Automated catalogue and storage for the data acquisition phase

Data Management Planning to prepare for data acquisition

**Planning**

**Acquire**
The creation, acquisition or capture of the science data needed to support business, operational and legislative requirements.

**Processing & Quality**

**Disposal**
The appropriate removal or archiving of science data in accordance with Section 24 of the Archive Act 1983. Publishing phase.

**Storage and Catalogue**
Relates to the retention and ongoing management of science data to ensure its continuing value, security and availability.

**Re-use and Preserving**

**Use/Maintain**
The responsible, ethical and lawful utilisation of science data to meet business, operational and legislative requirements.

**Access and Disclosure**
Ensures that science data is available to authorised users in a secure, consistent and controlled manner.

**Analysing**

eCat for publishing the data, objects and data products

**Access**

Cataloguing other Objects associated with data and data products

# Making our schema available

http://www.ga.gov.au/data-pubs/datastandards/cataloguestandard

# http://pid.geoscience.gov.au/

Applying geoscience to Australia's most important challenges

**Australian Government**
**Geoscience Australia**

GA website

# pid.geoscience.gov.au home

## About PIDs

This is the root of the persistent ID (PID) URI subdomain managed by Geoscience Australia on behalf of the Australian geoscience community.

This PID subdomain is used to both identify and provide access to a range of digital resources published by members of the Australian geoscience community such as datasets, vocabularies, terms within vocabularies, schemas, lists of people, systems, etc.

The choice of this domain, geoscience.gov.au, for PID use, rather than perhaps GA's web domain, ga.gov.au, is in accordance with 'Cool URI' principles, in particular that PID URIs shouldn't include organsational details that are liable to change with organisational change. Since this domain, geoscience.gov.au, is concept-based, it's likely to remain relevant to Australia for all time (or at least for as long as Australia exists!).

The choice to use a dedicated subdomain, pid, within geoscience.gov.au it to avoid any collisions of web addresses with other users of the geosceince.gov.au domain name, such as the AusGIN website at www.geoscience.gov.au. This

## PIDs in use here

The following collections of PIDs are in use within pid.geoscience.gov.au:

- **Definitional**
  - /def/ -- a register of all definitional resource registers
    - /def/ont/ga/ -- ontology register
    - /def/schema/ga/ -- schema register
    - /def/voc/ga/ -- vocabulary register

- **Instances**
  - /dataset/ga/ -- GA's Dataset register
  - /feature/ga/ -- GA's Feature register. This is conceptual only; there is no service delivering a list of individual features yet
  - /org/ga/ -- GA's Organisation register
  - /sample/ -- a national Samples register using IGSNs for the Sample IDs

# Making our schema available

## ISO19115-profile

### Description

This repository contains the files used to describe and generate Geoscience Australia's profile of the ISO19115-1:2014 'Geographic information -- Metadata' standard. This is the development area for the GA profile - the production release is deployed to GA's Schemas Register at http://pid.geoscience.gov.au/def/schema/ga/ISO19115-3-2016.

This profile extends the base ISO19115-1:2014 standard in accordance with Annex C "Metadata extensions and profiles" of the ISO19115-1:2014 'Geographic information -- Metadata' documentation. Our profile is used to ensure that some elements of the standard which are normally optional are compulsory so that those elements, such as dataset lineage, are always filled out by GA staff recording metadata for datasets. Our profile also implements several new type codes for things such as AssociationType which allow us to characterise relationships between objects described with this profile in ways not possible with the un-profiled standard. These include type codes such as 'derivedFrom' which indicated that a dataset is generally derived from another dataset. This is more general than the standard's own 'revisionOf' which implies that the revised dataset is substantially the same as the original.

Some elements of this profile allow us to relate elements modelled here to other models. For example, the 'derivedFrom' relationship mentioned above can be interpreted as a 'wasDerivedFrom' relationship as used in the PROV ontology for provenance.

### Profile Extensions

Following are the ISO 19115-1:2014 elements that have been extended by the GA profile.

Metadata entity set information (MD_Metadata)

| Name | Definition | Change from ISO 19115-1 |
|---|---|---|
| metadataIdentifier | unique identifier for this metadata record | Optional -> Mandatory |
| parentMetadata | identification of the parent metadata record | Conditional -> Conditional (changed condition) |
| referenceSystemInfo | description of the spatial and temporal reference systems used in the resource | Optional -> Conditional |
| metadataConstraints | restrictions on the access and use of metadata | Optional -> Mandatory |

# Data Strategy - Data Management inventory prioritisation
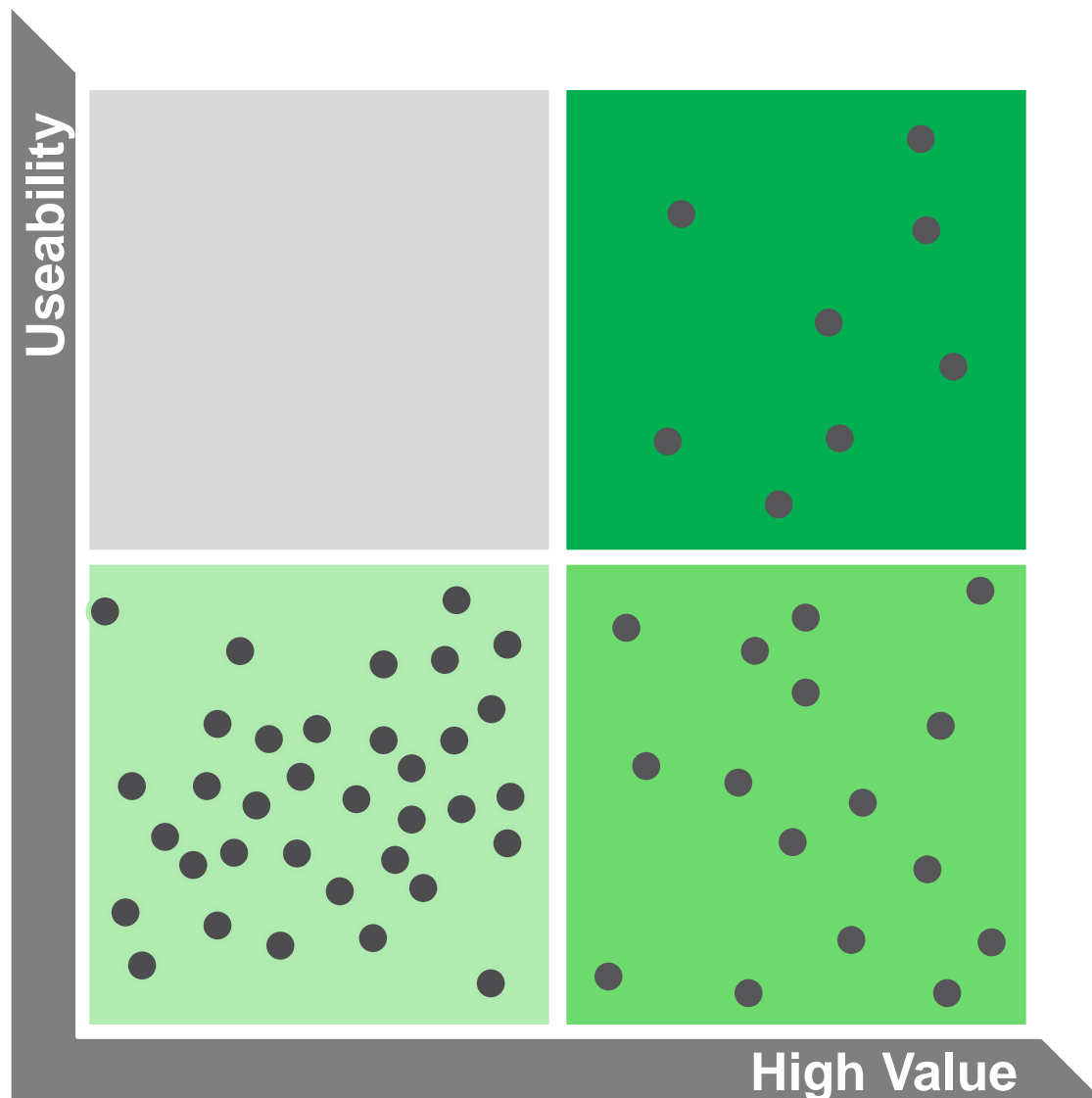
# Thank you.
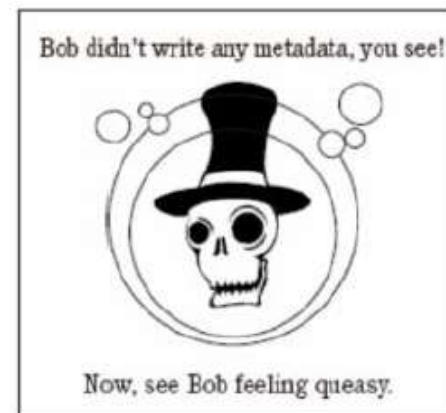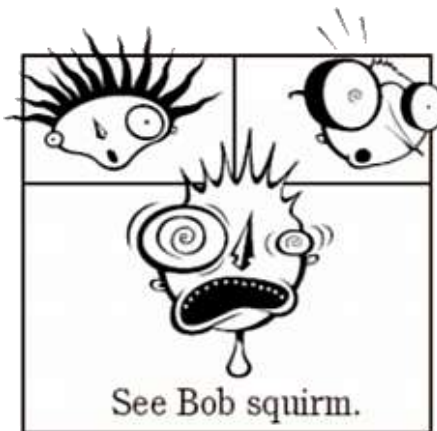
data@ga.gov.au

*With permission: NOAA 2006*