# GSQ Dataset Profile

Guidance document

By Nicholas Car
Senior Experimental Scientist
CSIRO Land & Water, Dutton Park, Qld.
nicholas.car@csiro.au | https://orcid.org/0000-0002-1963-3508

This document is a part of the declaration of the Geological Survey of Queensland's *dataset* profile [1]. It is a *guidance* document which means it is one of many parts of the profile and it intended to assist users of the profile with using it.

## The dataset model

Figure 1 shows the dataset model diagram.

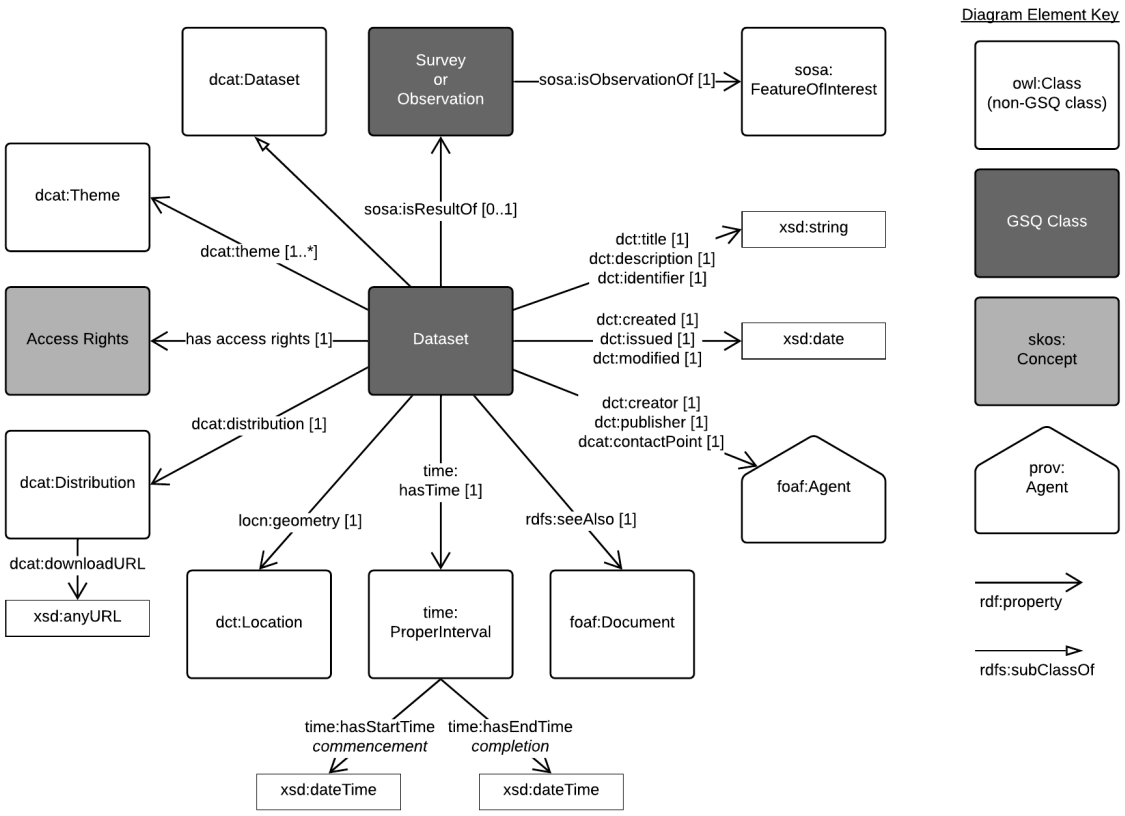


**Figure 1**: The GSQ Dataset Profile's model

## When and how should I use this profile?

You should use this profile when you want to represent basic information about a *datasets* (see discussion about what a *borehole* is below) such as what is it, who made it and when. Likely this information will be used within a catalogue of all of GSQ's datasets and for exchange with other agencies.

## What are all the parts of this profile?

See the profile's index page [1]. It lists the various guidance, validation and specification resources that together make up this profile.

## What is a dataset?

A *dataset* is a management unit of data.

There can be no size, data category, file type or other definition of a dataset as one person's file within a dataset is another's whole dataset. Datasets are best declared as needed to manage all of GSQ's data.

## What granularity of dataset should I create?

Initially GSQ should aim for total coverage of all of its data within datasets. This might mean a few, very large, datasets are created such as "all the old data on network drive X" in order to have it present in the dataset catalogue at all. Over time though, finer-grained datasets should be declared as GSQ gets a better hold on its data.

Data coming into or going out of GSQ in recognisable chunks should be distinct dataset.

Substantially altered forms of data due to processing should be declared as different datasets with relations about derivation (*wasDerivedFrom*) declared between them.

## How does a dataset relate to other GSQ information objects?

Every bit of data in GSQ should be within a dataset. Data from within a dataset may be used within a collection or a custom database, for example, notifications of rock samples analysed by mining companies might come in to GSQ in reports that recorded individually as datasets but then the sample and analysis metadata might then be copied into sample and analyte databases for aggregated use.

## References

[1] Geological Survey of Queensland (2019) *Dataset Profile*. A profile of the DCAT (rev) dataset metadata vocabulary. https://github.com/geological-survey-of-queensland/gsq-dataset-profile