

Issue 1: Individual units vs "unit types"

ABS Problem Statement

The ABS (Australian Bureau of Statistics) is currently developing a Metadata Registry and Repository (MRR) aligned with GSIM. While aligned with GSIM, the definition of objects for the ABS MRR will require attributes additional to those which are defined in GSIM as a generic model.

A question has arisen in this regard when it comes to the modelling of *Unit* in the GSIM V1.0 specification.

The three examples given all relate to individual people, companies etc. Being able to identify a specific unit in this way is important when, eg, populating a register/frame or the unit identifier field in a unit record file.

It is very common also, however, to talk about "types" of statistical units (eg persons, families, enterprises, registrations, transactions)

While related, the concept of an individual unit and the concept of a type of unit don't seem equivalent. It seems unlikely, therefore, that "persons" is simply a *Unit* in the same way as an individual person.

The alternative of saying "persons" is a *Population* also seems awkward. A *Population* will usually have at least some form of temporal and spatial scope. In theory "persons" as a unit type refers to all persons ever born, or yet to be born in the future. This seems different in concept to the population, eg, of persons on Planet Earth (which would typically only include people who were alive at a particular time).

A couple of possibilities come to mind for capturing "unit type" as an attribute. The attribute might be, eg, populated from an extensible controlled vocabulary of recognised unit types (at an agency specific or international level).

Option 1

The attribute could be applied to *Unit*. This might (or might not) imply the list of unit types should be limited to mutually exclusive "base types". (At the conceptual level, are Dan Gillman as a Person and as an Employee referring to two different *Units*?)

Option 2

Perhaps it would be more appropriate to attach the attribute to *Population* to describe the "type" of *Units* (which would be subject to more detailed scoping in the definition of the *Population*) from which the *Population* is composed. For example US persons and Australian persons are

- different *Populations*
- composed of different individual *Units*
- built from the same "type of unit"

Option 3

Unit Type could be added as an object in its own right. This would allow *Unit Type* to have attributes and relationships of its own - including with other *Unit Types*. For example, it would be possible to define "Employee" as a specialisation of the "Person" *Unit Type*. (If "Legal Entity" were ever a *Unit Type* it might be a specialisation applicable to either "Person" or "Corporation"?)

I think that adding a new object tends to add more complexity to the model than adding a new attribute to an existing object. Option 3 would need to add significant advantages over Option 1 or Option 2 in order to justify the extra complexity?

Would you recommend other options be considered instead, or would you recommend one of the options listed above instead of the others?