

Survey Design and Implementation Team Report: Weighting Subgroup 2012/13

Submitted by Peter Granda (Chair)

Team Members

- Steven McEachern, Australian Data Archive
- Dan Gillman, Bureau of Labor Statistics
- Peter Granda, Inter-university Consortium for Political and Social Research (ICPSR), Chair
- Amanda Norton, Australian Bureau of Statistics
- Anita Rocha, University of Washington
- Wendy Thomas, University of Minnesota (DDI Technical Group Representative)

The goal of this subgroup was to build on the existing set of descriptors about weighting that currently exist in the DDI 3.1 and what is contemplated from 3.2.

The group decided to explore a model-based approach and includes a short list of slides to describe in general terms the model and its definitions, elements, attributes, and relationships in addition to some tentative controlled vocabularies. The Group took this approach because:

- 1) Considerations about weighting may enter into the documentation process during several stages of the survey life cycle. For example, design weights can be formulated early in the cycle based on the sample design while non-response weights are estimated later in the cycle upon completion of data collection.
- 2) A belief that weighting is best described in a modular way as a reusable scheme that can fit into different DDI schemas such as Data Collection and Logical Product as the need arises.
- 3) Future DDI development may hinge on a model-driven system and this set of recommendations is an experiment in this space.

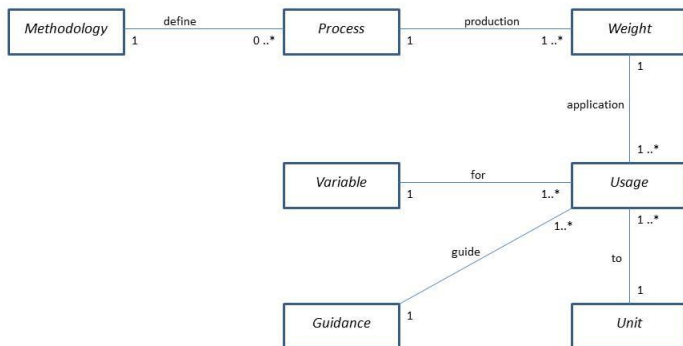
Introduction

- Model for Description of Weights
- General – for any kind of weight
- Model
 - Definitions
 - Class diagram
 - Relationships
 - Attributes

Definitions

- Unit – Sampling unit for which the weight was developed
- Methodology – General description of the method(s) used to develop weights
- Usage – Rules and guidelines for using weights with Units
- Process – Steps from an algorithm applied to generate weights
- Weight – A number assigned to each Unit indicating the number of times that Unit is expected to occur in a population

Model – Class Diagram



Relationships

- **Define**
 - A Methodology defines zero or more Processes
 - A Process is defined by one and only one Methodology
- **Production**
 - A Process produces one or more Weights
 - A Weight is produced by one and only one Process
- **Application**
 - A Weight is applied by one or more Usages
 - A Usage provides the application for one and only one Weight

Relationships

- **Guide**
 - A Guidance guides zero or more Usages
 - A Usage is guided by one and only one Guidance
- **For**
 - A Usage is applied for one and only one Variable
 - A Variable (for which) has one or more Usages
- **To**
 - A Usage is applied to one and only one Unit
 - A Unit (to which) has one or more Usages

Attributes

- **Assumption**
 - Every class has an ID attribute
- **Variable**
 - Defined elsewhere in DDI
 - Those attributes apply
- **Unit**
 - Defined elsewhere in DDI
 - Those attributes apply

Attributes

- **Methodology**
 - Type
 - Set of pre-defined kinds of weighting, such as cross-sectional or longitudinal
 - Reference
 - Reference (e.g., URL) or citation to methodological source
- **Usage**
 - No attributes defined

Attributes

- **Process**
 - Algorithm
 - Description of algorithm applied
 - Literal or reference
 - Intention
 - For what purpose is an adjustment being applied (non-response, population control, etc)
 - Rationale
 - Reason for applying the chosen algorithm to achieve the purpose

Attributes

- Weight
 - Number
 - value for the weight
- Guidance
 - Examples
 - Relevant illustrations
 - Restrictions
 - Rules for appropriate use
 - Recommendation
 - Best practice

Limitations

- Process
 - Process in this weighting model is strictly about the implementation of a methodology for weighting.
 - It does not apply to the survey life-cycle
- Therefore
 - Survey life-cycle process considerations are outside the scope of the model presented.

Methodology - Type

CV Elements:

- Cross-sectional weight
- Longitudinal weight
- Pooled data weight
- Replicate weight
- ... (others?)

Comments:

- Equivalent to TimeMethod?

Intention/purpose (Attribute of "Process")

- Non-response
- Population control/adjustment
- Design effects
 - Oversampling
- Non-coverage
- *Unequal selection probabilities (covered in Methodology??)*
- ... (Other?)