Describing Subsamples in Assessment Data

Many types of assessment ratings require multiple values to be recorded per plot, and is beneficial when there is a lot of variability present. These values are entered in ARM as subsamples.

Describing the assessment's subsamples involve a handful of assessment header fields, depending on the subsampling technique. In GDMdef, these header fields include:

- Sample Size, Unit
- Collection Basis, Unit
- Reporting Basis, Unit
- Number of Subsamples

Depending on the technique, not all of these fields are necessary to describe an assessment. We will provide some examples of different assessments and how to describe them using these header fields.

Example 1: No subsamples

We wish to count the total number of fruit for each experimental unit (perhaps a single tree). This is the simplest case - when no subsampling occurs.

The Sample Size field could be filled as "1 PLOT" or "1 TREE" in this case, but leaving the field empty also implies that there was no subsampling. Thus only the Number of Subsamples field needs to be filled in, with the number 1:

Number of Subsamples	1

Example 2: Basic subsampling

We wish to rate the % disease on the surface of fruit. Instead of inspecting only 1 fruit from the entire plot, we sample 20 fruit in each plot.

In this case, the 20 fruit is our sample size, giving us 20 subsample values to enter for each plot:

Sample Size, Unit	20	FRUIT
Number of Subsamples	20	

Example 3: Adding detail on subsampling techniques

In the previous example, no explanation was given about *how* the 20 fruit were sampled. Perhaps we just randomly chose fruit, trying our best to sample from different parts of the tree as we made our way through the assessment.

On the other hand, maybe we were systematic about how we chose the fruit to rate. Let's say we chose 4 different branches of the tree, and then randomly chose 5 fruit from each branch. Now our sample size is 5 fruit, collected from 4 different branches - our collection basis. This still results in 5*4=20 subsample values to enter per plot.

Sample Size, Unit	5	FRUIT
Collection Basis, Unit	4	BRANCH
Number of Subsamples	20	

Another way to read this is "We are sampling 5 fruit per branch, and doing that 4 times". Or generically:

We are sampling _[Sample Size, Unit]_ per _[Collection basis unit]_ , and doing that _[Collection basis]_ times.

Example 4: Using Reporting Basis to normalize data

Sometimes, sampling must be done at a non-standard basis. Or perhaps there is not a de-facto standard, and the sampling depends on the equipment available at that location. The Reporting Basis can be used when normalizing the data as it is entered into ARM.

For example, suppose the number of nematodes per cm3 of soil is to be recorded. At trial site A, 250 cm3 soil samples are collected, while at site B 200 cm3 soil samples are collected. Each trialist can fill in their own sample size (200 or 250 cm3), but then normalize the data to 500 cm3 when entering the data. Thus trialist A would *record* values 2 times the raw *collected* values (250*2=500), and trialist B would record values 2.5 times their raw collected values (200*2.5=500).

Trialist A:

Sample Size, Unit	250	cm3	∨
Collection Basis, Unit	1	PLOT	∨
Reporting Basis, Unit	500	cm3	∨
Number of Subsamples	1		

Trialist B:

Sample Size, Unit	200	cm3	∨
Collection Basis, Unit	1	PLOT	∨
Reporting Basis, Unit	500	cm3	∨
Number of Subsamples	1		