

Plot	sub-plot	Tmnt	Sp	Ind	wt
1	A	X	Aus	1	10
1	A	C	Bus	1	20
1	B	X	Aus	3	10
1	B	C	Bus	4	10
2	A	X	Aus	1	20
2	A	C	Bus	4	10
2	B	X	Aus	5	20
2	B	C	Bus	4	10

Table 1: A dataset with more complex information

How to annotate the above dataset using “key yes”, “distinct yes” and “identifying yes”.

Given the dataset in Table 1, users have different situations to catch.

- Requirement 1: *Plot* with label “1” should refer to the same one physical plot (i.e., the *Plot* in the first 4 row means the same thing); similarly, *plot* with label “2” should refer to the second physical plot (i.e., the *Plot* in the last 4 row means the same thing).
 - This can be captured in annotation by putting “Distinct yes” for observation type *Plot* and “key yes” for its measurement type *PlotLabel*.
- Requirement 2: *sub-plots* with the same label should refer to the same physical sub-plot if they are within the same plot; but the sub-plot with the same label with different *Plot* label are different sub-plots. E.g., Row 1 {*Plot*=1, *sub-plot*=A} refers the same sub-plot as that in Row 2, but is different from the one in Row 5 {*Plot*=2, *sub-plot*=A}.
 - This can be captured by putting “Distinct yes” for observation type *SubPlot*, “key yes” for its measurement type *SubPlotLabel*. We need to denote *Plot* is its context and with *identifying yes* specified on this context.
- Requirement 3: *Tmnt* with the same label should refer to the same treatment process (So that we can aggregate on different treatment process, e.g., on “X” or on “C”.) But the treatment at different sub-plot should refer to different treatment.
 - The first requirement can be captured by treating all the Treatment with value “X” as the different entity instances with the same type (*Tmnt-Type*). The second requirement can be captured by treating the treatments in different sub-plots as observations of the different entity instances. I.e., treatments in row 1 and row 3 are of but are different observation instances which are of different entity instances.
 - *Tmnt* has the context *sub-plot*. We don’t need to specify “key yes”, “distinct yes”, or “identifying yes”.

Summarization questions:

- (1) Give me the average weight of the individuals with treatment “X”. How can this question be answered after the annotation and materialization? This need to be answered after we annotate *Sp*, *Ind*, and *wt*.

- Requirement 4: *Sp* with the same name should refer to the same species (e.g., a bird named *Aus* flies from sub-plot (1,A) to (1,B).) But the *Sp* with the same name at different sub-plot should refer to different observations of a specie.
- Requirement 5: *Ind* with the same label and and the same species name should refer to the same species. But the individual (with the same label and the same species name) at different sub-plot should refer to different species observations.

The annotation can be done as follows.

- Observation type: Plot *distinct yes*
 - Measurement type: PlotLabel *key yes*
- Observation type: SubPlot *distinct yes*
 - Measurement type: SubPlotLabel *key yes*
 - Context: Plot *identifying yes*. This means that the key measurement for SubPlot (SubPlotLabel) and the key measurement for Plot (PlotLabel) together form the key of SubPlot observation (and its related entity too.)
- Observation type: Treatment *distinct yes*
 - EntityType: TreatmentProcedure
 - Measurement type: TmntType *key yes*
 - Context: SubPlot *Identifying yes*
- Observation type: SpeciesIndividual
 - Measurement type: Species, *key yes*
 - Measurement type: IndLabel (represent *Individual*), *key yes*
 - Measurement type: Weight
 - Context: SubPlot. Note: (1) if the users want to treat the individual (with the same lavel and the same species) at different sub-plots as different observation instances, here, *identifying yes* should be specified. Otherwise, (2) if the users want to treat the individual (with the same lavel and the same species) at different sub-plots as SAME observation instances, here, we don't need to specify *identifying yes*.