

EasyDMP and RDA DMP Common Standard

Hanne Moa (Uninett / EasyDMP), Adil Hasan (Uninett Sigma2), Joakim Philipson (Stockholm University)

EasyDMP is a tool for creating DMPS that has simple and nested question types, in a linear structure. Meanwhile, the DCS has simple and nested data types, in a tree structure. The main goal of the Datatypist team during the hackathon was to encode some of the missing data types as question types, and ponder how to encode the nested tree structures, like dataset. During the hackathon, we included support for the cost complex data type and the “Date” simple data type used in `dataset:distribution:license:start_date`, `dataset:issued` and `dataset:distribution:available_until`. Furthermore, there’s now a rudimentary plan for how to encode datasets.

Simple data types: Date

Implementing Date was quite easy and may serve as a good example of how to implement future simple question types. For that reason we plan to document how to write a simple question type using Date as an example.

Complex data types: *cost*

A cost in RDA consists of a currency code, a description, a title and a value. According to the DCS, a DMP can contain zero or more costs. Fortunately, all of the sub-items have simple data types. EasyDMP already has a way to encode such multiples of simple sets, even though it is a much more involved process than making a simple question type. Better documentation for how to write such a type is needed though.

Complex data types: *dataset*

EasyDMP has no notion of datasets. In the DCS, there can be multiple datasets, each being a combination of some simple data types and some nested tree structures, such as `dataset:distribution`.

Existing design

Each Plan has a single, linear template, as shown in Figure 1. The template contains the questions, and the plan contains the answers.

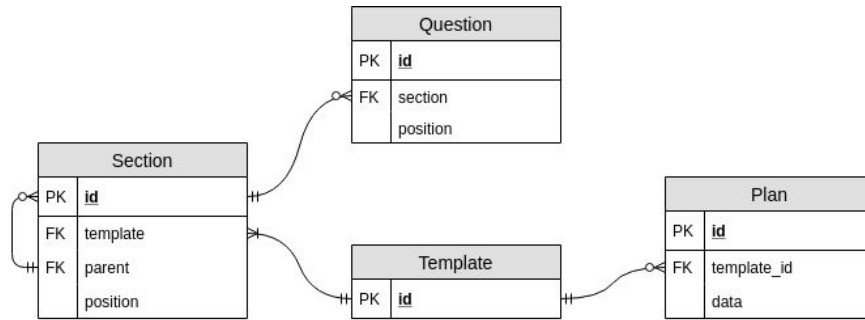


Figure 1. Existing design template

Revised design

A section, which already may have a parent section, is given a role which maps to an RDA tree structure, allowing for a mapping. A Plan is connected to its Sections via AnswerSets. The template still holds the questions, but the answers move from the Plan to the AnswerSets. This redesign is shown in Figure 2. .

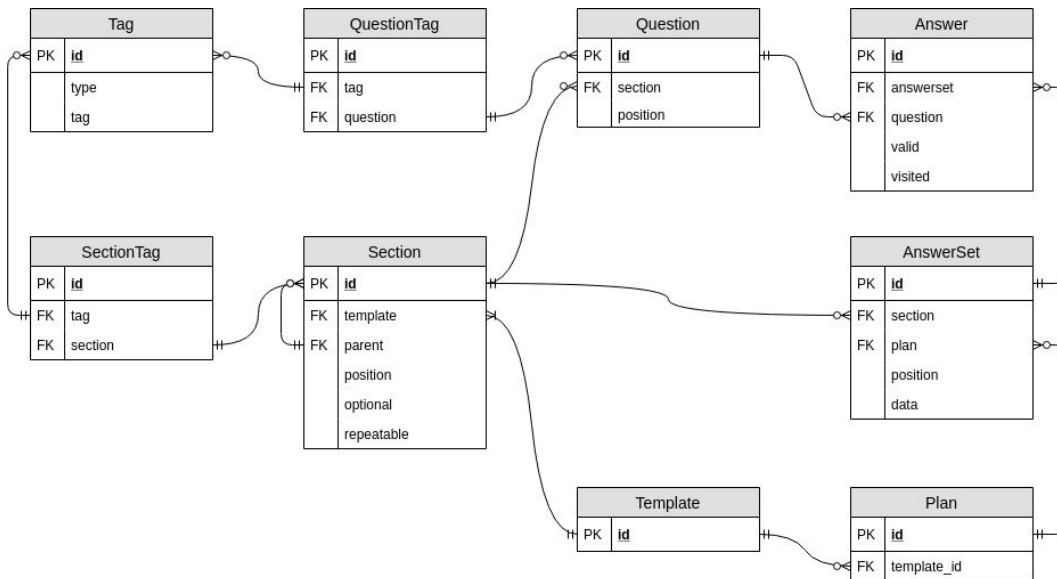


Figure 2. Revised design

Another puzzle will be how to mark cardinality on the sub-sections, since some of them, such as dataset, can occur more than once.

Getting to here from there will be quite a bit of work since much of the business logic on Plan will have to move to AnswerSet, and additional logic is needed to serve it all up as a single endpoint in the API.

Skipped

We had planned to also make a question type for the simple type `DateTime` and something to cover both `dataset:distribution:byte_size` and tools that ask for wanted quota and supplies used quota.

It turns out `DateTime` is underspecified for time zone and how to handle that is uncertain at this point. Tackling this issue would require that a plan worked for collaborators in different time zones; meaning a creation timestamp in one zone could appear to happen after its modification timestamp. We will await clarification on this point before we finish making the question type.

Having a simple type that includes both the unit, as defined at the International System of Units (SI), and magnitude was defined as out of scope since `dataset:distribution:byte_size` can be covered by the existing `PositiveInteger` question type.

Further work

We plan to make a template in EasyDMP that covers all of the DCS application profile, both for testing purposes and to show funders what is covered by the standard, so that the funders can improve their own templates. This will require properly implementing and documenting a tree support via nested templates. When the RDA template is finished we can use that as a starting point to work on importing plans from other systems using the DCS as the interchange format.