ICH DMP-SS

DMP Model and Application Overview

Jack Gager, Pascal Heus
northamerica@metadatatechnology.com
http://www.metadatatechnology.com / northamerica@metadatatechnology.com

DRAFT
Version 2012.03
(This internal project document is in progress and subject to change at anytime)

Approach

After reviewing the DMP Online materials and template (from https://dmponline.dcc.ac.uk/documents), it became clear that although there are some overlaps with the DDI, many extensions would be necessary to meet all of the requirements for a data management plan solution in a purely DDI implementation. Therefore, our approach has been to create a Data Management Plan model that can interact with DDI, ISO27001 and other standards where necessary, but is otherwise unrestricted. In order to create a generic solution, other data management plan implementations were researched, in particular the US DMPTool (https://dmp.cdlib.org). Although this work is still ongoing, we feel that enough information has been gathered to understand what the commonalities are across the various data management plans in use today.

The one thing that became clear after the first round of modelling the data management plans is that a generic solution needs to allow for flexibility in how the plan is structured and what content is included. The investigation of data management plans in use today revealed everything from simple plans with very little structure and only a few fields to more complex plans with many levels of nesting and strong typing for some response fields (The UK DMP being one of the coplex one). The existing plans in use today are almost exclusively used for documentation purposes- that is to say the information contained in the plan is not used in a machine actionable manner. In essence, the data management plan was simply structured textual metadata associated with a study or collection of studies (some of which is referenced directly from the study metadata).

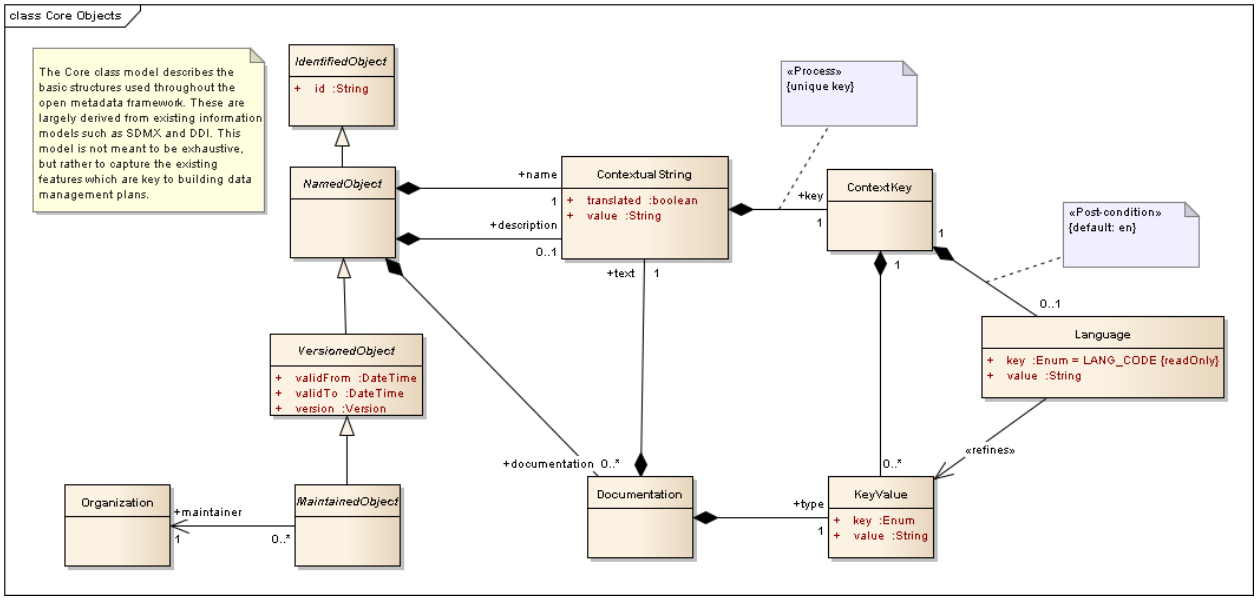
This bore a striking resemblance to the SDMX reference metadata mechanism which allows for a report structure to be described and instance based on the structure to be attached to existing data or metadata. The structure of the data management plan defines the information which is required for a plan (e.g. which fields are required, how they should be populated, and how they should be organized). The instances of the plans provide the actual information for these fields as they related to a particular study or collection of study. With this in mind, a model was created which allows for the structure of data management plans to be defined and for instances of these plans to be created based on these structures.

Model

The model is broken into three basic sections, each described here.

Core Object

This section of the model describes the core objects as they relate to the data management plan tool. These are largely derived from the SDMX and DDI information model and are already in use today in the open metadata framework (the framework on which the DMP solutions will be built).

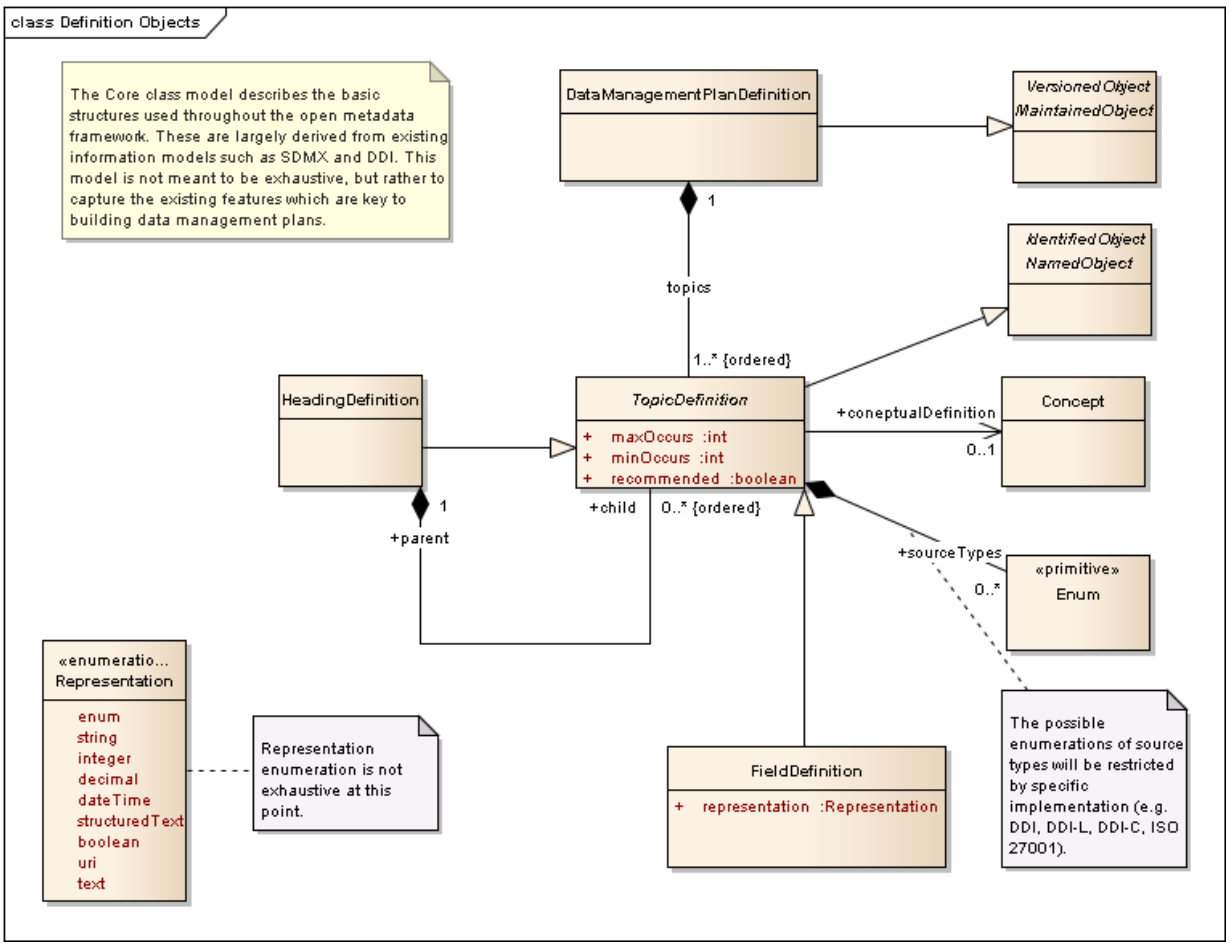


Definition Objects

This section of the model defines the constructs used to structure a data management plan. The abstract structure of a plan definition is actually very simple; it is simply an ordered collection of topic (headers or fields) definitions. A topic definition has cardinality, a flag to indicate that it is recommended (in the case the minOccurs is 0), and an collection of source types. The source types define the type of resources that the information in a report instance can be derived from. For example, one may have a topic definition for a study description. The source types for this might be DDI-L or DDI-C meaning that the value for the topic in a report instance could be mapped from a DDI-Lifecycle or DDI-Codebook resource. A topic definition also allows for an association to a Concept. This association is in place so that various data management plan structures can describe their similarities through conceptual modelling. For example, the concept of “disclosure risk” might be referenced from two different data management plan definitions. Although the nature of the topic in the different plans might be different, they are documenting the same thing. The conceptual relationship allows for this to be inferred, if desired.

A heading definition is a specialization of a topic definition which is a grouping of other heading and/or field definitions. The recursion of the heading definition allows for a hierarchical data management plan structure to be built.

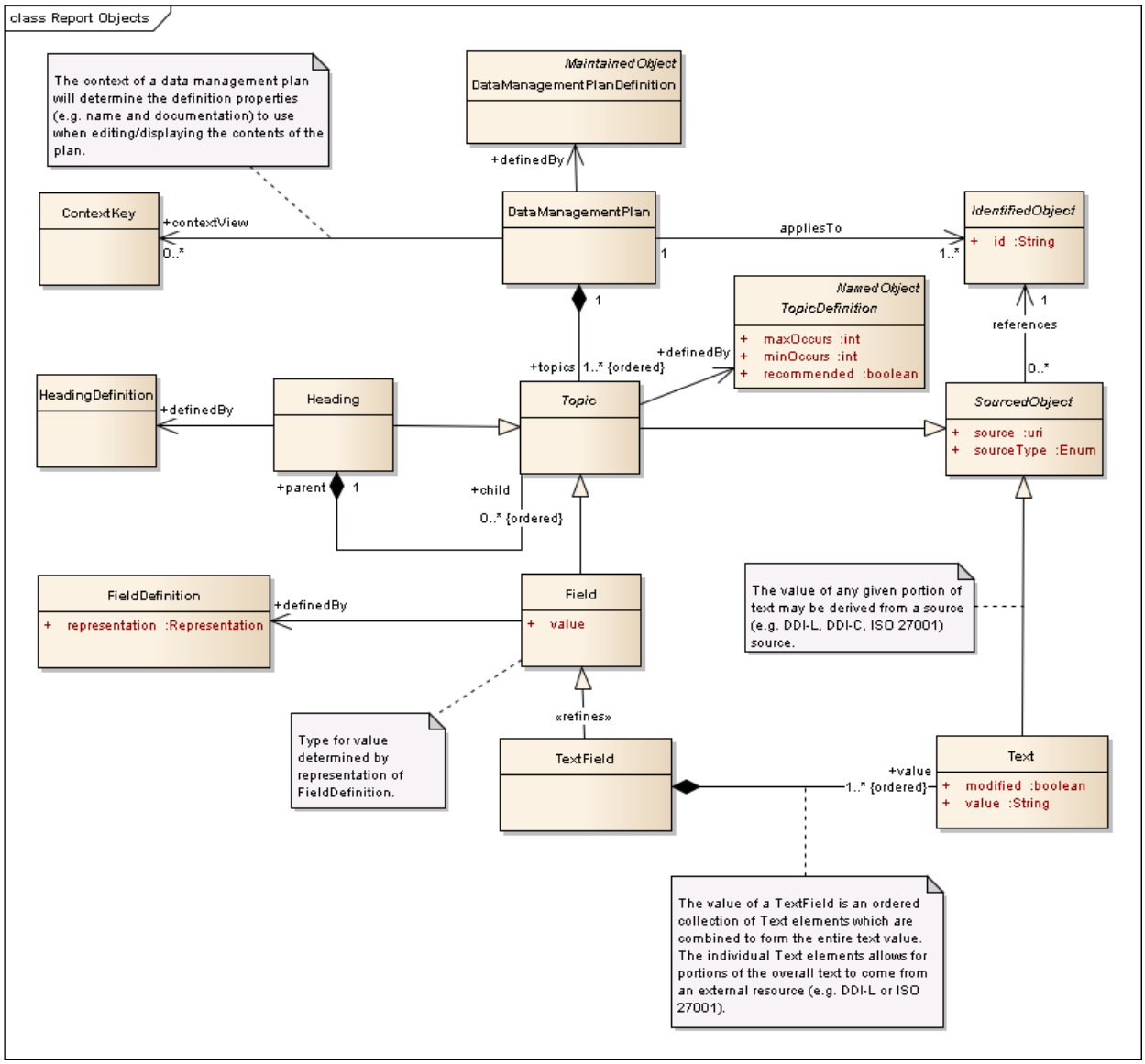
A field definition is another specialization of a topic definition which is meant to hold a literal value. The representation type of the field definition restricts the data type for all value of the field data management plan instances.



Report Objects

The section of the model defines the constructs for actual plan instances. It mirrors the definition structure. A data management plan has an association to one or more identifiable objects. Typically, this will be a study or collection of studies that are defined within the system. A data management can also have an association with a context key. A context key can be used in implementations to determine the specific values from the topic definitions to use. For example, the name of a field might be provided in multiple languages or perhaps the documentation for the field might vary on whether the plan is being developed or disseminated. This relationship to the context can be used to select the appropriate values for the definition metadata.

Every topic in a plan instance can be sourced from an external resource such as a DDI-L instance. The data management plan field has a value associated with it. Although there are many potentially refinements of the field class (based on the representation type of the field definition), only the text field is shown. This is meant to show how the relationship of a field in a report to an external resource (such as DDI-L or ISO 27001) can be further broken down to specific portions of text in a report.



Tools Implementation

Generic Data Management Plan editing tools are being built in the Open Metadata framework around the above data management plan model. These tools will work at the abstract level so that a modification in the data management plan definition is easily translated into an updated version of the data management plan tool. Essentially, the tool user interface will align itself on the plan specific structure.

Based on the information provided from DMP Online, a data management plan definition will be created. The data management plan definition will be used to create editors specific to the ICH data management plan, although the basis of the editor will be built such that they are applicable to any data management plan definition. The general idea is that much like reference metadata in SDMX, an editor can be built on the fly based on a structure definition. The goal is to have a flexible data management editing tool specific to the ICH needs that can also be

readily used to others.

Generic editing components will be built based on the heading and field definitions. The general idea is that the editor will look much like a document template. However, it will be greatly enhanced by the metadata contained in the plan definition. For example, the documentation about a heading or topic will be readily accessible to users so that they can quickly find the instructions on how to populate the plan. And since this will be retrieved directly from the plan definition, the documentation can be very dynamic. Changing the metadata for a topic in the plan definition will be immediately translated to the editors.

What is more, since the editor is being built in the open metadata framework it will already understand DDI. Special editor components will be built for the purpose of populating the data management plan topics from referenced DDI-L resources. These resources will include other material citations of ISO 27001 documents. The manner in which the information is extracted from the referenced resource will be designed in a way that is flexible enough to allow someone with minimal technical expertise to modify how the information is extracted.

For all fields, the editor will have components specific to the representation. For example, if the plan definition defines a field as have a date representation, this will translate to a calendar control in the editor. And as with the documentation relating to the fields, this will be based on the definition metadata so it is dynamic and can be modified by simply changing the plan definition. For text fields, the editors will allow for rich text editing where appropriate as well as allowing for specific components of the text to be extracted from external resources.

Based on the generic data management plan model, an XML serialization will be created for use in web services. The generic nature of this XML will allow tools to be written that can essentially process any data management plan, regardless of the underlying definition. For example, by combining the plan instance with a definition, an XSLT can be written which can generate PDF or HTML renderings of a plan without knowing the specific structure of the plan.