EBA European Banking Authority

European System of Financial Supervision

ITS on supervisory reporting

Data Point Model database

May 2012

Introduction

When developing the draft Data Point Model (DPM) for the draft Implementing Technical Standards (ITS), put forward in consultation papers CP 50 and CP 51, the EBA decided to further enhance the methodological approach, by introducing a relational database as the repository for the DPM metadata, instead of relying solely on MS Excel data structures. For convenience reasons, MS Access was chosen to support this database.

One of the main purposes of this new technical component is to strengthen the coherence of the model, by imposing some logical constraints through referential integrity, and to enable the realisation of a series of automatic consistency checks that were not available before. We believe this is contributing decisively to reduce the effort needed to achieve the desired level of quality, on a DPM that categorises nearly 22,000 table cells, and nearly as many data points.

Another major benefit from the database is the possibility of defining many different views on the same metadata content, in an easy and secure way, to facilitate the interpretation of the DPM and the understanding of the reporting framework. The dimensional data points are now explicitly defined in the DPM, and the link to/from the business templates is direct, requiring no interpretation.

The database structure is reflecting the DPM modelling concepts, not actually the business concepts of COREP/FINREP, which are only represented by the metadata content. By having a generic meta-model structure, the database can be used in any reporting domain, beyond COREP/FINREP, and not limited to supervisory frameworks.The meta-model focuses directly the main concepts of data point modelling (e.g. framework, table, table cell, dimension, member, domain …), keeping a low level of abstraction, and therefore should be easy to query.

The dimensional concepts used in the DPM are the same ones used in analytical systems, which makes possible a very straightforward connection between both ends of the reporting chain.

The meta-model is not bound to any particular technical solution for data exchange, and no XBRL specific constraints are reflected in the DPM, which is technology neutral as required. However, in order to streamline the process of automatic translation from the DPM to XBRL taxonomies, an extra layer of XBRL properties (e.g. namespaces), will be added to the database.

Other enhancements will be introduced on the next version, to deal with problems like the versioning of metadata. Both the templates and the data points' categorisation are expected to change in the future, and keeping track of history of the unique data points is a fundamental requirement for data warehousing and time series analysis.

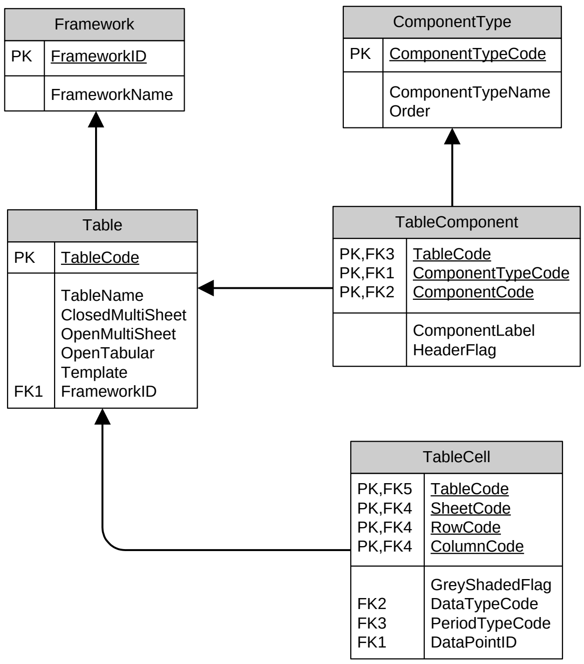
For now, at this introductory stage, the meta-model is also ina draft version, showing only the core concepts, kept in its most basic format.

Supporting the core DPM process

The meta-model is basically structured around three groups of concepts: (1)the representation of the templates' metadata, (2)the dimensional concepts used to categorise the data, and (3)the links between them, which is the actual categorisation.

Tables belong to a Framework (currently either COREP or FINREP); most of the times the concept of *table* will be the same as business *template*, except when, for modelling reasons, the templates had to be normalised and split into two or more tables (e.g. CA 5.8).

(In the following relational diagrams, the arrows represent *many-to-one*relationships,pointingfrom the foreign key to the primary key).

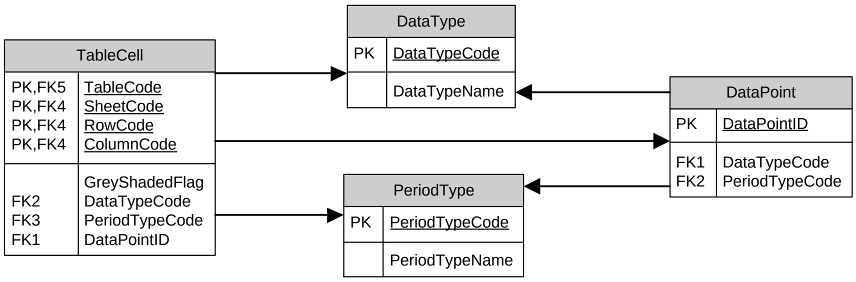


A TableComponent is either, a *row*, a *column*, a *sheet*, or the *header* of a table, and these are the only four instances of ComponentType.

This decomposition of tables is required by the modelling process, which categorises each individual component (3,866 in total), instead of each individual cell (22,000 data cells).

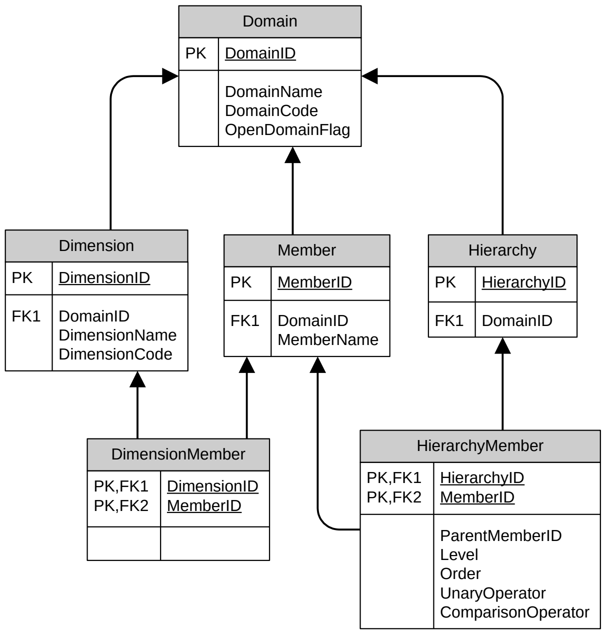
There are in the framework different kinds of tables. Most have a fixed structure, with one single sheet, while others can have multiple sheets with the same structure (e.g. CR IRB 1), or even a variable number of sheets (e.g. CR IRB GB). Also, some tables have a 'tabular' format, that is, an open structure where rows are identified by typed key data, and repeating an indeterminate number of times, depending on the data being reported (e.g. MKR IM 2).

TableCells are generated in the database, by crossing the components of each table. Around 50% of all COREP/FINREP table cells are grey-shaded, either because data is not being required, or because the *row x column* combination has no logical meaning. Later in the modelling process, data type (e.g. *monetary, percentage, integer*…) and period type (*stock* or *flow*) are automatically assigned to table cells.



Each table cell (not considering the grey-shaded) corresponds to one and one only DataPoint; there are,however, some data points represented in different table cells. In the later case, the table cells are sharing exactly the same categorisation in the DPM.

The dimensional concepts represented in the second group are Domains, Dimensions, Members, and Hierarchies.



Dimensions are the different categories used to describe the data points (e.g. *Counterparty*), and Members are the actual instances of those categories(e.g. *Central banks*).

For instance, the cell in FINREP table 5, row 100, column 030, is categorised in the DPM by the following five pairs *[dimension].[member]:*

[Base].[Liabilities]

[Amount type].[Carrying amount]

[Main category].[Deposits. Redeemable at notice]

[Accounting portfolio].[Trading financial liabilities]

[Counterparty].[Central banks]

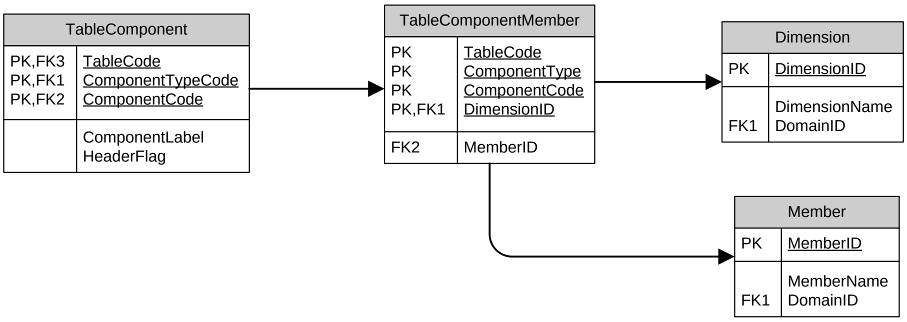
All members of a dimension must belong to the same Domain. A domain groups together members of the same type, corresponding to concepts with a similar semantic nature, either abstract like *Type of risk*, or more concrete like *Currency*. Some domains are closed, i.e. have a predefined and restricted number of members (e.g. *Countries*), and others are open, since we cannot enumerate all possible instances (e.g. *Legal Entities*).

Dimensions are not always equivalent to domains. Dimensions of the same domain can play different*roles* in the model. For instance *Residence of the counterparty, Location of the activities*, and *Country of the market*, are all different dimensions that take their members from the *Geographical area* domain. Thus the same domain member can belong to different dimensions, and two dimensions from the same domain can categorise the same data point (although a dimension cannot be associated twice to the same data point).

Hierarchies specify how members relate to each other (parent-child relationships), and define the aggregations from lower to upper levels in the hierarchy. At this stage, the hierarchies are not yet included in the database, but will be part of the final DPM.

When analysing the templates, the business experts define the set of dimensions and members that categorise each column and row. In case some common categories apply to the whole table, they can then be assigned to the header component.

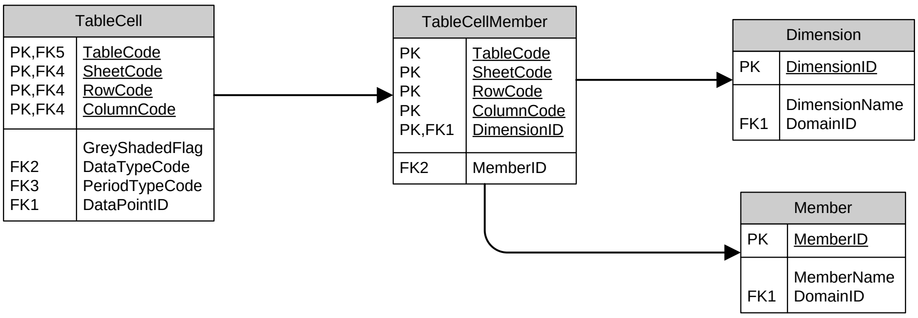
There are also templates (e.g. CR IRB,FT 10) that requiredata input at the header level,(e.g. *Exposure class*, *Residence of counterparty*), in which casethe same table structure is being replicated in different sheets.The specific dimension and members that differentiate each sheetmust also be specified in the DPM, and will apply to all cells of the sheet.



Supporting the enhanced DPM process

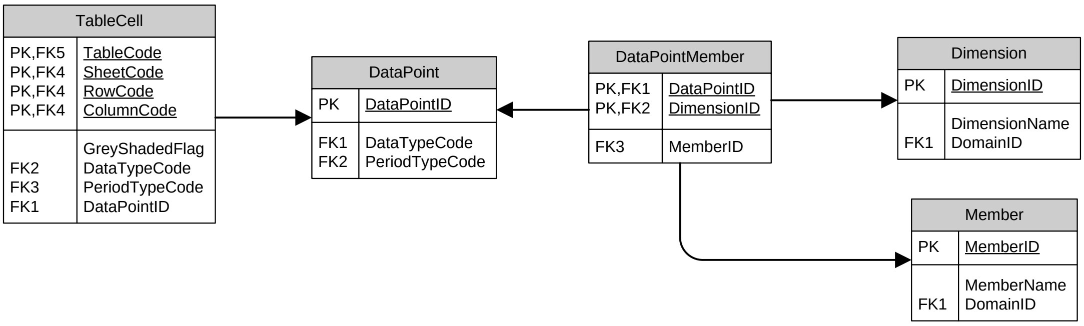
Data point modelling is an iterative process. Each cyclebegins withtemplate analysis & modelling, continues with loading& generating metadata, and ends with running checks& analysing the errors.

Generation of metadata consists of defining the categorisation of each individual table cell, based on the categorisation of its components (header, sheet, row, and column).



Consistency checksare appliedto all the cells in the framework, to validate the model from a logical perspective, checking for cases of missing mandatory dimensions, or duplicate dimensions, for instance; the resulting validation reports are sent to business experts for analysis.

Data points are “discovered”bymeans of identifying the unique combinations of pairs *[dimension].[member]*throughout the complete set of categorised cells.



The outcome of this last automatic process needs also to be analysed by the business experts, as there are two possible reasons for different cells being pointing to the same data point: either the result was expected (or then realised as true), or there is a mistake that needs to be corrected in the model, requiring new analysis of templates.

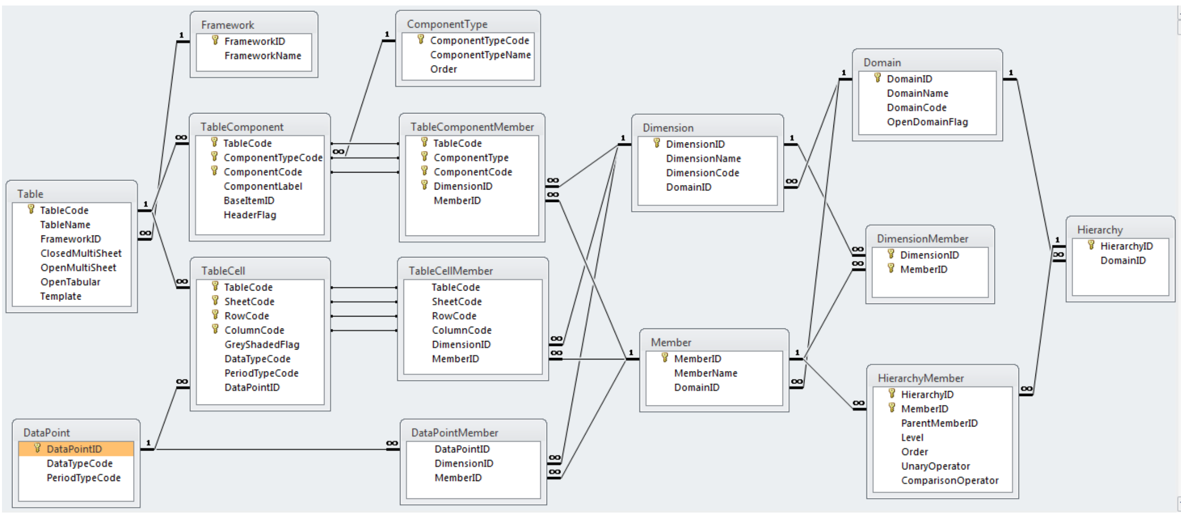
Conversely, there is also the possibility of finding two cells, which are known *a priori* to be the same data point, that are not showing the same categorisation, which also calls for another look at the analysis.

Evolution of the DPM meta-model

The next release of the database will keep the basic structures of the current version, but the meta-model will be extended to support additional features like versioning, validation rules, XBRL properties, and references to the underlying legislation.

The database is published in the EBA Internet web-site with the file name:"CRD4 DPM – Database.accdb"

Under menu 'Database Tools/Relationships' you'll find the database diagram:



General comments and questions on the database design options, or relatedsubjects,that are <u>outside the scope of the DPM consultation</u>, can be addressed to: *carlos.martins@eba.europa.eu*