Annex I: Non-technical introduction to SMCube methodology

SMCube is a methodology for defining metadata to describe datasets. Its pivotal role is in defining cubes, which define the structure of a dataset, intended as a set of data organised as a table with fields (columns) and records (rows).

Table 1 provides an example of a dataset with information related to securities.

Table 1: Securities dataset.

ISIN	Type of instrument	Inception date	Legal final maturity date	Subordinated debt	Nominal value	Fair value	Accrued interest	Currency
YY7365 923618	F_32	16/06/2015	16/06/2035	1	1.000	1.020	10	EUR
ZZ3941 829354	F_511	05/10/2015			6.000	6.050		USD

The minimum information required to describe the structure of the dataset can be summarised with the following three questions:

1. What are the fields (columns) of the dataset?

In the SMCube methodology, the fields of a cube are called variables. The variables are defined independently of the cube, so that a cube will refer to as many variables as it needs; but the same variable may be referred by many cubes, thus allowing for reusability of concepts.

Referring to Table 1, the variables are: “ISIN”, “Type of instrument”, “Inception date”, “Legal final maturity date”, “Subordinated debt”, “Nominal value”, “Fair value”, “Accrued interest” and “Currency”.

2. What are the allowed values for each field?

In the SMCube methodology, each variable is defined on a domain, which states the allowed values for the variables. Domains can be enumerated, if they provide a list with the allowed values, or non-enumerated, if they provide a data type.

In the example dataset, some variables are defined on non-enumerated domains, like the “ISIN” (string domain), the “Inception date” (date domain) or the “Fair value” (monetary domain). Three variables are defined on enumerated domains: “Type of security”, “Subordinated debt” and “Currency”; for each of these variables, it is necessary to provide the list of allowed values. The possible values for an enumerated domain are called members in the SMCube methodology. For each member, the dictionary shall provide additional information, like the description of the concept (e.g. what does F_32 mean?)

But one variable can have different allowed values in different datasets. For instance; the example dataset refers to securities, so the allowed values for “Type of instrument” should only refer to securities, but in a dataset covering more instruments the list of allowed values could be enlarged. In order to allow reusing the variables, the domain will contain all the possible members for variables.

The same holds true for non-enumerated domains, where the subdomain will serve to constrain the allowed values, even if they are not enumerated. For instance, the variable “Fair value” is a monetary variable, but in the context of the securities dataset, its allowed values may be limited to only positive amounts.

In the context of a cube, one variable has to be associated to a concrete subdomain, i.e. the subset of the domain that is allowed in the context of the cube.

3. What is the role of one field within one dataset?

One of the most relevant aspects of the structure of the dataset is what the identifier of the record is or, in other words, what combination of fields makes a record unique. In the example dataset, if nothing is said regarding the structure, applying some business knowledge one may conclude that each record is uniquely identified by the ISIN. But it is also true that one security with one ISIN may be denominated in more than one currency, so if the variable currency is referred to the denomination of the security, it could be the case that more than one record per ISIN is allowed, as long as the currency in each record is different. Thus in order to get a thorough understanding of the described dataset, the role of the variables has to be explicit. In the SMCube methodology, variables that serve as identifiers of the records take the dimension role.

Let’s suppose that in the example the only dimension is the ISIN. The rest of variables may have also different roles, depending on the variables to which they add information. Let’s take again the currency variable as an example, once it is known that it is not a dimension. Without additional information, a user could have two different interpretations of it: It may be adding information to the ISIN (i.e. to the combination of dimensions, in this case only one) specifying the currency in which the currency is nominated, or it may be referred to the monetary amounts, specifying the currency in which the amounts are nominated. In the SMcube methodology, the variables that add information to the combination of dimensions take the observation value role, while the variables that add information only to one variable of the dataset take the attribute role.

It is worth highlighting that one variable can take different roles in different datasets. For instance, the example dataset is disaggregated to the level of security. Another dataset may contain, for one reporting institution, the total securities broken down by instrument.

Table 2: Aggregated dataset.

Reporting institution	Type of instrument	Nominal value	Fair value
ABC	F_32	1.000	1.020
ABC	F_511	6.000	6.050

In the context of this dataset, the type of instrument is a dimension, since it is part of the identifier (one record is uniquely identified by the combination of “Reporting institution” and “Type of instrument”).

Summary

With the SMCube methodology, one cube serves to define the structure of a dataset. One cube is a set of variables, for which the allowed values are specified by a subdomain, and that have a role in the context of the cube. The SMcube definition of the dataset 1 could be summarised in the following table:

Table 3: Cube describing the securites dataset.

Variable	Subdomain	Role
ISIN	12-character alpha-numerical code	Dimension
Type of instrument	Types of instruments for dataset 1	Observation value
Inception date	All dates	Observation value
Legal final maturity date	All dates	Observation value
Subordinated debt	Boolean including not applicable	Observation value
Nominal value	Positive monetary amounts	Observation value
Fair value	Non-negative monetary amounts	Observation value
Accrued interest	Non-negative monetary amounts	Observation value
Currency	ISO 4217 codes	Observation value