DESIGN OF DATA STRUCTURE DEFINITIONS FOR MICRODATA

Report of Experiences by the European Central Bank and Deutsche Bundesbank

Contents

1 Introduction

In the SDMX[1] Roadmap 2020 (1) the SDMX sponsors have outlined their vision of where SDMX should head in the coming years. One of the main objectives of the roadmap is to make the use of SDMX easier and more widespread. So far, the usual application for the SDMX standard has been the time series data exchange. Recently, there is an increasing focus on microdata. To give a first starting point, this report of experiences of using SDMX for microdata has been elaborated. It complements the already existing "Guidelines for the Design of Data Structure Definitions" where the focus was on the exchange of macrodata (2).

This report is written from the perspectives and practices of the European Central Bank and the Deutsche Bundesbank with the House of Microdata (3) and AnaCredit. The House of Microdata is the central integrated microdata collection at Deutsche Bundesbank. AnaCredit stands for "Analytical Credit Datasets" and is a data set containing detailed information on individual bank loans in the euro area. Therewith, the report focusses on financial and real economic data in central banking. Survey data are not covered.

The aim of the document is to flag the issues and challenges which are typical for microdata and give the reader an orientation how to proceed with microdata when designing a Data Structure Definition (DSD) using the current version 2.1 of the SDMX standard.

Prerequisites

The reader of this document is assumed to have a basic knowledge of the SDMX standard. For more information the reader is referred to the existing documentation, e.g. the SDMX website[2] or the SDMX InfoSpace from Eurostat[3]. Specifically, the documents on the Content Oriented Guidelines (4) and the Guidelines for the Design of Data Structure Definitions (2) are recommended as a prerequisite.

Terminology

Before going into details regarding the construction of a DSD for data exchange it is important to frame the definition for macro- and microdata in the context of official statistics.

The United Nations Statistical Commission and the Economic Commission for Europe define

- *Statistical Microdata* as "An observation data collected on an individual object - statistical unit" and
- *Statistical Macrodata* as "An observation data gained by a purposeful aggregation of statistical microdata conforming to statistical methodology" (5).

The G20 Data Gaps Initiative elaborated in their Workshop on Data Sharing 2017 a terminology focussing on the confidentiality of the data (6). They define

[1] SDMX stands for Statistical Data and Metadata Exchange
[2] www.sdmx.org
[3] www.ec.europa.eu/eurostat/web/sdmx-infospace

- *aggregated data* as “data aggregates that have a low likelihood of identification of individual reporting units, such as those found in traditional data sets (e.g., those covered by most of the DGI-2 recommendations)” which is equivalent to the above definition of macrodata,
- *disaggregated data* as “data below the level of aggregated data and with a higher likelihood of identifying individual reporting units than in the aggregated data”,
- *microdata* as “data on individual reporting units or specific transactions/instruments, which in most cases allow the identification of individual entities and [are] therefore considered confidential. In addition, publicly available data on individual reporting units are considered non-confidential although they can still be subject to data sharing limitations due to commercial property rights” and
- *granular data* as both “disaggregated data and micro data”.

The G20 Data Gaps’ terminology complements the definition by the United Nations Statistical Commission and the Economic Commission for Europe by also defining the terms *granular* and *disaggregated* data. Furthermore, they complement the definition of micro- and macrodata by elaborating on the confidentiality of the data. In this document we will not focus on the confidentiality of the data although it is one import aspect which will be briefly touched in the next chapter. The terms *macrodata* and *aggregated data* will be used synonymously throughout this document.

In the central banking context, within microdata one can further distinguish between microdata at the level of reporting agents (e.g. Monthly Balance Sheet Statistics) and transactional data (e.g. Money Market Statistical Reporting). The two types of microdata differ in the time dimension: Microdata at the level of reporting agents normally take a snapshot of all the variables / indicators at a moment in time and have a pre-established frequency (just as macrodata). Transactional data is recorded whenever the transaction occurs. Even if the data has a fixed reporting frequency (e.g. “daily” in the case of the Money Market Statistical Reporting), the timestamp of the transaction is the one that is included in the data.

Structure of the document

The document is structured as follows: in Chapter 2 the challenges of microdata compared to macrodata are reviewed. Chapter 3 describes the process for DSD definition for microdata according to the experiences of the European Central Bank and Deutsche Bundesbank. Chapter 4 touches on the topic of easy-to-use SDMX formats. The given theoretical explanations are then completed by the use cases of the House of Microdata and AnaCredit.

2 Challenges of Microdata

After the financial crisis microdata became more and more important in central banking. The financial crisis had shown that it was no longer sufficient to look at aggregates but also the heterogeneity of data had to be taken into account. Furthermore, distributions of the data as well as complex economic relationships have to be analysed. In financial stability for example, in order to assess systemic risk looking at individual financial institutions is crucial. These examples show that microdata are imperative for central banks to ensure performing their core tasks of maintaining price stability and contributing to financial stability. (7) (8)

In the following section the challenges which accompany the shift from macro- to microdata are outlined: starting from general challenges of microdata and going over to challenges specific to the DSD's design.

2.1 General Challenges

The general challenges for microdata compared to macrodata are comprised of several microdata characteristics concentrated around data volume, masterdata, metadata, revision practises and validation.

Volume of the Data

Usually, the volume of microdata is several orders of magnitude higher than that of aggregated data. This is due to the fact that microdata include data on individual reporting units or specific transactions and instruments. Most of the time, microdata is high frequency information (events, tick-data, raw-data).

The commonly used SDMX message format (SDMX-ML) is XML-based and in consequence, a high volume of data may result in large files for the data exchange. Therefore, effort has to be put into how to cope with the data volume, for example by using compact exchange formats (see Chapter 4).

Data Confidentiality

In the case of microdata, data is considered at the level of individual units instead of looking at their aggregates. This often leads to higher requirements regarding the protection of the data.

The SDMX standard itself guarantees neither data integrity nor confidentiality. However, outside the standard there are well known mechanisms allowing such issues to be tackled[4].

In order to manage the access to the data, SDMX allows for flagging confidential data. In the SDMX standard it is possible to define different confidentiality types[5] which can be set at observation level, at the level of time series and for the data set. Setting confidentiality types at the attribute level is not possible. These confidentiality types are flags and do not guarantee the protection of the data. The compliance with the confidentiality types has to be ensured by the interpreting tools and processes.

Master Data

With microdata we have observations at the level of individual entities instead of looking at groups such as banking groups or whole sectors. Hence, data describing these individual entities becomes more relevant in particular for the analysis of the data. This describing data is called *master data*[6]

[4] For example, within the context of message integrity, the exchanged XML-file can be validated against the format's XSD (XML-schema). Additionally, checksum mechanisms could be put in place. Concerning the confidentiality as well as integrity of the data, one could leverage the HTTPS protocol to ensure that the communication link is encrypted and that the source is the one expected.

[5] The "Guidelines for Confidentiality and Embargo in SDMX" (11) cover the confidentiality aspects in SDMX data exchange, including embargo scenarios.

[6] According to the DAMA Dictionary, *master data* is "the data that provides the context for business activity data in the form of common and abstract concepts that relate to the activity. It includes the details (definitions and identifiers) of internal and external objects involved in business transactions, such as customers, products,

and the challenge is how to link the master data to the data itself. One possible modelling solution is to attach the master data directly to each data point. In this case the master data is linked to the data by construction which allows a high performance when using the data. However, it entails a high redundancy of the data. Another possibility is to create a separate DSD or MSD for the master data which is more compact but then has to be linked to the data outside of the standard.

Reference Metadata

Reference Metadata (such as information on underlying concepts, methodology or quality) are essential for interpreting data correctly. From numbers without metadata we cannot deduce any insights, they are just numbers. Compared to aggregated data, metadata for microdata are especially important. As the data is no longer aggregated, more granular and additional information is necessary in order to interpret the data correctly. Together with the general high volume of microdata there is a need for a standardized documentation of the metadata since it is becoming infeasible to get the relevant information from the responsible data expert on call.

Back Data Revision Mechanisms

There are different revision mechanisms for macro- and microdata.

The compilation methods followed by macrodata producers may have different degrees of reliability and macro estimates are often subject to revisions. Initial aggregate estimates are released with the expectations that these may be revised and updated as further data becomes available. This is due to the fact that with time aggregated data is enriched and revised based on better underlying source data thus aggregated back data[7] is frequently revised. Revisions may also occur when methods or systems are changed (for example change in the aggregation method or change of data source). Typically such back data revisions are not very common in the case of microdata. There do exist revisions/corrections by the reporting agents (e.g. in Money Market Statistical Reporting) but these are with respect to a time period which is close to the present and not with respect to historical data.

Validation

Validation continues to be important for microdata. Although the topic will not be elaborated on, the most recent developments are mentioned: until lately, validation was conducted as part of the data exchange process. SDMX includes a package for transformations and expressions which is present in the information model but until 2015 no specific language existed. In 2015, the Validation and Transformation Language (VTL) was published by the Technical Working Group and it is foreseen to be implemented in the next version of the SDMX standard (SDMX 3.0). More information including the VTL 2.0 package can be found at the SDMX website[8].

---

employees, vendors, and controlled domains (code values)". In the context of central banking and reporting, master data are mainly data on institutions and instruments.

[7] Back data in macroeconomic statistics is the generation of improved longer time series (back in time) based on the current observations and newly available data referring to the past values.

[8] www.sdmx.org

2.2 DSD Specific Challenges

In addition to the general challenges noted above there are issues specific to DSD creation:

Multiple Measures

With microdata comes also the need for multiple measures. The question is how to model several observations for one information item in a specific reporting period. In Money Market Statistical Reporting for example, there is not only the nominal amount of the transaction as an observational value but also the spot rate and forward points are observational values for the market segment of the FX Swaps.

The use of several measures has not been integrated in the current SDMX 2.1 version of the standard but it is scheduled to be considered in the next version. Currently, there are several ways to deal with multiple measures:

One possibility is to add a dimension to the DSD describing the type of the measure and create a new key for each observation. An additional dimension allows for a harmonized treatment and an easy retrieval. The disadvantage is that it increases the number of dimensions and it is less compact. This concept of an additional dimension is also foreseen by the standard. SDMX 2.1 offers a so called "measure dimension" which allows to specify different types of measures. However, only one primary measure can be defined.

A second possibility is to exploit attributes. In general, the use of attributes for multiple measures has the advantage that storage is compact. The first option here is to add an attribute which defines the type of the measure. A second option is to use the main observation as the measure and then integrate the remaining observations as attributes. In this case however, the treatment of the observations is not harmonized and retrieving the value of the attribute and comparing it with the actual observation value is more difficult. Difficulties also occur when one would like to attach different attributes' values for the actual multiple observation values since we cannot attach sub-attributes on attribute level.

Which of the options should be chosen depends on the business case.

Uncoded Concepts and Rapidly Increasing Codelists

For aggregated data, code lists usually have a manageable number of codes; the codes are known in advance and the code lists are static. In the rare event that a new code needs to be added, a new version of the code list is created.

With the substantial volume growth coming along with microdata, the number of codes in a code list may increase radically. Furthermore, the codes – together with their descriptions - that would need to be in a code list may change frequently (e.g. new codes may be added on a daily basis) or they may not be known in advance.

Examples are

- the International Securities Identification Number (ISINs) in Securities Statistics,
- the Universal Transaction Identifier (UTI) of transactions in financial markets,

- the RIAD-Code in the European System of Central Banks (ESCB) and the Legal Entity Identifier (LEI) from the Global Legal Entity Identifier Foundation (GLEIF) for identifying legal entities participating in financial transactions.

In these cases, not all of the possible values that would cover a concept can be anticipated and consequently, the feature of *uncoded concept* becomes more relevant for microdata. Uncoded concepts are part of the SDMX data model and can be either associated to a dimension or an attribute. The drawback of un-coded concepts is that by not defining a code list (i.e. code values and their associated description) non-valid codes may be exchanged and the values exchanged or stored in a database do not have a clear description associated. Therefore, uncoded concepts should be used for concepts that can be interpreted unambiguously by the data compiler or end-user for example.

Regarding, the rapidly increasing code lists, the rising number of codes would lead to an explosion in the number of DSD versions. This challenge will be tackled in the next version SDMX 3.0 of the standard.

Groups

Another aspect of microdata is that the definition of groups becomes more relevant. To work with the data, individual entities need to be grouped, e.g. individual countries into the group of the European Union (EU) or individual institutions into Monetary Financial Institutions and Non-Financial Monetary Institutions. For this requirement, the SDMX standard offers Hierarchical Codelists, where codes can be arranged into simple hierarchies by referencing another code as its parent (10).

3 DSD Design Principles for Microdata

The SDMX standard provides a generic model. Hence, there is no need per se to differentiate between different data types like macro- and microdata for creating the data model.

Therefore, when working with microdata one can follow the same approach for the DSD definition as for macrodata. A short overview of the steps can be found in Figure 1. For a more detailed description the reader is referred to the “Guidelines for the Design of Data Structure Definitions” (2).

For macrodata we have the tendency to group the data and in most cases have fewer DSDs in order to ensure harmonisation and integration of the structures. For microdata, often, several DSDs are created due to the numerous dimensions/concepts characterising each data point.

The advantage of using one single DSD is that the data is already linked by construction. The SDMX Standard 2.1 allows to link different dataflows through the Category Scheme by defining a common category. However, it is not able to establish the variables based on which the data sets should be linked (i.e. it does not allow to define foreign keys in terms of database architecture). Hence, this linking of different data sets has to be defined outside of the SDMX standard.

Several DSDs have the advantage that the model is denser and less sparse. The data is less redundant while the integrity increases (i.e. the model is normalized in terms of database architecture). This is especially favourable with a huge amount of data.

To conclude, there is no golden rule to decide on how many DSDs to use. Instead, for each business case one has to find the balance between redundancy and integrity respectively versa the number of DSDs.

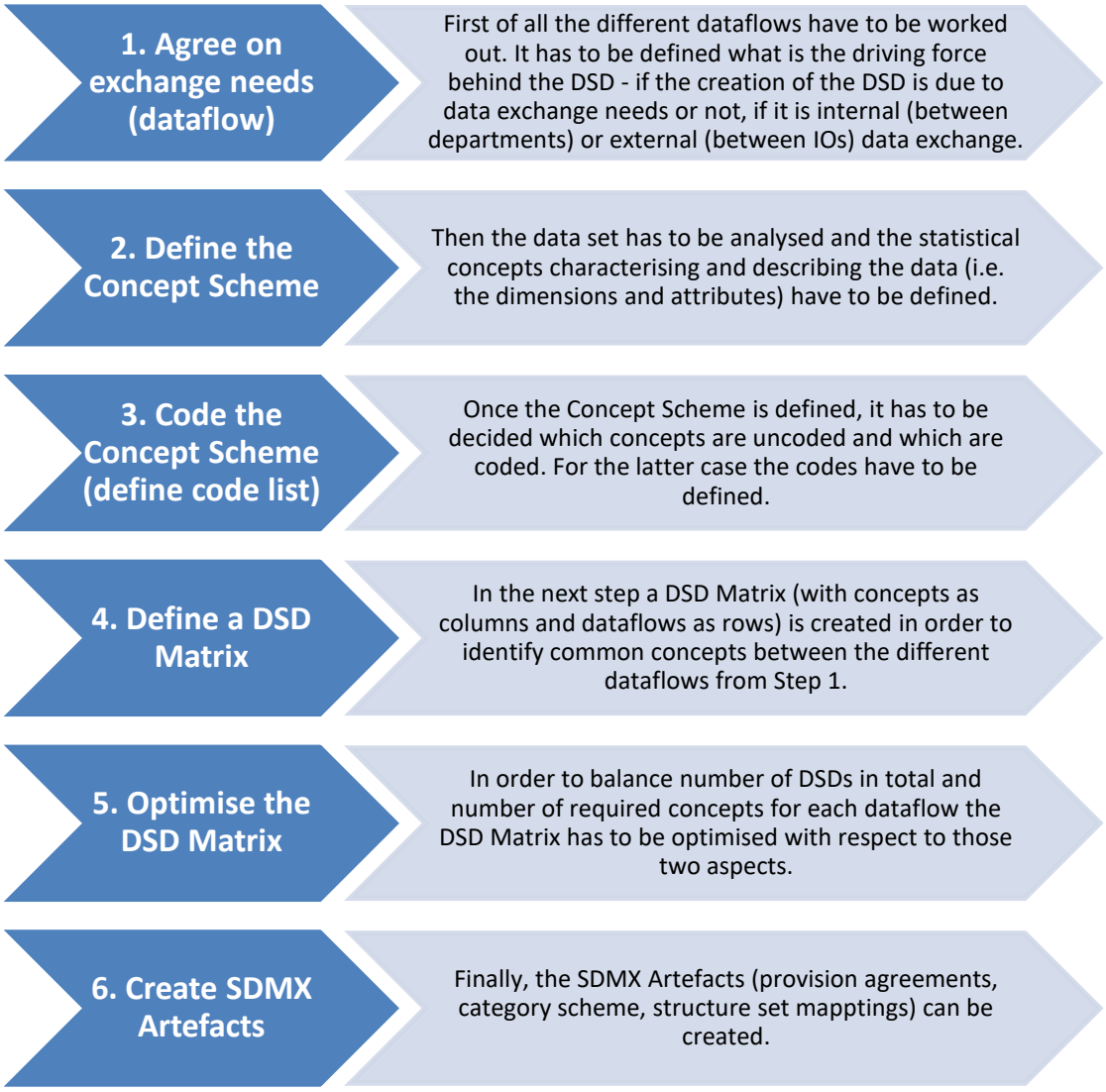

Figure 1 Step-by-step guide for DSD design.

4 Easy-to-Use SDMX Formats

The sizes of microdata sets are several orders of magnitude larger compared to macrodata. Hence, the compactness of the exchange format as well as the user-friendliness play an important role.

The SDMX standard 2.1 offers two XML formats: the generic schema and the structure specific format. The generic format uses XML elements to represent the information while the structure specific format replaces the XML elements with attributes where the attribute's names are derived from the dimensions' codes. Therewith, the generic format has an abundance of structure description (dimensions, attributes, …) which allows structuring the data and which may also be important for the later analysis of the data whereas the structure-specific format allows for a more compact representation of the data.

Furthermore, the SDMX-ML 2.1 offers to use “flat formats”. To write a flat data set, the dimension at observation level must be set to “AllDimensions”. With this approach it is possible to write data sets that do not have any series explicitly defined in the XML. The data set is essentially “flattened” so unlike the standard output, where there is a series containing observations, there are no series, but each observation describes the series key and series attributes. It then does not matter what is written as a series key, and what is written as an observation, it all gets written at the level of the observation. An example is given in Chapter 5.2.

The flattening of the SDMX artefacts and hence the SDMX neutral representation of the data allows to handle the data without knowledge about the SDMX concepts. However, when working without a DSD one withdraw of this approach (and also of the later introduced SDMX-CSV format) is that it does not include information about dimensions, observations and attributes. For getting this information the corresponding DSD is needed.

In addition to the common SDMX-ML 2.1 format based on XML, there exists the SDMX-EDI format. It arose from the GESMES/TS (preceding the SDMX Standard) and uses the UN/EDIFACT syntax. It was the first SDMX format and is restricted to time series but it is still used for reporting financial data. In contrast to the SDMX-ML it is very compact.

Furthermore, there a two new compact and easy-to-use formats. The SDMX Sponsors released an SDMX-CSV format[9] based on the RFC 4180 specification for CSV files with the aim of having a representation which is optimised for both public data dissemination and for usage in common statistical software as well as for creating pivot tables in spreadsheets applications. An example is shown in Chapter 5.2. Furthermore, the SDMX-JSON format[10] was defined which conforms to the JSON standard specification and targets data discovery and visualisation on the web.[11]

Compared to the often used generic format of the SDMX-ML, the above presented formats are more compact and therewith better suited to cope with the high volume of microdata. Each of them was developed for a different application scenario and hence, the best format has to be chosen based on the use case.

5 Use Cases: the House of Microdata and AnaCredit

After having described the theoretical process for constructing a DSD, we present in the following two use cases on how these theoretical thoughts have been realized in practice.

5.1 The House of Microdata

With the increasing importance of microdata, Deutsche Bundesbank has launched a large-scale initiative to make better use of existing microdata both, for fulfilling its duties and responsibilities as well as for internal and external research projects. This initiative is called IMIDIAS - Integrated MIcroData based Information and Analysis System - and aims at creating a central integrated data collection at Deutsche Bundesbank.

[9] https://github.com/sdmx-twg/sdmx-csv
[10] https://github.com/sdmx-twg/sdmx-json
[11] A more detailed description of the mentioned data formats can be found at: https://metadatatechnology.com and https://ec.europa.eu/eurostat/web/sdmx-web-services/data-rep which served as source for this chapter.

The underlying microdata collection is called the House of Microdata. It is based on the existing Central Statistical Infrastructure of Deutsche Bundesbank with its tools for data management and analysis.

The House of Microdata includes data sets from different Directorate Generals. The resulting data variety needs standardization. Therefore, SDMX is used as a common data model for the House of Microdata to standardise and harmonize the data from the different data sources.

For each data set which is integrated into the House of Microdata an SDMX classification is created. In the process we design a DSD as described in Chapter 3 balancing for each case the number of DSDs versus the redundancy and integrity of the data. The common type of data exchange for all data sets is internal since the data is integrated into the House of Microdata from different operative systems within the Deutsche Bundesbank.

In the following we give two examples:

The **Monthly Balance Sheet Statistics** provide an overview of the business of German banks (MFIs). Comparatively, it is a clear example for microdata with a low data volume of around 2 million time series without dynamic dimensions or multiple measures. It is an example, where a DSD for the aggregated data already existed. For the microdata, we created a new DSD with almost the same structure – with the exception that the dimension using the reference sector breakdown (e.g. „Monetary and Financial Institutions“ or „Credit Institutions“) for the aggregated data uses the individual banks (e.g. „Bank 1“ or „Bank 2“) for the microdata.

We created only one DSD for the Monthly Balance Sheet Statistics. However, we created a second DSD for the reference data of the banks containing for example the bank's location, the bank code and the banking group. This DSD can then be reused for other banking statistics like the Statistics of the Bank's Profit and Loss Accounts.

The **Money Market Statistical Reporting** is a transaction based example with a higher volume of microdata. It collects transactions carried out by monetary financial institutions on the euro money market on a daily basis. In this case we decided to use only one DSD and keep the number of the dimensions small in order to cope with possible corrections of transactions. Therefore, most of the information was specified in the attributes. The DSD is given in Figure 2. We see that in this case we handle multiple measures by using the nominal amount of the transaction as the main measure and integrating other observations like the interest rate in the attributes.

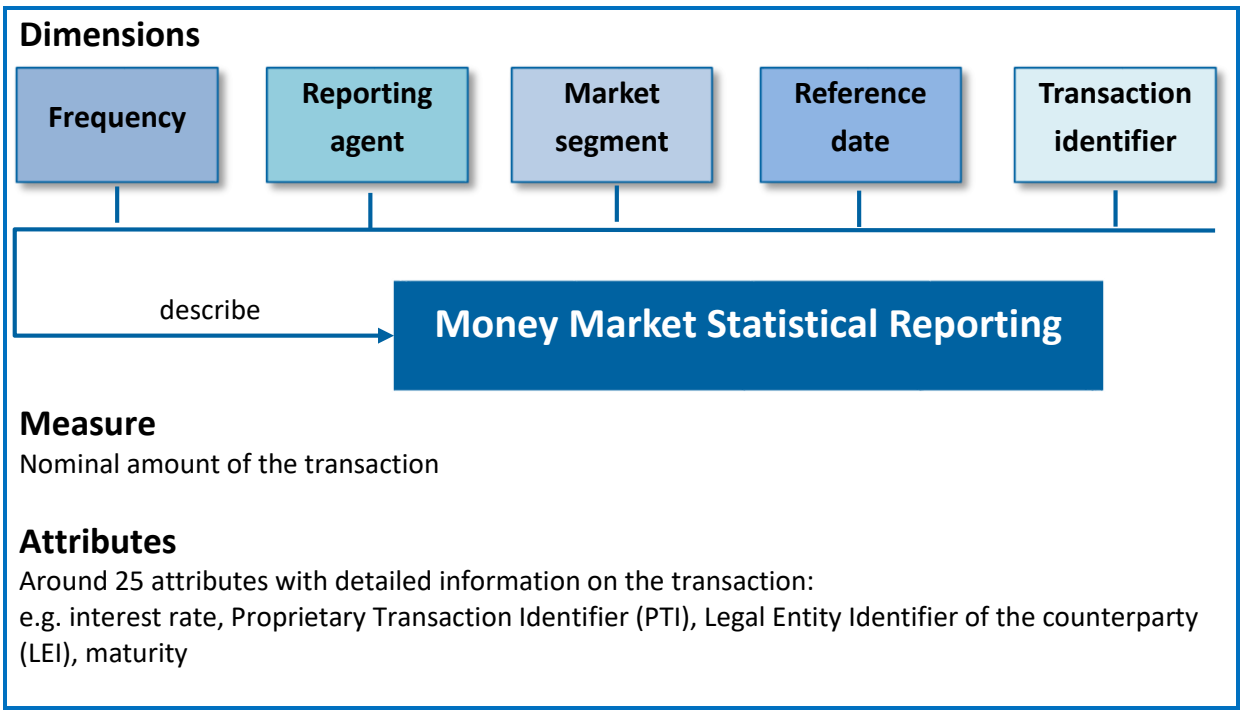

Figure 2 Concepts of the DSD for Money Market Statistical Reporting

As mentioned in Chapter 2 one important aspect in the case of microdata is **data confidentiality**. In the House of Microdata it is ensured by reusing the access right concept which has been already in place on the level of aggregated time series. In addition, a central unit has been created in the Directorate General Statistics which administrates the user access to the data according to the underlying legal and contractual rules.

Another challenge for microdata which we presented is the increasing need for standardized **referential metadata**. For this purpose, a dedicated DSD for metadata has been created both for micro- as well as macrodata. It is based on a metadata model developed in cooperation between the data providing departments, the Research Data and Service Center and the division for Statistical Information Management, Mathematical Methods. The DSD is depicted in Figure 3. We do not use a Metadata Structure Definition (MSD) as foreseen by the SDMX standard since it is not implemented in our infrastructure. Using a DSD for metadata allows us to use the tool set which we developed for data also for metadata (e.g. for data input, maintenance or presentation). We use the same DSD for micro- and macrodata but created two separate Data Set Identifier (DSI) for allowing a strict

separation regarding access rights.

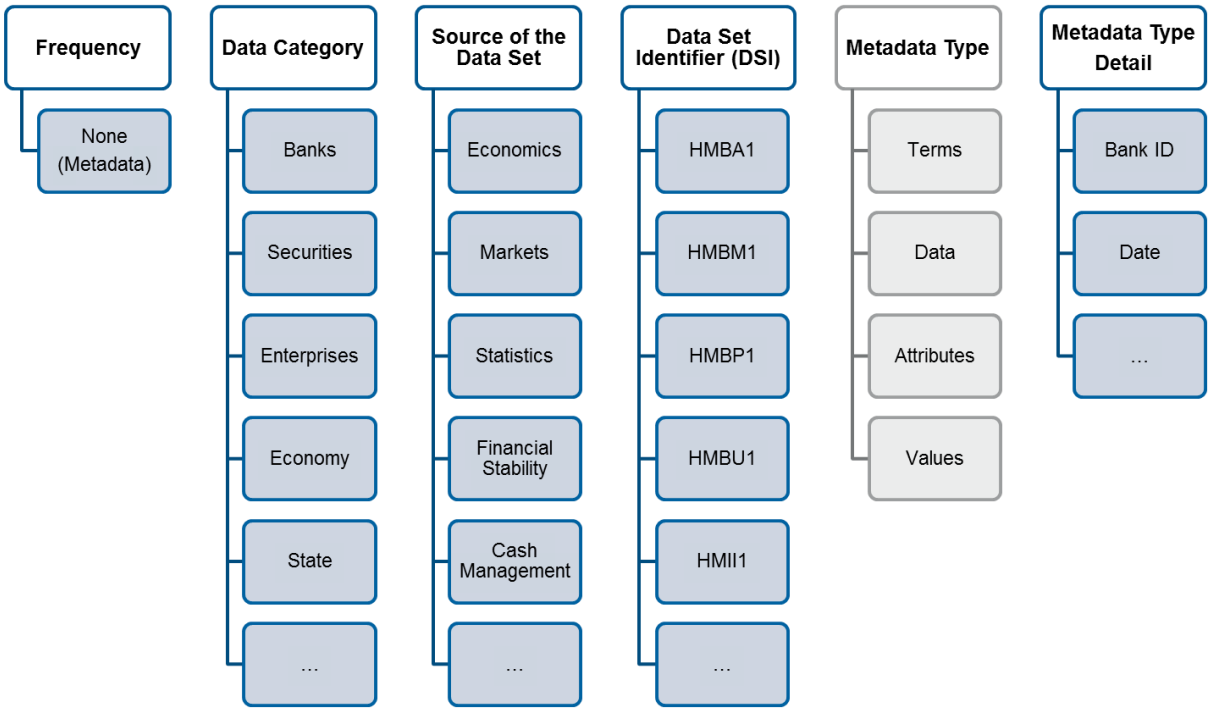

Figure 3 DSD for Reference Metadata

5.2 Bundesbank's Implementation for AnaCredit

In the AnaCredit project, the European Central Bank (ECB) collects microdata on credits on a loan-by-loan basis from National Central Banks (NCBs) of the Eurozone and beyond. NCBs collect these microdata from the monetary financial institutions (MFIs) of their jurisdiction.

The ECB published an SDMX format in flattened form (see Chapter 4) for collecting the microdata from NCBs. Bundesbank's AnaCredit project (AnaCredit-BBk)[12] chose to also use this format for its reporting agents (an example is given in Figure 4). As mentioned in Chapter 4, the flat format can be used without knowledge about the SDMX standard. Due to different timelines and to provide more simplicity for the reporting agents, Bundesbank's AnaCredit project decided to use the flat format without using a corresponding DSD.

[12] More information about AnaCredit Bundesbank can be found at https://www.bundesbank.de/de/service/meldewesen/bankenstatistik/kreditdatenstatistik--anacredit--611424

Figure 4: Exemplary SDMX flat format for data collection for AnaCredit Bundesbank[13]

As a data model AnaCredit Bundesbank decided to use a relational model creating a table for each entity. A draft impression about the complexity of the AnaCredit relational model with its different entities can be gained in Figure 5. The number of key dimensions of the entities is rather low (3 to 6), and the number of values per dimension is high (e.g. the instrument entity comprising the credit instruments issued by the MFIs or the counterparty information for all credit contract counterparties, consisting of a unique counterparty identifier and a vector of about 30 observation attributes).

[13] Version 2.0 of the XML-specification can be found in https://www.bundesbank.de/resource/blob/748914/f12c1dd52d6f454bb236563a89820605/mL/anacredit-technisches-meldeschema-version-2-0-data.zip

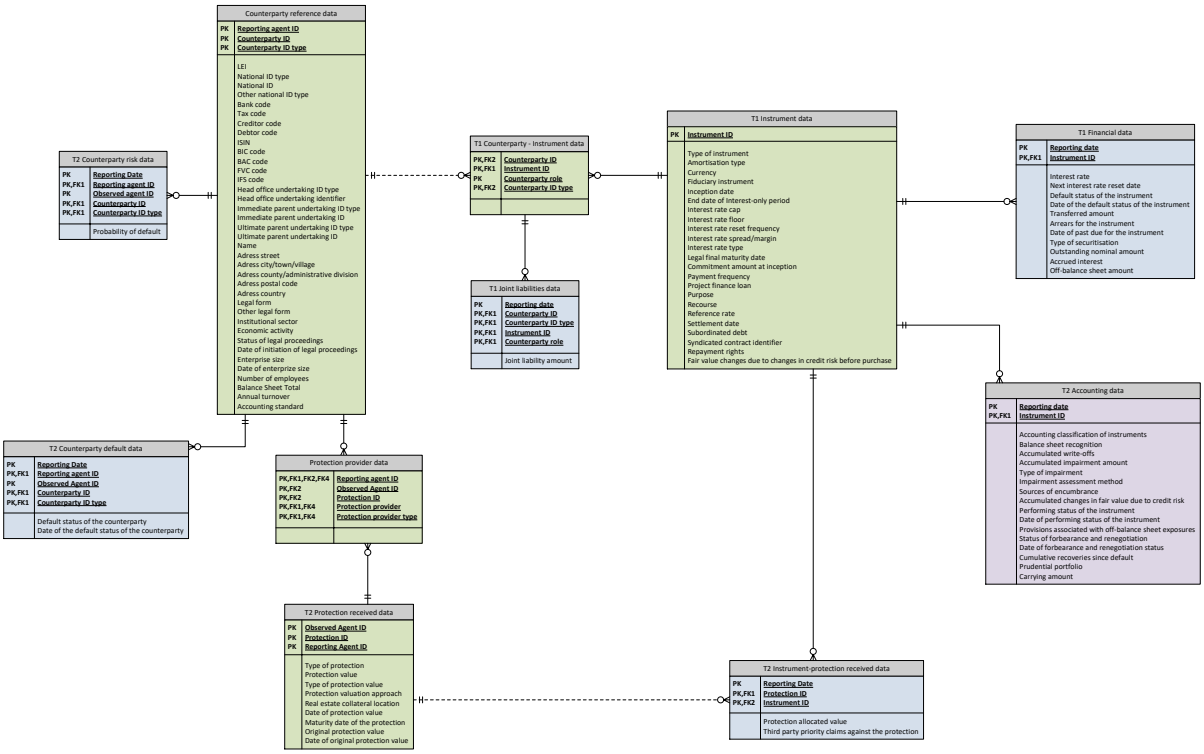

Figure 5: Relational data model of tables reported to AnaCredit Bundesbank

As discussed in Chapter 4 this flat format approach without a DSD allowed the AnaCredit reporting agents to manage their reporting obligations without being forced to handle SDMX concepts. They just had to sort their data into the tables of the relational data model and for each point pack keys and attributes in an observation without any classification.

For the (point-to-point) interface to the internal Business Intelligence system, the SDMX-CSV format as described in Chapter 4 is used. As shown in Figure 6 the easy-to-use CSV is based on the relational model with many dataflows (one for each entity).

```
DATAFLOW;RPRTNG_AGNT_CD;OBSRVD_AGNT_CD;DT_RFRNC;CNTRCT_ID;INSTRMNT_ID;TYP_INSTRMNT;TYP_AMRTSTN;CRRNCY_DN
BBK:BBK_ANCRDT_INSTRMNT_C(2.0);DEICR19;DEICR19;2018-06-30;A910;BE02;20;5;EUR;2;2015-11-23;NOT_APPL;NOT_A
BBK:BBK_ANCRDT_INSTRMNT_C(2.0);DEICR19;DEICR19;2018-06-30;E100;BE01;1000;4;EUR;2;2015-12-06;NOT_APPL;NOT
BBK:BBK_ANCRDT_INSTRMNT_C(2.0);DEICR19;DEICR19;2018-06-30;A500;BE50;1003;4;EUR;2;2015-11-01;NOT_APPL;NOT
BBK:BBK_ANCRDT_INSTRMNT_C(2.0);DEICR19;DEICR19;2018-06-30;A750;BE01;1004;3;EUR;2;2015-06-16;NOT_APPL;NOT
```

Figure 6: SDMX-CSV format for internal communication in AnaCredit Bundesbank

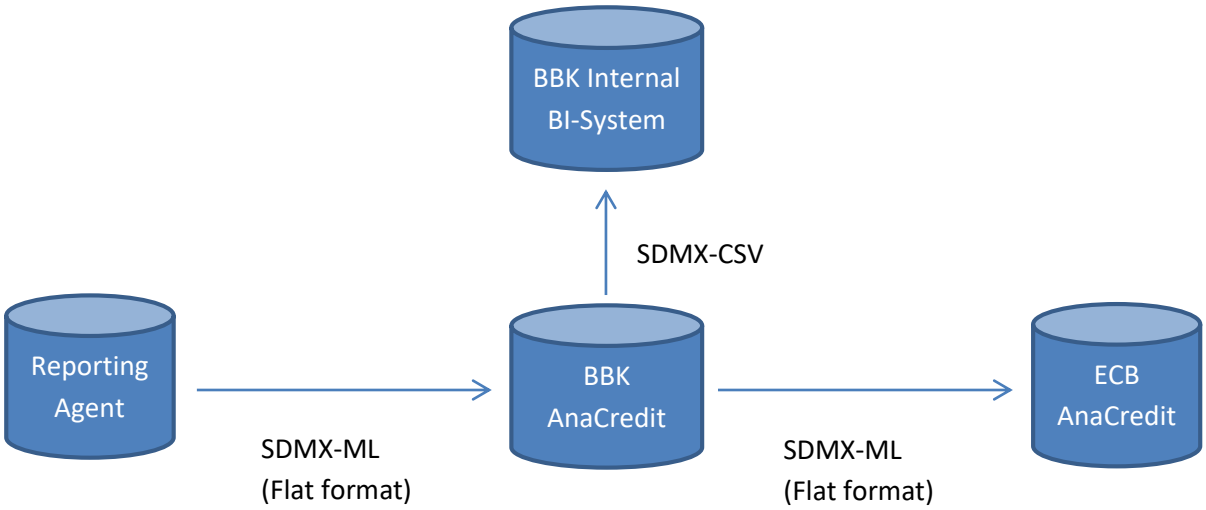

Figure 7 Easy-to-use SDMX formats within the AnaCredit Bundesbank

In sum, Bundesbank's AnaCredit uses SDMX formats for microdata on the input and output side as well as internally (see Figure 8).

6 Conclusion

So far, the focus of the SDMX standard has been on aggregated data. After the financial crisis microdata became more and more important and the need arose to investigate how SDMX can also be used for microdata.

This report reflects the experiences with microdata from the European Central Bank and Deutsche Bundesbank to give a first starting point. To begin with, it reviews challenges of microdata. Some of the challenges can be tackled within the SDMX standard, for others the SDMX standard has to evolve or means outside of the standard have to be found.

With respect to DSD design, due to the genericity of the SDMX standard, one can apply the same process for micro- and macrodata. Compared to macrodata, it is in particular relevant for microdata to balance for each business case the number of DSDs versus data redundancy and integrity.

To conclude, two use cases were presented. Within the House of Microdata we elaborate for two business cases the process for DSD design and how some of the mentioned challenges have been tackled. The second use case of AnaCredit gives an example for employing easy-to-use SDMX formats.

7 Bibliography

1. **SDMX Sponsors.** SDMX Roadmap 2020. [Online] February 2016. [Cited: 19 11 2018.] https://sdmx.org/wp-content/uploads/SDMX_roadmap2020_FINAL.pdf.

2. **SDMX Sponsors**. *Guidelines for the design of data structure definitions.* June 2013. Version 1.0.

3. *The Bundesbank's House of Micro Data: Standardization as a success factor enabling data-sharing for analytical and research purposes.* **Staab, Patricia.** Basel : 8th IFC Conference on "Statistical implications of the new financial landscape", 2016.

4. **Statistical Working Group .** *SDMX Content-Oriented Guidlines.* February 2016.

5. **United Nations Statistical Commission and Economic Commission for Europe.** Terminology on statistical metadata. *Conference of European Statisticians Statistical Standards and Studies, No 53.* Geneva : CONFERENCE OF EUROPEAN STATISTICIANS, 2000.

6. **Inter-Agency Group on Economic and Financial Statistics.** *Update on the Data Gaps Initiative and the Outcome of the Workshop on Data Sharing.* March 2017.

7. **Professor Claudia M. Buch, Deutsche Bundesbank.** Focus on micro data: Potential benefits for the industry? *Speach at the 8th European Central Bank Conference on Statistics. Central Bank Statistics: moving beyond the aggregates.* 2016.

8. **Deutsche Bundesbank.** *Microdata – paradigm shift in central banks' statistics.* Frankfurt am Main : Annual Report, 2015. pp 47-59.

9. **Sponsors, SDMX.** *Modelling Statistical Domains in SDMX.* June 2018. Version 2.0.

10. **SDMX Sponsors.** *SDMX Glossary.* October 2018. Version 2.0.

11. **SDMX Sponsors**. *Guidelines for confidentiality and embargo in SDMX.* January 2018. Version 2.0.

12. **SDMX Sponsors**. *SDMX Implementors Guide.* November 2005. Version 2.0.