



# **Linking data for MFIs**

## Technical Report 2019-01

Deutsche Bundesbank, Research Data and Service Centre (RDSC)

Deutsche Bundesbank, Research Data and Service Centre (RDSC)

# **Linking data for MFIs**

## **Technical Report 2019-01**

Harald Stahl - Deutsche Bundesbank

Citation:

Stahl, Harald (2019), Linking data for MFIs, Technical Report, Deutsche Bundesbank, Research Data and Service Centre (RDSC).

## Contents

<b>1</b>	<b>General approach</b>	<b>3</b>
<b>2</b>	<b>Multi-office banks</b>	<b>3</b>
<b>3</b>	<b>Banks' profit and loss accounts</b>	<b>3</b>
<b>4</b>	<b>Different frequencies</b>	<b>4</b>
4.1	Linking master data . . . . .	4
4.2	Linking yearly GuV and monthly Bista . . . . .	4

## **Abstract**

We describe how to link several research datasets on banks provided by the Bundesbank's Research data and service center. That is generally easily done. However, exceptional with respect to the identifier are the monthly balance sheets statistics for multi-office banks and the banks' profit and loss accounts.

## 1 General approach

Linking data for MFIs is in general easily done by a common identifier. However, there are some caveats. The monthly balance sheet statistics (BISTA), the external position of banks (AUSTA), the quarterly borrower statistics (VJKRE), the MFI interest rate statistics (ZISTA), and the selected master data for MFIs (MaMFI) are all panel data. They have a common identifier BAID ('bank identifier') and a common representation for the date. It is delivered in separate variables YEAR, MONTH, and DAY so that even different frequencies can easily be linked.<sup>1</sup> The drawback is, that the user has to create his or her own time variable. Exceptional with respect to the identifier are the monthly balance sheets statistics for multi-office banks and the banks' profit and loss accounts.

## 2 Multi-office banks

A multi-office bank (Gesamtinstitut) is a bank with legally dependent branches abroad. Their data is consolidated. For the monthly balance sheet statistics there is data for the domestic part of a bank, for its legally dependent branches abroad, and consolidated data for the multi-office banks. Multi-office banks have their own identifier, i.e. a BAID that is unrelated to the identifier of their domestic part (Inlandsinstitut). So we need a mapping between the identifier of the multi-office bank (BAID) and the identifier of its domestic part (BAIDdom) if we want to work with multi-office banks.

## 3 Banks' profit and loss accounts

The situation for the banks' profit and loss accounts (GuV) is related to the multi-office banks. The results from the profit and loss accounts are based on the published annual reports of the individual institutions in accordance with the provisions set forth in the German Commercial Code (Handelsgesetzbuch or HGB) and the Regulation on the Accounting of Credit Institutions (Verordnung über die Rechnungslegung der Kreditinstitute, or RechKredV). Data for legally dependent branches abroad for example should be included in a bank's profit and loss accounts whereas they are not accounted for in other statistics. For the external position of banks and for the monthly balance sheet statistics legally dependent branches abroad have their own reporting obligations.

Therefore, if a bank has legally dependent branches abroad it is advisable to link its profit and loss account with the data for the multi-office bank and not with its domestic part.<sup>2</sup> The banks' profit and loss accounts use the identifier of the domestic part of the bank even if it owns legally dependent branches abroad. The variable gvBAID identifies a bank in the banks' profit and loss accounts over time and is identical to BAIDdom. The banks' profit and loss accounts have to be linked with other data using gvBAID. Please notice, that not for every institute in BISTA a profit and loss account is available. After renaming BAIDdom to gvBAID in the mapping between the identifier of the multi-office bank (BAID) and the identifier of its domestic part (BAIDdom) we can link the data.

The variable gvBAID identifies a bank in the banks' profit and loss accounts over time and BAID is the respective entity in the other data. If a bank has no legally dependent branches abroad both identifiers are identical. If a bank owns legally dependent branches abroad the BAID of the banks'

---

<sup>1</sup>The quarterly borrower statistics (VJKRE) is provided for the last month of the quarter (and the last day).

<sup>2</sup>The variable reporting the bank's capital in the GuV nonetheless is taken from the domestic part's monthly balance sheet, however.

profit and loss accounts is the BAID of the multi-office bank in the other data. Notice that a MFI in the banks' profit and loss accounts may have two different BAIDs since at times it may have legally dependent branches abroad and at different times not.

## 4 Different frequencies

Data comes in different frequencies. There is the monthly balance sheet statistics (BISTA), the quarterly borrower statistics (VJKRE), and the yearly statistics on the banks' profit and loss accounts (GuV). Master data (MaMFI) is monthly. The easiest way to proceed is to aggregate high frequency data to the lowest frequency available while sacrificing the information contained in the high frequencies. A more cumbersome approach would be to disaggregate the low frequency data. It is best to start the explanations with linking the master data.

### 4.1 Linking master data

For time-invariant variables frequency is no problem. The location, here federal state, may change within the year or quarter, not at its end. One may choose the most frequent case or the end of period value. Things are getting trickier with mergers and acquisitions. It may happen that bank 'B' absorbs bank 'A' and bank 'C' absorbs bank 'B' later but within the same period. In the lower frequency data bank 'C' has to appear as the only 'customer'.

### 4.2 Linking yearly GuV and monthly Bista

Aggregation to the lowest frequency available can be done by averaging or by taking the end of period value of the high frequency data. At the beginning of a time-series, low frequency data may cover a time-period that is not available for the high-frequency data. E.g. if the reporting obligations for the monthly balance sheet statistics start in April 2010 but the bank existed before, the bank's profit and loss accounts cover the months January 2010 until March 2010, too.<sup>3</sup> At the end of a time-series, high frequency data may still be available, while there is no low frequency data anymore. E.g. if a bank ceases its business in April 2010, high frequency data is available for the months January 2010 until March 2010 but the bank's profit and loss accounts data ends in 2009.

A merger or acquisition is a special case of the aforementioned. Assume bank 'B' absorbs bank 'A' in October 2010. For the absorbed bank, its monthly balance sheet data is available for January 2010 until September 2010 but its profit and loss accounts data ends in 2009. On the other hand, the customer's profit and loss accounts data for 2010 includes bank 'A's' data even for the months January 2010 until September 2010. If there is no major interest in the period of the acquisition one may predate the merger, in the example to December 2009. The high frequency data of the absorbed bank for the suppressed period then has to be assigned to the customer's data.<sup>4</sup>

---

<sup>3</sup>E.g. the balance sheet total reported in the GuV is calculated by averaging the monthly balance sheet total available assuming that the average monthly balance sheet total for the months not available does not differ. If the balance sheet total is increasing (decreasing) the balance sheet total in the GuV is underestimated (overestimated).

<sup>4</sup>E.g., in the year of the acquisition the balance sheet total reported in the GuV for the customer is calculated by summing up the balance sheet total of the customer and the acquired bank and dividing by twelve.