

Securities holdings statistics in Germany: A flexible multi-dimensional approach for user-targeted data provision

(Paper presented at the 58th World Statistics Congress of the International Statistical Institute (ISI 2011), Dublin, Ireland, August 2011)

Matthias Schrape

Deutsche Bundesbank, Statistics Department,

Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main, Germany

E-mail: matthias.schrape@bundesbank.de

Abstract

The high speed Information Technology developments of the last years enables data capturing to levels never possible before. IT Systems can generally reveal everything from the most granular element of a financial activity to systemic patterns of global markets – at a frequency much higher than before. However, data alone do not necessarily give insights as every statistician well knows. In order to turn a numerical ocean into valuable information streams for users it is essential to have “agile” statistical tools such as micro-databases and software for performing advanced analytics. Starting with this in mind the Bundesbank set-up a securities holdings statistics database for Germany in 2005 which currently covers comprehensive information for more than a million different securities. The securities holdings data are provided quarterly by 2,000 reporting agents on a security-by-security level. On one hand the system serves as a hub providing standardised statistical products for e.g. Financial Accounts, Government Finance Statistics, Balance of Payments Statistics, International Investment Position, Statistics for Investment Funds and Insurance Corporations. On the other hand, the granular data are stored in a multi-dimensional system to enable high service level for ad hoc user requests for e.g. information on financial risks. Finally, user-tailored data are provided for a number of research projects in the area of financial stability, international economics and for studies on investor behaviour. Future development plans include a move to a monthly frequency and the integration of consolidated data for large financial groups to satisfy further information needs addressed by users in the course of the current financial crisis like in the context of the G20 recommendations.

1 Introduction

Today, agile statistical tools such as micro-databases store and process amounts of data which could not be even imagined a few years ago. Information can be sliced and diced in many ways, supported by software for performing advanced analytics enabling statisticians to serve evolving user needs in a very flexible manner. Item-by-item reporting systems emerged as a result of this high speed Information Technology progress. Meanwhile, these systems are regarded as best practice and are implemented all over the place.

The underlying idea of an item-by-item reporting system is that statistical compilers receive highly granular data to produce all relevant output rather than relying on pre-structured data received from reporting agents. With the collection of data on single items such as individual securities, loans or investors a high flexibility is achieved. New requirements can be adapted easily without increasing the burden on reporting agents. In addition, the provision of raw, less aggregated data can help to reduce costs for reporting agents and at the same time contribute to an enhanced data quality.¹

Data alone, however, rarely speak for themselves. Thus, a user-oriented statistical approach demands more than “just shift the data and pass it on”. To gain informational value data have to be well organized, transformed, and provided in a way that gives insights to users. In that respect it is necessary to understand user-by-user their needs, interests, expectations, abilities and knowledge.

Having this in mind, the Bundesbank set up a highly flexible security-by-security information system for Germany to support various users in achieving their goals in the most efficient and effective way. The system uses an internationally standardized securities identification number – the ISIN code – to cross reference the data on individual securities holdings with the respective reference and price data.²

This paper aims to give an overview of the Bundesbank’s securities holdings information system, describing the input data captured, the processing steps carried out, and the main user needs fulfilled. Finally, the benefits of such a system as well as the future developments plans are briefly outlined.

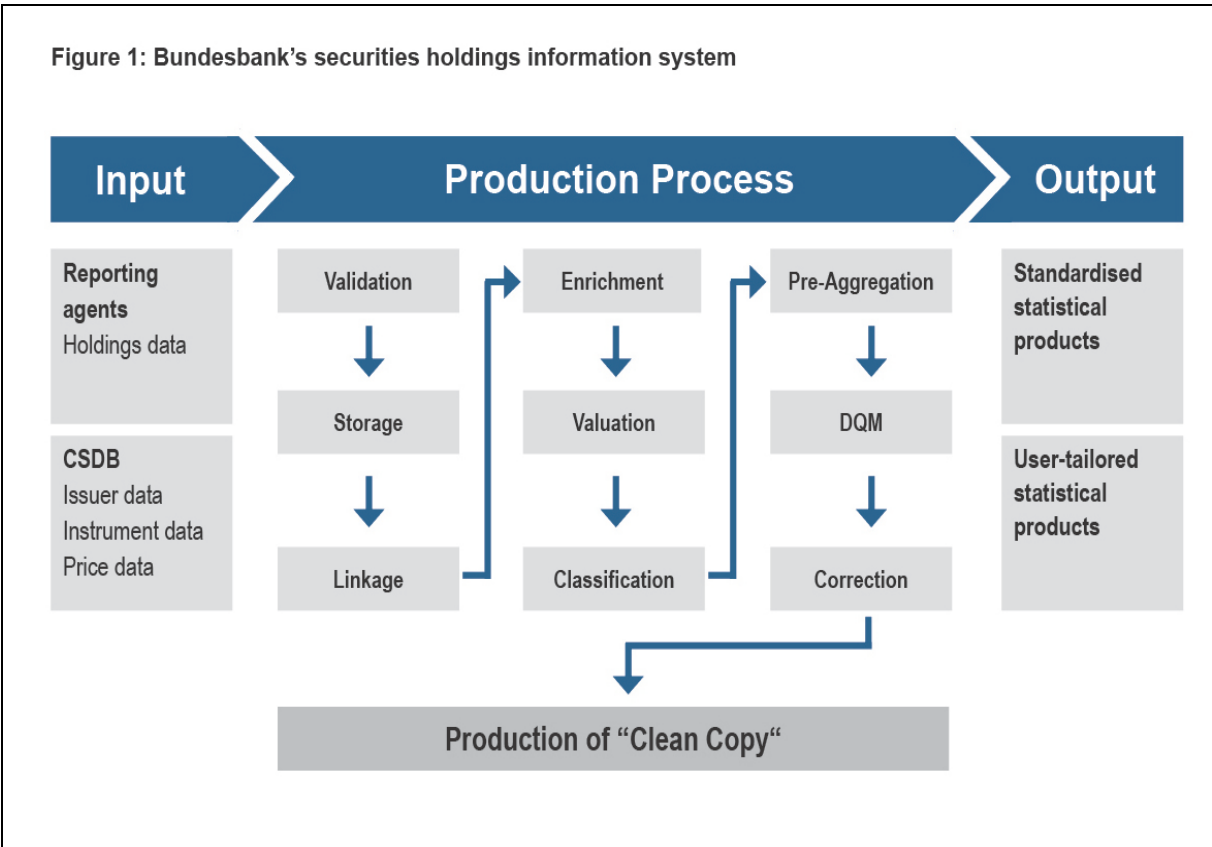
¹ See Brunken (2009).

² See also Amann, Baltzer, and Schrape (2011).

2 German securities holdings information system

The Bundesbank set up a modern, high-performance IT infrastructure in 2005 for the transition to a quarterly security-by-security collection system.³ The basic idea was to link individual securities holdings data transmitted by reporting agents with relevant issuer, instrument and price data extracted from the Centralised Securities Database (CSDB) in an efficient and smooth way. In order to enable a timely and flexible analysis of large data volumes, the securities holdings statistics database was supplemented by a business intelligence solution.

The database currently (1st quarter of 2011) processes data of around 450,000 different securities per quarter. 2,000 reporting agents transmit data on approximately 6 million individual holdings of securities. Today – after being in operation for 5 years – the database comprises information on more than one million securities for which over a hundred million data on individual holdings have been stored. This chapter continues with an explanation of the input data, production of securities holdings information and the research tool available for data analyses. Figure 1 provides an overview of the data flow within the securities holdings information system.



³ The Bundesbank has been compiling data on securities holdings since 1962. Initially, the data collection was conducted annually on an aggregated basis.

Data-input: Collection of a very basic set of information

The securities holdings data in Germany are collected on the basis of a mixed approach, combining the advantages of direct and indirect reporting schemes. Financial institutions located in Germany report any securities they are holding for domestic and foreign customers. The financial institutions obliged to report comprise domestic banks (monetary financial institutions excluding money market funds), domestic investment companies and “other” domestic financial companies. In addition, domestic banks provide information about their own holdings, irrespective of where the securities are held, possibly also abroad. The data are collected by a full census, i.e. no reporting thresholds apply. The data collection involves holdings of debt securities, equities and mutual fund shares, no matter where the securities were issued and in what currency they are denominated.

In order to keep the reporting burden at a minimum, only a very basic set of information is required to be reported. This includes the ISIN, the nominal amount or number of units held, and the sector and residency of the holder. Furthermore, as the significance of securities repurchase and securities lending transactions has increased strongly in recent years, securities holdings which are passed on or acquired as part of such contracts are to be flagged separately. This information is mandatory for the own holdings of domestic banks only. The same applies for the distinction between direct investments and portfolio investments to be made. Identifying the ultimate holder of securities, however, can pose a challenge to statisticians, as custodians often hold securities through other intermediaries. In order to ensure that holdings of securities are not captured in the statistics more than once, only holdings from end investors are to be included.

The issuer and instrument information needed, including the sector and residency of the issuer and the instrument features are extracted from the CSDB. In addition, CSDB price information is used for the calculation of market values. The CSDB covers individual reference data for all securities relevant for statistical purposes of the European System of Central Banks (ESCB), containing at least all securities issued by euro residents and those securities issued by residents outside the euro area and denominated in Euro. Moreover, all non-euro area securities (not denominated in Euro) most likely held by euro area residents are covered.⁴ The existence of such a large reference database with information on securities issued worldwide was a prerequisite for the transition to the security-by-security collection system. With its decision to extract all reference and price data from the CSDB from 4th quarter of 2005, the Bundesbank became the first National Central Bank using the CSDB as single data source.

⁴ See ECB (2010).

Production process: Quality management and data provision

The production process includes the reception and storage of reported securities holdings data and the extraction of relevant reference data from the CSDB, before being linked, enriched, valued, classified, pre-aggregated, checked and corrected.

The production process starts with data reception and automated format checks. If reported data do not meet formal requirements the report is rejected with an auto-reply indicating the reason for the failure. As a result, only valid securities holdings data are entering the production environment. In a next step, the reported data are cross referenced via the ISIN code against the relevant issuer, instrument and price data extracted from the CSDB and then stored in the database. In case of missing and incomplete CSDB data automated default algorithms enrich the data set. The next production phase includes the valuation, classification and pre-aggregation of data to allow second-level checks on micro data as well as third-level checks on aggregated data.

Quality analyses on micro data are supported by automated filters indicating significant changes or unexpected patterns which require further investigations. For instance, some filters reveal improbable combinations of sector and country of holder. With the quality control on aggregated data structural breaks are monitored and cross checks with other statistical sources are carried out. For example, the comparison of the total holdings for one security across all investors with the amount in circulation indicates gaps or double-counting. The data can be easily browsed to either detect the source(s) that caused an error or to access the impact of a specific failure on the overall results. However, due to time and cost constraints it does not seem adequate in a security-by-security world to go for a data accuracy of 100 %. The Bundesbank therefore implemented a balanced approach with well defined thresholds. Major outliers and implausible information revealed in the data production process are finally discussed with the reporting institutions. In case of erroneous data revised reports are submitted. If securities reference information has to be corrected the Bundesbank changes the CSDB data. Corrections, if necessary, are being performed on the level of individual securities. The additional information gathered throughout the production process like qualitative information on certain securities or investors are stored separately. This information serves as basis for future production cycles as well as for answering questions of researchers on historical data.

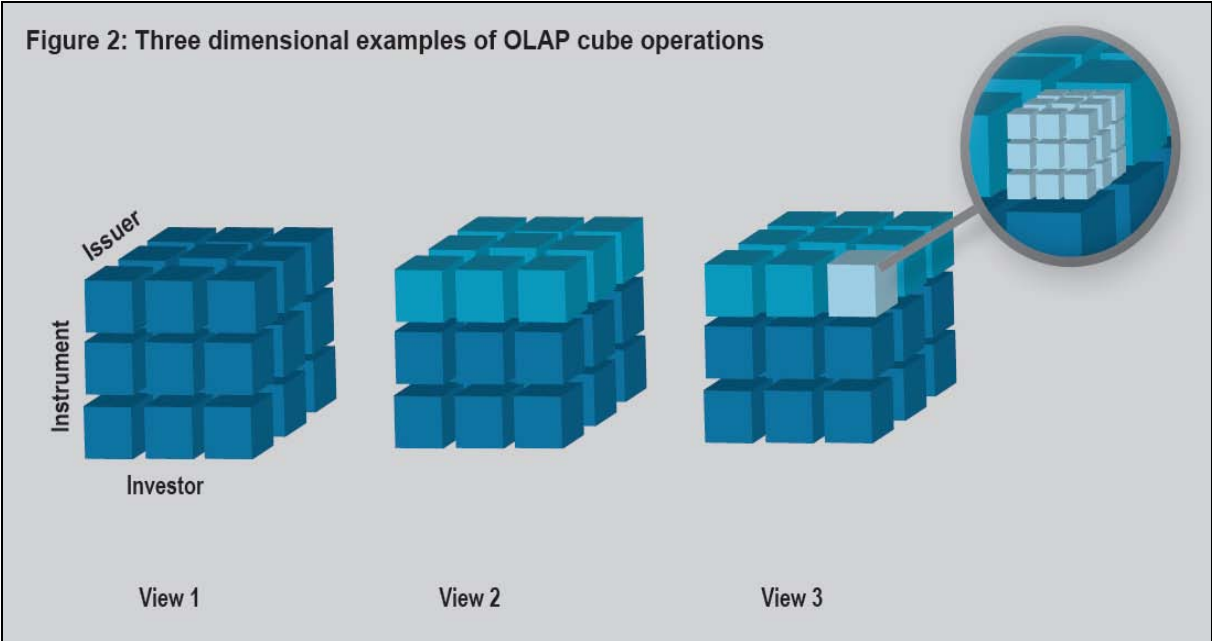
Once the data are analysed and cleaned, they are loaded into the time series database and into the Online Analytical Processing (OLAP) system to be made available to users.

OLAP system: An analytical tool for customised data provision

Due to the vast amounts of data stored in the securities holdings statistics database it was crucial for the Bundesbank to implement a business intelligence system like an OLAP tool. The high-performance OLAP system allows for complex analytical and ad-hoc queries on granular data from multiple perspectives with a short response time. OLAP systems are used for analysing multidimensional data which could be structured according to cubes.

The OLAP system covers facts or measures and dimensions extracted from the securities holdings database. The dimensions represent the focus of a specific analysis. They are organised hierarchically according to the level of granularity, whereas the members of one level can be aggregated to members of higher levels. For example, the types of instrument “quoted shares“, “unquoted shares” and “other equities” are aggregated to “shares and other equities” and all of them are part of the dimension “instrument classification”. The measures are numerical attributes, i.e. the stock data reported can be analysed in different dimensions. To keep pace with the growing size of data as well as the number and complexity of user requests Bundesbank’s OLAP system was extended beyond a standard OLAP cube design. This includes that data are stored on different aggregation levels to further increase the query performance.

The following example aims to illustrate the functioning of the OLAP system (see also figure 2). For simplicity reasons it is assumed that there are only three dimensions available for research.



View 1, for instance, supports an analysis of total amounts held for all securities and investors (for all countries and sectors). Whereby the edges of the cube represent the dimensions “investor”, “instrument” and “issuer”. If the information needed is more specific like the total amounts held of a certain type of instrument, the search would focus on a slice of the cube (see view 2). More detailed data research would require a search which is embedded in a broader environment – the so called dice (see view 3). With view 3 it is possible, for example, to extract the total amounts of a certain type of instrument launched by a specific issuer and held by a selected group of investors. The OLAP tool also allows to drill down and to roll up the data, i.e. to switch between different data aggregation levels. Furthermore, there is a sorting and pivot function which helps to rotate the data to provide alternative data presentations. With these operations it is possible to flip through the data and extract, aggregate, sub-divide the relevant information and to change perspectives.

3 Use of securities data

Due to the fast growing importance of securities markets, macro and micro information on issues and holdings of securities are attracting a great deal of attention. Securities data are indispensable for monetary analysis as any shifts in financing between the banking system and the securities markets may affect the transmission of monetary policy. In addition, data are essential to monitor the development of amounts held and the distribution among different (groups of) investors. Financial stability analysis makes use of the data to measure the risks associated with different types of instruments and exposures to individual counterparties. In addition, information on the composition and quality of the portfolios of holders is used to better understand investor behaviour. The recent financial crisis has raised this interest and further increased the demand for securities data. With the securities holdings information system the Bundesbank serves a wide range of user needs either by providing standardised or user-tailored statistical products.

Standardised statistical products

Standardised products are compiled quarterly and are mainly used as building blocks for producing further statistics and for publication purposes. Users can decide if they prefer to receive the data via time series or an application-to-application file transfer. The data are provided for Financial Accounts to analyse the flow of funds in the German economy as well as for the International Investment Position to assess the volume and structure of German external assets and liabilities. The data also allow cross checks with the transactions reported under the scope of Balance of Payments Statistics. Governmental holdings are provided for Government Finance Statistics, which serve as basis for compiling the German

Maastricht debt level. Securities information is also provided for preparing Investment Funds Statistics and Statistics on Insurance Corporations and Pension Funds. Data on German investors' holdings of foreign securities are provided to the International Monetary Fund (IMF) for the Coordinated Portfolio Investment Survey (CPIS). Finally, a selection of most wanted securities holdings data is published on the Bundesbank's website.

Ad hoc requests

The securities holdings information system was set up not only for the compilation of standard products but as well for flexible and timely response to ad hoc user requests. Most of these requests refer to different risk aspects and deal with the question of “who-holds-what-from-whom-and-how-much”. This includes analyses on risks regarding certain securities (e.g. real estate funds or asset-backed securities), issuers (e.g. Lehman Brothers), countries (e.g. in the context of the current sovereign debt crisis) and currencies. Further issues addressed are risks of interest rate changes or reinvesting risks.

Research projects

The granularity of securities holdings data and the highly flexible IT infrastructure have led to a number of research projects in recent years. A research network was established as the securities holdings data of the Bundesbank has gained quite some popularity. Members of this forum are researchers and statisticians meeting on a regular basis to present and discuss studies on securities data and to exchange views on future activities. This platform also offers the opportunity to statisticians to get a better understanding of the needs and constraints of users. The following list presents an overview of some research projects which are based on individual securities holdings data:

- **“Is proprietary trading detrimental to retail investors?”⁵**

The researchers studied the conflict of interests that might arise at universal banks between their proprietary trading and their retail banking. Using a set of data that covers the stock investments of individual domestic banks and of its respective retail customers on a security-by-security basis the return characteristics of those stocks that flow from a bank's proprietary portfolio into its respective customers' portfolio were analysed. The results show that stocks which are transferred from banks' proprietary portfolio into the portfolio of their customers are of lower performance than the average stock held in banks and customers portfolios.

⁵ Fecht, Hackethal, and Karabulut (2010).

- **“Home-field advantage in equity investments? Empirical evidence on local bias among German individual investors.”⁶**

This project was set-up to analyse the effect of geographic proximity on portfolio choice. The data for security-by-security stockholdings of private households at regional banks in Germany suggest a strong and consistent over-investment in geographically close companies. However, comprehensive performance analysis shows that households do not earn excess returns on their local stock investments. Obviously, individual investors do not possess value-relevant information about local stocks.

- **“Are there disadvantaged clienteles in mutual funds?”⁷**

This research project further investigated the flow-performance relationship of three different investor groups – financial corporations, insurance corporations and pension funds, and private households – in mutual funds. The paper concludes that financial corporations have a strong tendency to chase past performance. Insurance companies and pension funds show some evidence of performance chasing, but (probably due to investment restrictions) are underrepresented in the best performing funds. Households chase to some extent past top performer but they are also driven by advertising and status quo bias.

4 Summary and outlook

The Bundesbank’s securities holdings information system opens up a broad spectrum of potential data uses. The information is currently provided for the compilation of various statistics, for data publication, to monitor financial stability, for supervisory purposes, and for research projects. As statistical data are produced from a granular basis, new requirements can be adapted easily without addressing reporting agents. The detailed data also allow sophisticated micro and macro quality analyses. Furthermore, only a very basic set of data has to be submitted by reporting agents. As the relevant reference data is obtained from the CSDB, a consistent valuation and classification of statistical units is ensured.

The demand for securities holdings information in and outside the Bundesbank has further increased since the onset of the financial crisis. As the importance of securities markets is likely to grow and periods of financial imbalances are increasingly prevalent securities holdings data are more and more used to answer questions on financial stability aspects and for supervision. Owing to the high degree of detail and the possibility of combining numerous dimensions, securities holdings data continues to offer significant potential for new research

⁶ Baltzer, Stolper, and Walter (2011).

⁷ Jank (2010).

studies. Proponents of granular data collection systems sometimes indicate that the very extensive data can also provide answers to questions which have not been yet been posed.

However, the current financial crisis proved as well that there are still information gaps which need to be closed.⁸ Therefore, further development projects are planned in the area of securities holdings statistics for the coming years. It is intended to shift the German data collection from a quarterly to a monthly frequency. In addition, there are plans at European level to collect consolidated data of large financial groups. Finally, in cooperation with the ECB, the Bundesbank is developing the Securities Holdings Statistics Database (SHSDB) covering information on securities investments for the entire euro area/ESCB. A number of improvements and new avenues for future analysis and research will open up following the introduction of the SHSDB in Europe.

⁸ See International Monetary Fund and Financial Stability Board Secretariat (2009 and 2010).

References

Amann, M., M. Baltzer, and M. Schrape (2011). Micro-database: Securities Deposits Statistics. A flexible multi-dimensional approach for providing user-targeted securities holdings data. Technical documentation.

Baltzer, M., O. Stolper, and A. Walter (2011). Home-field advantage in equity investments? Empirical evidence on local bias among German individual investors.

Brunken, S. (2009). Cooperation to improve European and national securities statistics. IFC Bulletin No 29.

ECB (2010). The Centralised Securities Database in Brief. European Central Bank.

Fecht, F., A. Hackethal, and Y. Karabulut (2010). Is Proprietary Trading Detrimental to Retail Investors?

IMF and FSB (2009). The Financial Crisis and Information Gaps. Report to the G-20 Finance Ministers and Central Bank Governors. International Monetary Fund and Financial Stability Board Secretariat.

IMF and FSB (2010). Progress Report Action Plans and Timetables. International Monetary Fund and Financial Stability Board Secretariat.

Jank, S. (2010). Are There Disadvantaged Clienteles in Mutual Funds? Deutsche Bundesbank Discussion Paper Series 2: Banking and Finance Studies No 11/2010.