

A Research Data and Service Centre (RDSC) at the Deutsche Bundesbank – a draft concept

Ulf von Kalckreuth¹

Abstract

This paper summarises the state of discussion in the Statistics Department of the Deutsche Bundesbank on setting up micro data based information systems. Currently, the production systems are not geared to micro level data analysis. Changing them, will be a time and resource consuming endeavour. The suggested approach features a short term enhancement on the basis of existing infrastructure. A "House of Microdata" will be created, essentially a Statistical Data Warehouse, together with a Research Data and Service Centre. The pre-existing statistical processes will be left intact, minimising operational risk. This is at once a huge leap forward towards availability of micro data for analysis and research, and a step on the way to a longer term reform of the informational infra-structure and data-management. The enhancement will be greatly helped by the fact that the Deutsche Bundesbank has a long-standing tradition of providing micro-data to internal and external researchers on-site.

¹ Deutsche Bundesbank, Statistics Dept., Central Office, ulf.von-kalckreuth@bundesbank.de.

A Research Data and Service Centre (RDSC) at the Deutsche Bundesbank – a draft concept

This paper summarises the state of discussion in the Statistics Department of the Deutsche Bundesbank on setting up micro data based information systems. Currently, the production systems are not geared to micro level data analysis. Changing them, will be a time and resource consuming endeavour. The suggested approach features a short term enhancement on the basis of existing infrastructure. A "House of Microdata" will be created, essentially a Statistical Data Warehouse, together with a Research Data and Service Centre. The pre-existing statistical processes will be left intact, minimising operational risk. This is at once a huge leap forward towards availability of micro data for analysis and research, and a step on the way to a longer term reform of the informational infra-structure and data-management. The enhancement will be greatly helped by the fact that the Deutsche Bundesbank has a long-standing tradition of providing micro-data to internal and external researchers on-site.

1. Mandate

In line with the objectives set out in the Bundesbank's Strategy 2016, micro data are to be made available, in a new setting and "philosophy", for both analytical tasks and scientific research. This is partly in order to meet the increased data needs of the Bundesbank's departments and the ESCB resulting from the expansion in their range of duties and the new issues they face. The following problems need to be solved:

- Data management in the Statistics Department (and elsewhere in the Bundesbank) is not primarily designed to support micro-level analysis; it is usually based on process data used to create statistical aggregates or data for prudential or administrative purposes. Micro data are generally not available as closed data sets;
- Documentation for use by externals is often absent or incomplete;
- The use of micro data is restricted by data protection rules; different legislation applies in different contexts. The level of confidentiality has to be systematically assigned to user groups;
- The data are created through various processes and are not integrated. For any given financial institution, the Bundesbank has data from various sources, eg the bank balance sheet statistics, the borrowers statistics, the external position and prudential reports; it is very difficult to analyse these data simultaneously;
- For both internal and external researchers and analysts, working with unknown and/or complex micro data requires investing a lot of time and effort. It takes long to understand the content of a data set – including restrictions and sources of error – well enough to draw reliable conclusions regarding distributions.

The proposed solution is an interdepartmental project coordinated by the Statistics Department to set up a "**House of Microdata**" (**HoM**) and a **Research Data and Service Centre (RDSC)**.

1. The “House of Microdata” will be the infrastructure component, receiving “clean copies” of process data from data production entities, formatted in such a way as to facilitate their joint use. The data clusters will be selective and geared to the use for analytic purposes. In the House of Microdata, the data can be retrieved by anybody within the Bank who holds the access rights. An existing, powerful platform for storage and retrieval of time series can be extended, and the well-established SDMX-standard will be used to create data models.
2. The Research Data and Service Centre will **facilitate data access** for researchers and analytics, as well as **provide important services**:
 - compiling, documenting and archiving informative micro data sets,
 - enhancing existing data through record linkage,
 - supplying internal and external researchers and analysts with micro data sets in line with the relevant legislation,
 - methodological and content-related support and services for users,
 - analyses and customised processing of requests from internal analysts in cases where legal constraints prevent them from accessing the data or where it would be too time-consuming for them to process the data themselves,
 - supplementary methodological and descriptive research based on the data sets created by the unit.

Looking ahead, potential is to be created for

- receiving and supplying micro data from other departments and external sources,
- exchanging research data within the ESCB and with other research data centres.

Strategically speaking, the production and provision of micro data sets fit for micro level analysis is a new and unique type of statistical product; the outcome of the statistical process is not the indicators themselves, but a **new possibility for users to produce their own statistical results following their own criteria**. To support this process, the unit will provide documentation and advisory services, clarify organisational and legal issues, and create and operate a technical infrastructure which allows micro data to be processed within a governance structure that fulfils all the relevant legal requirements and the principle of “participate and contribute”.

By contrast, in conducting its own analyses on behalf of others, the RDCS will essentially be providing one of the Statistics Department’s classic products: it will aggregate the available individual data to derive moments and indicators for distributions. Users generally want information on distributions – such as risk measures or correlations – rather than individual data. However, the RDCS would provide information flexibly according to varying user requirements. This service is particularly suitable for meeting the growing data needs of other Bundesbank’s departments and the ESCB. Where certain analyses come up time and again and can be standardised, the unit will also assume development functions for the entire department.

Given the RDCS’s role in development and dissemination, it will need to carry out a certain amount of own research. The unit needs to compile research data sets

using process data. This is a constructive service which will have to be geared to the users' needs. Developing a research data set is, in itself, a form of research. It requires explorative work and examination of current issues and standpoints in the world of research. A minimum level of active research is one of the prerequisites set by the German Council for Social and Economic Data (RatSWD) for recognising an institution as a research data centre.

2. Developing the RDSC

Initially, the RDSC's work would focus mainly on data structures for which the Bank already has experience in micro-level scientific analysis. These data structures would take the form of scientific use files (SUFs), would be saved periodically (data freeze), archived in line with DOI standards and documented. User rights and restrictions would be clarified and an application procedure set up. Record linkage would be carried out on the basis of available characteristics. The statistical process routines used at present would remain unchanged; individual data would be selected periodically according to certain rules and – where necessary – manual record linkage would be carried out using available common identifiers. The trade-off for this would be a fairly long time lag between the reporting or recording of the original data sets and collective availability. This would be less of a problem for academic users than for analytical work. Users will also generally need to be provided with sample routines for generating aggregates both at the level of observational units and for sub-populations.

Setting up this unit will be greatly helped by the fact that the Research Centre of the Deutsche Bundesbank has a long-standing tradition providing micro-data for researchers on-site. In the past, micro data sets have been provided on special requests, concerning

- Corporate Balance Sheet Statistics,
- MFI Monthly Balance Sheet Statistics (BISTA) and profit and loss accounts,
- External position of banks,
- Statistics on Securities Investments (formerly Securities Deposit Statistics),
- German Panel on Household Finances (PHF),
- Micro-Database Direct Investment (MIDI),
- Several merged data sets.

It could be a suitable point of departure to fully develop these micro data for analysis. While supplying users with an initial menu of data sets, the unit would simultaneously start to develop more and better integrated micro data sets and well-organised statistical support for use by internal and external analysis.

3. The structure of an RDSC

Apart from a manager, an RDSC would require

- a) **experts for the existing data sets**, who would be responsible for the creation and development of microdata sets, documentation, data archiving and user support. They would act as a liaison point between users and the departments responsible for the data in question. Additionally, they would carry out special analyses on request and conduct a limited amount of research.

- b) **qualified staff for the technical provision of data** by means of
- teleprocessing
 - a secure on-site computer
 - de facto or fully anonymised scientific use files
 - monitoring data output
 - (possibly) an exchange with other research data centres
- c) **qualified staff for merging and validating data**, as well as supplementing data with additional information, such as sector identifiers.

4. Integrated micro data management

With an integrated approach to processing micro data, each characteristic should in principle only be recorded once – this characteristic is then added to the existing data on the respective observational unit and made available with all others. Such a state of integration largely **eliminates statistical redundancy**, reduces the workload of the data providers and increases the volume of information that can be analysed simultaneously. Record linkage would be a by-product of data processing.

From an internal perspective, it is of particular importance that different units in the department, working with the same observational units (banks, enterprises, securities), are able to **exchange data more easily**. This will facilitate more efficient, rapid and flexible processing of statistical data as well as enhancing quality control in the production of the data. Ideally, the goal is to make all statistical information **available as both aggregated and granular data**, allowing them to be combined in various ways depending on the requirements of the task at hand. This would e.g. be very useful for examining financial risk issues from a new angle.

It will be easier to apply this to banking statistics than to firm-level data, for which there are currently no common indicators in the Bundesbank. An integrated approach to firm level data is severely affected by the fact that currently, for legal reasons, the Bundesbank has no access to the Federal Statistical Office's business register. Much progress in the integration of data has already been made in the area of securities. If need be, the processes used in the production of statistical data will have to be adapted or replaced. Ultimately, the reporting system will have to be overhauled.

The existing production systems answer diverse analytical and administrative questions by equally diverse and independent statistical processes, very often operating under different legal and institutional frameworks. Any step of further integration must be **realised under the strict condition of continuous functioning** within the pre-existing organisational and legal context.

A very important step towards integrated data management will be achieved by **setting up the House of Microdata**, which is essentially a Statistical Data Warehouse for micro data. The SDMX based data models will be created in such a way as to facilitate the joint use of the data, and, where possible, there will be joint reference data for the micro units. This will enable the statistical data production units to use the finalised data of other units for consistency checking and outlier analysis. However, this is not tantamount to integrated data management on the process level. A full integrated data approach can be approached only gradually, and as a **result of an evolutionary process**.

The associated work will involve many units and can by no means be carried out entirely (or predominantly) by a separate unit. However, the data repositories that will be created as a result of a reorganisation of data could be operated centrally for all other units. The RCDS's potential would be amplified by employing an integrated approach to data processing – particularly because the data which have already been integrated would no longer have to be merged.