New Developments in the Bundesbank's Research Data and Service Center (RDSC):

Linking Various Sources as a Key for Unleashing More Potential than Ever Before

Jens C. Orben, Stefan Bender (both Deutsche Bundesbank)

Since mid-2014, Deutsche Bundesbank operates its Research Data and Service Center (RDSC) in order to provide researchers access to its vastly available high-quality microdata, unleashing its potential to the scientific community. In the last three years, Bundesbank's RDSC has increased the number of available standardized microdata sets by more than 50 % (15 in total by now), covering information on banks, securities, enterprises and households. Setting the latter aside, these data are all able to be linked to other internal as was as external data. In order to foster the linkage process, the RDSC now offers key metadata sources as "selected master data of banks" (MaMFI) and "Centralized Securities Database" (CSDB) and further ID-matching tables, created in a sophisticated record linkage process.

Especially microdata on securities (trading) are of great interest for research in finance. That is for instance the "securities holdings statistics" (SHS-Base plus), available for the time period beginning in 2005 covering all securities on a security-by-security basis, held by a notifiable financial institutions (located in Germany) in safe custody for national or foreign depositors. In addition, national monetary financial institutions (MFIs) have to report their own holdings, stacking up the number of observations to around 6 million per month.

By itself the SHs-Base plus can be used to conduct household finance, capital market and banking research. However, its analytical potential upsurges when combined with additional information on banks, e.g. matching detailed core banking business information from the monthly balance sheet statistics (BISTA, offering around 3,500 different positions for each bank). For more than 10 million different securities, there are more than 80 instrument (eg. debt type, yield, maturity, issuance, redemption, coupons, splits) and issuer attributes (eg. sector, country, IDs) from the above mentioned CSDB ready to be merged and analyzed together with the CSDB. Permeating the investments' characteristics even closer, assets being investment funds may be analyzed, using the investment fund statistics (IFS-Base). The IFS-Base entails on a fund-by-fund and security-by-security basis even richer information than the SHS-Base, additionally including other asset types than securities and the liability structure.

The above mentioned datasets show exemplarily the merit of all available datasets individually and the enormous possibilities when merging them. The success of RDSC microdata since 2014 is illustrated best with the following figures:

- Project applications: 565 (355 successful)
- Active projects: 312
- Mean number of datasets per project in 2018: 2.68
- Number of collaborating institutions: 160 worldwide, of which 90 abroad

An overview of papers that have successfully used Bundesbank microdata is available online (BISTA, SHS-Base plus). Among those are publications in the Journal of Financial Intermediation, The Journal of Finance, Journal of Financial Economics, Journal of Financial Stability, Review of Finance, The Quarterly Review of Economics and Finance, Journal of Banking and Finance and International Journal of Central Banking.

Besides the data itself, RDSC staff provides standardized documentation in English for each dataset - minimizing entry cost for researchers, harmonizes variable names and formats - simplifying the linkage, does a DOI registration for each dataset - making it quotable and archives projects after publication – allowing replication and underlining good scientific practice.

Due to legal requirements and in order to meet data protection rules, individual data can be made available only under certain restrictions: An application entailing a research question in need of microdata and a CV by the respective researcher(s) are to be sent to the RDSC. If successful, Bundesbank is happy to welcome researchers to the save environment/center in Frankfurt, Germany to work with the data. A growing number of datasets is ready for remote execution, where researchers write their code - after at least one research visit at the premises of Bundesbank - anywhere, send it via email to the RDSC and get their results back (only anonymized output is leaving the RDSC).

Corresponding author: Jens Orben, Deutsche Bundesbank, Central Office, Wilhelm-Epstein-Straße 14, D-60431 Frankfurt, Germany, email: jens.orben@bundesbank.de.