29 May 2024

Guidance Note 5 — Development and maintenance of a security-by-security database

Archana Dilip and Bruno Tissot, Bank for International Settlements[1]

This guidance note sheds light on various aspects related to the development and maintenance of a security-by-security database (SBSDB), particularly:

- *the increased interest expressed by central banks in setting up a SBSDB;*
- *the scope of a typical SBSDB, and the associated key benefits and challenges;*
- *the different stages involved, especially as regards the collection, processing, quality management and dissemination of data;*
- *some recommendations for strengthening the related data governance when maintaining a SBSDB.*

1. Introduction

The micro data revolution witnessed in recent decades has **made available large amounts of granular and complex data sets to public authorities.** Such information allows focussing on specific areas of interest in a targeted, flexible and agile way.[2] The developments in this context have been particularly relevant for central banks, reflecting their dual roles as compilers of official statistics and as users of data in pursuing their public mandates.

One pivotal factor driving the shift from conventional "macro" frameworks to more micro-level approaches has been the **Great Financial Crisis (GFC) of 2007–09**. This event underscored the importance of making better use of disaggregated data to accurately identify financial stability risks. As a result, the various phases of the G20-endorsed Data Gaps Initiative (DGI), set-up after the GFC to address the most pressing data needs, comprised a whole range of initiatives to expand and make use of more granular-level data on financial markets' participants, products and transactions.[3]

---

1 Respectively Visiting Statistician on secondment to the Bank for International Settlements (BIS) from the Reserve Bank of India (archana.dilip@bis.org; adilip@rbi.org.in); and Head of BIS Statistics and Research Support and Head of the Secretariat of the Irving Fisher Committee on Central Bank Statistics (IFC) (bruno.tissot@bis.org; ifc.secretariat@bis.org).

This guidance note is based on the voluntary responses received for an internal survey organised in the context of the knowledge centre of the IFC of the BIS at the request of interested central banks.

The views expressed here are those of the authors and do not necessarily reflect the views of the affiliated institutions and the various other organisations mentioned in this guidance note. Bilyana Bogdanova, Tracy Chan, Edward Lambe and Olivier Sirello (BIS), Alfredo Fuentes (Central Bank of Chile), Johannes Kleibl and Johannes Micheler (European Central Bank (ECB)), and Kuniko Moriya (Bank of Japan) are very much thanked for their comments and suggestions.

2 See Israël, J-M and B Tissot (2021): "Incorporating micro data into macro policy decision-making", *Journal of Digital Banking*, vol 6, no 3, pp 233–50.

3 The first phase (2009–15) of the DGI targeted "*the possibility of improved collection and sharing of information on linkages*

Among these efforts, one important development has been the **global push for collecting more detailed information on securities**.[4] The demand for such statistics has been driven by various policy factors, especially in the areas of monetary and financial stability. Yet, there is no unique and comprehensive data source covering global securities, despite the claims made by a number of commercial vendors. As a result, various domain users within a central bank (or concerned government agencies), tend in practice to maintain decentralised (partial) securities databases, set-up in a rather ad hoc way. This can raise significant data quality issues, particularly in terms of timeliness, coverage, consistency and accuracy.

To address these challenges and offer a centralised response to the diverse data users' requirements (both internal and external), **a growing number of central banks as well as government agencies have shown interest in setting up a security-by-security database (SBSDB).** The primary aim is to collect data on individual securities, first from the **issuer** perspective (ie debtor, with precise identification of the issuers and details of their securities issued), and increasingly also with a **holder** perspective (ie investor, with identification of the holders of the securities and details on their holdings). This information plays a pivotal role in supporting the compilation of key macroeconomic statistics, such as the Balance of Payments (BoP) and the International Investment Positions (IIP), especially as regards their portfolio investment component – in many jurisdictions, these compilation exercises are generally shared between central banks and national statistical offices as part of their contributions to the National Statistical System (NSS). There are other important compilation exercises that can be effectively supported by a SBSDB, for instance, statistics on investment funds (IFs), financial vehicles and monetary financial institutions (MFIs). Moreover, central banks that have developed a SBSDB also use it for policy, research and other non-statistical tasks, such as prudential supervision.

Central banks' reported initiatives underline that **developing a SBSDB is a fundamental project**. In practice, it requires, firstly, the availability of adequate and unique identifiers, such as the ISIN (International Securities Identification Number) and the LEI (Legal Entity Identifier); secondly, the storage and set-up of an end-to-end data processing flow at a highly disaggregated, security-by-security level; and, thirdly, the classification of this information based on a spectrum of well-identified attributes and on specific criteria serving the intended purpose of setting up a SBSDB. The necessary information is generally collected from different custodians as well as from various entities potentially involved in a security's life cycle (eg its issuance, trading and holding), as well as commercial data providers. Setting up a SBSDB also calls for a clear understanding of the underlying scope, benefits and challenges as well as of the various steps involved in statistical processing. Lastly, it often requires a legal framework (at the national level).

Looking forward, global experience suggests that countries that maintain a SBSDB would have an interest in sharing more micro-level information in the context of an international micro database. One key reason is the internationalisation of the investors' base, meaning that a large number of securities

between individual financial institutions". The second phase (2016–21) aimed to "*promote and encourage the exchange of data and metadata among and within G20 economies, and with international agencies, to improve the quality of data, and availability for policy use*"; the G20 economies were also "*encouraged to increase the sharing and accessibility of granular data, if needed by revisiting existing confidentiality constraints*". The third phase (2022–27) aims to address priority policy needs in the areas of "*access to private sources of data and administrative data, and data sharing to improve the timeliness and granularity of official statistics*".

4 According to the Handbook on Securities Statistics (HSS), "***securities** include debt securities, equity securities and, to some extent, investment fund shares or units*". **Debt securities** are "*negotiable financial instruments serving as evidence of a debt*". **Equity** securities are "*negotiable financial instruments that entitle holders to a share of both distributed profits and the residual value of the corporation's assets in the event of its liquidation*". **Investment funds** are "*collective investment schemes that raise funds from the public by issuing investment fund shares or units, in order to acquire financial or nonfinancial assets.*"

issued in one place can be traded and held in other countries. A global database would thus certainly help in improving the quality of the statistics compiled and in turn enhancing policy analyses, especially those related to risk exposures and financial vulnerabilities. Yet, the international sharing of such information requires the full support of interested jurisdictions and the set-up of adequate agreements, as was the case for the highly confidential data on large global banks collected by the BIS International Data Hub.[5] It also calls for applying common statistical standards, such as the Statistical Data and Metadata eXchange (SDMX) standard.[6]

The present note discusses these various issues, based on an **internal survey conducted by the Irving Fisher Committee on Central Bank Statistics** (IFC) in the context of its knowledge centre initiative to foster exchange of experience on statistical issues among central banks, as well as on the available literature describing the experience of different NSSs. It covers a presentation of the scope of a SBSDB (Section 2) and the related benefits and challenges (Sections 3, 4). It then reviews the main elements of end-to-end processing flows for securities data, from collection to processing, quality management and dissemination (Section 5). Lastly, it outlines some best data governance practices that could be followed to strengthen the management of SBSDBs (Section 6).

2. Scope of a SBSDB

The **general scope of a SBSDB is to cover all the relevant information for the securities of interest** (the so-called "reference database") which are required for a particular statistical compilation exercise.[7] Yet, the IFC survey of central banks shows that the objective of developing and maintaining such databases can in practice vary significantly across projects.

First, the **reference database may comprise several layers of information** on individual records. As a starting point, it classifies the types of securities as debt securities, equities, investment fund shares and even financial derivatives in some cases. Furthermore, it contains reference data on each instrument (eg outstanding amounts, market capitalisation, original and residual maturity, type of security, currency of denomination, coupon payments, dates, interest rate type, etc), issuers (identifiers, name, residence, institutional sector or sub-sector, etc), prices (market price, issue price, redemption price), valuations (face, nominal or market values) and ratings (of the issuer or the instrument).[8] In addition, security-related events like liquidations, stock splits, conversions from debt to share, takeovers, mergers and other changes can also be an important component of the reference database; these events may be recorded as specific attributes characterising outstanding amounts. The resulting complexity implies that the content of a SBSDB can vary significantly from one country to another.

The second important aspect is sectoral classification. This is essential to ensure a consistent reconciliation between micro and macro data, which is one of the key goals of setting up a SBSDB. In

5 See Bese Goksu, E and B Tissot (2018): "Monitoring systemic institutions for the analysis of micro-macro linkages and network effects", *Journal of Mathematics and Statistical Science*, vol 4, April.

6 This data and metadata standard has already proved beneficial for dealing with the transmission of macro information, and the latest SDMX 3.0 version comprises new features that support micro data; see IFC (2016): "Central banks' use of the SDMX standard", *IFC Report*, no 4.

7 According to the terminology of the Balance of Payments and International Investment Position Compilation Guide, the micro database that stores statistics at an individual equity and/or debt security level is known as the (securities) reference database.

8 See Annex 4 of the Handbook on Securities Statistics (HSS) for a discussion on the attributes of statistics stored in a SBSDB.

fact, the appropriate sectorising of economic institutional units is fundamental not only to "zoom out" from micro to macro (that is to derive ad hoc and more precise macroaggregates from individual observations), but also to "zoom in" from macro to micro, that is to "drill down" from macro-level phenomena to get more micro-level insights (cf Section 3). In practice, the choice of an appropriate sectoral classification system may depend on the types of statistical compilation exercises supported by the SBSDB. Yet, there are obviously merits to follow a consistent and internationally recognised standardised approach; for instance by adhering to the 2008 System of National Accounts (2008 SNA) when deciding to attribute specific codes to the sectors and sub-sectors of interest. A telling example is the Centralised Securities Database (CSDB)[9] developed for the European System of Central Banks (ESCB), whose classification is based on the European System of Accounts[10] – with the objective of ensuring consistency across the different statistical compilation exercises as well as with various prudential and regulatory reporting tasks conducted in involved jurisdictions. Similarly, the international securities reporting exercises launched in the context of the various DGI recommendations aims at ensuring compatibility with SNA sectorisation, even when their different scopes may require different types of classification, for instance, to cover fintech activities[11] or environmental aspects (cf Box 1).

Thirdly, **the ways to identify individual securities can vary significantly**. The choice of a unique identifier is typically influenced by national specificities, for instance with the use of the Stock Exchange Daily Official List (SEDOL) for London Stock Exchange-based securities or the Committee on Uniform Securities Identification Procedures (CUSIP) for facilitating the clearing and settlement of trades of North American financial securities. However, central banks are also actively advocating the use of international standards with unique and unequivocal identifiers, like the ISIN for individual securities and the LEI for issuers.[12] The objective is to ensure transparency and standardisation in the information collected, avoid duplication, support data integration and aggregation for statistical compilation; an important consideration is also to facilitate the linking of different databases as much as possible (cf Section 3.1).

A fourth important aspect is the possibility to **complement security-level reference database with holders' information**. The latter is collected in a so-called securities holding database – which typically contains information on the holder's residency and institutional sector/subsector and on the amount of securities held. These details provide valuable insights on a wide range of areas that are of key interest to central banks, including the compilation of external statistics as well as analyses supporting monetary and financial stability policies (see Section 3). One key example relates to the GFC, which emphasised the importance for public authorities to take into account information on securities' holders to analyse the potential impact of an issuer bankruptcy. To this end, the European Securities Holdings Statistics Database (SHSDB) was set up to present granular data on holdings of securities (equity, debt securities and investment fund shares) as a complement to the ESCB's CSDB. Other examples include the security-by-security system of the Bank of Thailand,[13] which allows the reference

[9] The CSDB contains data on instruments, issuers and prices for debt securities, equity instruments and investment fund shares issued worldwide, where each security is identified by its ISIN code. For more details on the ESCB's CSDB, see ECB (2010): "The "Centralised Securities Database" in brief", February; and Pérez, AC and J Huerga (2016): "The Centralised Securities Database (CSDB)-Standardised micro data for financial stability purposes", *IFC Bulletin*, no 41.

[10] The so-called ESA-2010 framework is consistent with the 2008 SNA. The data structure definition (DSD) used in this classification context can be found at https://sdmxcentral.imf.org/sdmx/v2/structure/datastructure/ESTAT/NA_SEC/1.15.

[11] For instance, one may want to collect information on fintech entities at a very detailed sectoral level that is not covered in the SNA framework; see IFC (2020): "Towards monitoring financial innovation in central bank statistics", *IFC Report*, no 12.

[12] See Legal Entity Identifier Regulatory Oversight Committee (LEIROC) (2016): "Collecting data on direct and ultimate parents of legal entities in the Global LEI System – Phase 1", March.

[13] Cf Weeraya, W (2014): "Coping with challenges in establishing security-by-security database in Thailand", *IFC Bulletin*, no 37.

database providing security-level information on transactions and outstanding positions to be linked to a securities holding database (using a unique identifier).

Table 1 provides an overview of the scope of a SBSDB (including its coverage and reporting features) based on the related experience of central banks.

General Scope of a SBSDB

Table 1

Coverage	Type of securities	Equity (stocks, other equity investments), debt (bonds, asset-backed securities, commercial papers) and other securities (investment fund shares, equity-linked securities, derivative-linked securities)
	Type of market	Primary and secondary
	Security events	Liquidation, stock splits and reverse stock splits, takeovers, mergers etc
	Valuation principles	Face, market and nominal
	Type of variables	Stock, redemption, revaluations, change in volume, gross and net issues*
	Unique identifier	eg ISIN, CUSIP etc (for securities); LEI (for entities)
	Residence	Residents, non-residents (issued in FX/domestic currency)
	Other breakdowns	Currency of denomination, maturity, ratings, instrument type, interest rate type, additional breakdowns of interest (eg green instruments)
Reporting	Sector of issuer	Generally includes institutional sectors and sub-sectors, as for instance those defined in the 2008 SNA (non-financial corporations, financial corporations, general government; households, non-profit institutions serving households, rest of the world); customised approach also possible
	Holding	Entity level, bank group level, sector level
	Data providers and other stakeholders involved	Central banks, issuers (eg government, non-financial companies), holders, custodians (for holdings), financial institutions (eg mutual funds, brokers, pension funds, insurance etc), commercial data providers, central securities depositories, government agencies, securities exchanges

(*): Typically includes all the information necessary for stock-flow reconciliation exercises.

Source: IFC internal survey on SBSDB (2023).

Box 1

Box 1: SBSDB and statistics on green finance

Establishing a **SBSDB can be instrumental for covering data needs related to green finance** and addressing the associated analytical exercises raised by the impact of climate change. The demand for statistics related to green finance, which can be broadly defined as the financing/refinancing (notably through bonds, equities and investment funds) of activities that foster environmental sustainability, has increased markedly since the 2015 Paris Agreement. Central banks and government agencies embarking on the journey to develop a SBSDB have thus an interest to broaden the scope of their project to cover the specificities related to green securities – and even to the broader spectrum of environmental, social, and governance (ESG) frameworks.

Yet, a **main issue relates to the way (the so-called "taxonomy") one should identify, describe, and classify the various financial instruments involved in sustainable finance**. The definition of what is "green" can depend on national specificities and/or preferences, and each jurisdiction may have different taxonomies. Work has already been initiated in international fora to address existing gaps, make a bridge between the different taxonomies that already exist, and reflect on the potential characteristics a global taxonomy could have. One important objective is to ensure consistencies between the various statistical frameworks being developed, especially as regards the ongoing elaboration of the new 2025 SNA and the new Balance of Payments Manual (BPM-7) as well as the implementation of the recommendations of the G20 DGI.①

In particular, it is envisaged **in the 2025 SNA to add the two "*of which: ESG*" and "*of which: green*" breakdowns to the four financial instrument classes** – ie debt, loans, equity and investment fund shares – in the financial accounts and balance sheets sections. It is proposed in this context to define ESG finance as finance for "*activities or projects that sustain or improve the condition of the environment or society or governance practices*" and green finance as finance for "*activities or projects that sustain or improve the condition of the environment*".②

In parallel, the **workplan of the third phase of the DGI includes a recommendation on climate finance** (no 4) led by the international Working Group on Securities Databases (WGSD)③ to promote statistical compilation of green debt securities and green listed shares. G20 economies have been invited in this context "*to provide (preferably on a quarterly frequency) experimental data on issuances and holdings of green debt securities and listed shares securities to the BIS, based on self-commitments and consistent as much as possible with the Handbook on Securities Statistics (HSS), and any other specific guidance developed in the context of the recommendation, and provide documentation on the compilation methodology*".

A first focus of this DGI recommendation has been on debt securities, with three main types identified: green bonds, sustainability bonds and sustainability-linked bonds.④ Green bonds are accordingly defined as "*debt securities whose proceeds are used to fund projects intended to deliver a positive environmental impact*". Sustainability bonds would be "*debt securities whose proceeds are used to fund projects intended to deliver a combination of positive environmental and social impact*". Sustainability-linked bonds relate to "*debt securities whose characteristics (eg coupon payments) can vary depending on whether the issuer achieves predefined environmental or other sustainability objectives*".⑤ Further work is planned on the definition of other financial instruments, especially green equity.

In addition to the work on definitions, DGI recommendation no 4 also aims to **develop reporting templates** to record stocks on the different types of green debt securities (and other financial instruments at a later stage). This reporting will be in the form of "*of which*" breakdowns that are consistent with the overall reporting of global debt securities (green and non-green) established in the G20 context – especially related to the recommendation no 7 of the second DGI phase; cf footnote 3 above. The reporting targets include a basic intermediate target for core data on green bonds by end-2025 and a final target for other data by the end of the DGI-3 phase (at end-2027). Moreover, interested countries outside of the G20 DGI framework have been associated with these discussions and would benefit from following similar definitions, should they wish to include climate finance securities under the scope of their SBSDB.

Lastly, there are several **other private and public international initiatives** related to the development of definitions, standards and reporting frameworks for green finance. These include the guidance published by (i) the International Capital Market Association (Guidance on sustainable financial instruments), (ii) the International Financial Reporting Standards (IFRS) (Initiative to promote global climate disclosure standards), (iii) the Climate Bonds Initiative (Climate bonds standard), (iv) the World Federation of Exchanges (Green equity principles), and (v) the European Commission (European green bond standard).⑥ In the absence of public / official standards in a given jurisdiction, this material as well as other privately-developed standards can be useful inputs to consider when starting to develop a SBSDB that aims to cover climate-related data needs.

① For a discussion on the complementarity of SNA, BPM and DGI-3 on green finance, please refer to SNA/M1.25/02 - Issues Note: Sustainable finance definitions in the 2025 SNA and Balance of Payments Manual 7.

② See the SNA/M1.23/14 Guidance note elaborated by the SNA 2025 Wellbeing and Sustainability Task Team on Environmental Classifications in 2023.

③ See Details and recommendations of the WGSD to improve the global availability of information on securities markets.

④ See Explanatory notes on the reporting templates for recommendation no 4 of the DGI-3.

⑤ As a result, sustainability debt securities and sustainability-linked debt securities are part of the broader ESG debt securities aggregate contemplated under the 2025 SNA.

⑥ See, respectively, the Principles, Guidelines and Handbooks; the IFRS International Sustainability Standards Board (ISSB); the Climate Bonds Standard V4.0; the Green Equity Principles; and how the European green bond standard are supporting the transition.

3. Benefits of a SBSDB

A SBSDB offers a wide range of analytical opportunities, not least because it allows the reuse of data for multiple needs spanning across several statistical and non-statistical domains. In addition, it provides a tool for ensuring a consistent reconciliation between micro and macro data, for instance to assess economic relationships as well as interconnectedness and spillover effects – issues that are of key importance for central banks in pursuing their public mandates.

3.1 Supporting statistical compilation

The development of a SBSDB can bring important benefits for the compilation of statistics. According to the experiences reported by central banks, these benefits include improving data quality; enhancing statistical agility, for instance to have more flexibility in deriving *customised/non-conventional* statistical aggregates to meet specific analytical needs; reducing burden on reporting entities, as some tasks can be more efficiently performed by data compilers; etc.

The starting point is that, **instead of reporting pre-defined aggregates and detailed breakdowns, reporting entities are asked to report individual records** (using a unique identifier) that can be stored in the SBSDB. Central banks and government agencies as data compilers are then responsible for the statistical classification of this information in a standardised and harmonised way. In turn, this approach can foster higher accuracy and consistency in the treatment of security-level data, in addition to the benefits of adhering to international standards (eg in terms of methodological soundness and comparability). This is in sharp contrast with traditional reporting practices, whereby different reporting entities use their own aggregation procedures to produce the information requested by data compilers, in a way that can be inconsistent and heterogeneous (potentially leading to miscalculations, misclassifications and inaccuracies). One telling example relates to the important area of BoP/IIP statistics: the use of a SBSDB allows for computing almost all of the required breakdowns in a comprehensive way, and this can prove particularly helpful for the compilation of the portfolio investment category (cf Box 2).[14] It also helps to mitigate the risk of double-counting and misclassification across BoP categories as well as to fill data gaps (eg for estimating "hidden securities assets").[15]

A second important point is the **greater flexibility for data compilers**, in terms of the level of details available (possibility to "zoom-in/zoom-out" the information), consistency between micro-level details and derived macroaggregates, and timeliness – as the "traditional" collection of pre-defined aggregates from data reporters may take more time compared to the direct receiving of individual records.[16] Moreover, once a SBSDB is established, one can more easily access the required information in case of new requests, instead of having to set up ad hoc surveys that may require important additional resources for their design, launch and actual reporting.

Box 2

Box 2: SBSDB as a supporting tool for compiling BoP/IIP statistics

An important benefit of SBSDBs for central banks and related government agencies is in the compilation of BoP/IIP statistics.① Securities statistics can indeed be critical for compiling major investment categories, such as foreign direct investments and portfolio investments.

In this regard, **SBSDBs can be beneficial for several reasons**. First, security-level data are particularly helpful to derive more granular breakdowns of the portfolio investment category, on top of the traditional ones (see Table 2). Second, they help reducing net errors and omissions, for example those that can arise because of mismatches between issuances and holdings data or of the inaccurate treatment of investment fund income and fees. Third, they enable to complement traditional sources, for example by providing information on foreign assets – which is typically missing from traditional external statistics data collection due to the difficulty to get data from foreign custodians (so-called "blind spot")② Fourth, data will be available with a higher frequency: this can in theory facilitate the compilation process and improve timeliness. Fifth, the information may be obtained from resident custodians (report positions in securities) and resident end-investors (own account on behalf of other residents).

Yet, **SBSDBs also present some limitations**. First, the reconciliation of different data sources is typically a very complex task in external sector statistics. The reason is that these usually depend on a wide array of data sources including both macro and micro data sets, including surveys – such as the Coordinated Direct Investment Survey (CDIS) and Coordinated Portfolio Investment Survey (CPIS) of the International Monetary Fund (IMF), or BIS data collections – and administrative sources (especially for the general government sector). This can prevent the perfect alignment between micro and macro-level data. Second, imperfections in the granular information available in the SBSDB can lead to notable inconsistencies between aggregates, depending on the statistical exercises considered (BoP, financial accounts, etc). And third, methodology and compilation processes can be intrinsically different between micro and macro sources, for instance as regards the inferring of transactions from changes in stocks (or the other way around).

As an illustration, the BoP/IIP attributes that can be usefully compiled using a SBSDB are provided in the following Table 2.

14 See a discussion on the possible data sources for compiling an IIP Statement in the BPM6 Compilation Guide.

15 See Gervais, E and P Bui Quang (2019), "How to identify "hidden securities assets" in the Balance of Payments: methods of Bank of France", *IFC Bulletin*, no 49.

16 One reported example is the possibility to extract granular information from the European CSDB to compute statistics on the rollover of government debt securities within one month, while similar information from aggregated sources would be available only after four months; see Perrella, A, and J Catz (2020): "Integrating microdata for policy needs: the ESCB experience", *ECB Statistics Paper Series*, no 33.

SBSDB as a source for compiling portfolio investment in IIP statements — Table 2

Portfolio investment	Assets	Liabilities
Equity and investment funds shares	X	X
Central Bank	X	X
Deposit-taking corporations, except the central bank	✓	✓
General government	✓	X
Other sectors		
Other financial corporations	✓	✓
Nonfinancial corporations, households and NPISHs	✓	✓
Debt securities	X	X
Central Bank	X	X
Deposit-taking corporations, except the central bank	✓	✓
General government	✓	✓
Other sectors		
Other financial corporations	✓	✓
Nonfinancial corporations, households and NPISHs	✓	✓

Note: The attributes that can be compiled using a SBSDB have been marked with ✓ (otherwise with X).
Source: Authors' illustration adapted from the IMF's *BPM6 Compilation Guide*.

① For a discussion on the possible data sources for compiling an IIP statement, please refer to IMF (2014): "BPM6 Compilation Guide".

② See Zucman, G (2013): "The missing wealth of nations: are Europe and the US net debtors or net creditors?", *Quarterly Journal of Economics*, pp 1321-64.

A third and important aspect is the **possibility of linking different databases**, particularly with credit and business registries. This information can be an important input supporting the compilation of key statistical exercises, such as statistics on MFIs and IFs; financial accounts (especially for the measurement of instruments like shares and securities); BoP/IIP statistics; international collections on debt securities such as those organised by the BIS; and new data collections on green finance, for instance in the DGI context. Further, and as mentioned above, the SBSDB reference database can be usefully combined with the securities holdings database, not least to facilitate the compilation of from-whom-to-whom (FWTW) SNA tables.[17]

Lastly, experience shows that the **development of a SBSDB provides a wealth of information than can be tapped for new statistical purposes that were not fully intended** at the first step;[18] in

[17] This can provide greater insights on cross relationships between issuers and holders of financial instruments, in turn supporting the assessment of the different underlying economic dynamics and relationships.

[18] For instance, important items of the European Macroeconomic Imbalance Procedure (which provides a framework for the coordination of economic policies), such as the compilation of headline indicators on consolidated credit flow and debt for

other words, the supply of new information can lead to new demand.[19] A SBSDB can also be useful for estimating valuations effects that are often difficult to capture. For instance, it can be used to measure revaluations and other changes in the volume of assets and liabilities by type of financial instruments. This can be very useful to derive information from high-frequency valuations measures on an accrual basis to operate stock-flow reconciliations or specific valuation exercises such as for nominal values (whose calculation often requires accurate information on accrued interests, coupon payments, etc which are often estimated at the level of individual securities.).[20]

3.2. Non-statistical uses

While the primary purpose for developing a SBSDB is typically to enhance the compilation of official statistics, there are also benefits for a number of non-statistical tasks performed by central banks. A key area is **financial stability**. From a **micro prudential** perspective, a SBSDB can provide valuable and consistent quantitative or qualitative information on the risk profile of individual institutions, which is an important input for stress-testing exercises. From a **macro prudential** perspective, information from a SBSDB can be used to measure systemic risks, interconnectedness and contagion channels in the financial system; for instance, when monitoring the growing role of non-banks in financing the economy.

More generally, a SBSDB can contribute to a **wide range of policy-relevant analytical works**, for instance, the production of financial stability reports, monetary and fiscal reviews as well as macroeconomic projections. Such analyses can rely on concrete evidence to answer questions such as "*Who is holding what?*", "*What is the total exposure of a given economic sector to specific classes of financial instruments?*", or "*What are the fiscal implications of financial connections with non-residents?*". One key aspect is that, being highly granular, SBSDBs offer the necessary flexibility and agility to meet changing and growing users' analytical needs.

A last area is **research**. The detailed data collected in a SBSDB can allow for a thorough analysis of new developments and trends in the financial markets. Granular information on securities' holdings is in particular being increasingly used in three research fields of key importance in central banking: (i) banking and finance literature (eg estimation of interconnectedness and contagion channels), (ii) international investment literature (eg impact of asset purchase programmes associated with quantitative easing policies,[21] development of granular portfolio insights), and (iii) monetary policy research (eg policy effectiveness and transmission channels).[22]

the private sector as well as households, largely benefit from security issues statistics; cf ESS-ESCB (2023): "Quality assessment report on statistics underlying the Macroeconomic Imbalance Procedure", July.

[19] One telling example relates to fiscal policy: for instance, the use of the ESCB's CSDB has facilitated the construction of new indicators on government securities, providing insights on the future evolution of the public debt service (encompassing both principal and interest payments) as well as on the interest rates committed to debt holders; see Pérez, A, J Diaz and D Lojsch (2015): "New and timely statistical indicators on government debt securities", *ECB Statistical Paper Series*, no 8.

[20] Cf Box 5.3 of the HSS (pp 42-3): "Debtor Approach and Creditor Approach to Recording Accrued Interest".

[21] See Koijen RS, F Koulischer, B Nguyen and M Yogo (2017): "Euro-Area Quantitative Easing and Portfolio Rebalancing" *American Economic Review*, vol 107, no 5, May.

[22] See, in the case of the ECB, Boermans, M (2022): "A literature review of securities holdings statistics research and a practitioner's guide", *De Nederlandsche Bank Working Paper*, no 757.

4. Challenges

The setting up and maintenance of the SBSDB is not devoid of **challenges**. The main reported ones cover: (i) the practical difficulties in compiling such a large and complex database, (ii) the associated costs in terms of development and maintenance, (iii) the need for adequate skilled resources to support data compilers, and (iv) the implementation of a legal framework.

- First of all, statisticians implementing a SBSDB face significant **challenges related to the practical handling of such a large and complex database**. The rich information available also implies that there can be multiple inconsistencies in terms of reporting and scope, for instance between the information collected from custodians versus firms.
- Turning to **costs aspects**, an important concern is that acquiring granular information from commercial database providers is often a must and can be (very) expensive. IT costs, related in particular to purchasing/developing the necessary hardware and software for data storage and processing, may also be significant. This is the case not only to set up the SBSDB, but also to maintain it: important investments are required to develop adequate quality modules to limit manual intervention, automate data validations, and take corrective actions on an ongoing basis. Hence, while in theory the direct availability of granular information through a SBSDB will improve statistical timeliness, the generally important resources needed for getting and processing the new data may lead to important delays in practice.
- In addition, **significant expertise and data literacy** are required to support a SBSDB. From the **producer** perspective, the compilation of raw granular data is often highly complex, notably to ensure the correct identification of the reported securities, the classification of their attributes according to statistical standards, and consistent data versioning. From the **end-user** perspective, data are typically disseminated on a "as is" basis and often require sound knowledge of statistical information models and advanced analytical capabilities to perform correct business aggregations.
- Lastly, as regards **legal framework issues**, there may be obstacles preventing the exchange of granular-level data between central banks, statistical agencies or other authorities. These limitations need to be correctly recognised and addressed, for instance by setting up adequate memorandum of understanding (MoU) with appropriate restrictions, particularly for handling and protecting confidential information or market sensitive data. However, when using commercial data, there may be licensing restrictions that prevent data sharing. Of key interest for central banks is therefore the exchange of best practices to facilitate data sharing and the set-up of appropriate safeguards.[23]

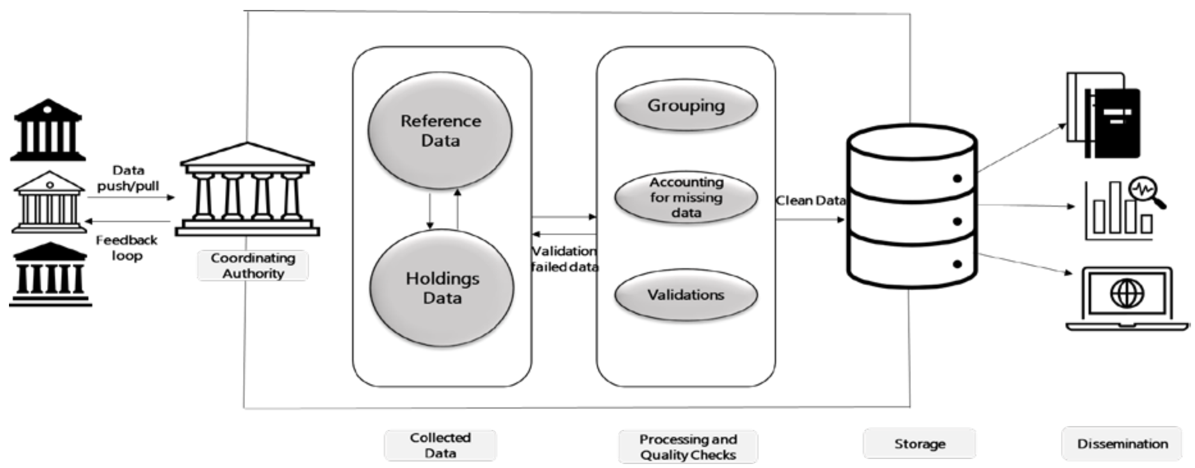
5. End-to-end statistical process flow supporting a SBSDB

The development of a SBSDB includes **various phases to cover the main stages of the data lifecycle**, which typically comprise their collection, their processing and quality management including editing (ie review of data consistency, detection of errors and outliers and correction), and their dissemination (cf

[23] Cf IFC (2023): "Data-sharing practices", *IFC Guidance Note*, no 3. It should be noted that the sharing of content can take different forms depending on the degree of confidentiality / legal restrictions. For instance, one could aggregate analyses or anonymise granular information. Another (potentially technically complex) possibility is to have a distributed SBSDB, where the microdata would reside in the collector jurisdictions and queries to aggregate content could be run locally and pushed to a central aggregator.

Graph 1).

Typical end-to-end pipeline for a SBSDB statistical business process — Graph 1



Note: This illustration relates to the network approach (ie with a single coordinating agency in charge of gathering information transferred from its partners to set up the SBSDB).

Source: Authors' illustration based on IFC internal survey on SBSDB (2023) and available literature describing the experience of different NSSs.

The starting point is the **identification of users' requirements and the available information** to meet their needs. This requires mapping (i) the existing data sources and the securities' attributes they cover, (ii) the level of investor details (investor-by-investor or sectoral level for covering information on holdings), and (iii) the coverage of the data in terms of sector, instrument and geography (for instance, split between resident and non-resident issuers and holders when the SBSDB includes information on holdings). In practice, the approach can be progressive, starting first with a minimal data set based on the data provided by reporting agents, before augmenting it with other available information for the compilation of more accurate statistical aggregates and meeting new data demands.

Once the scope of the targeted SBSDB has been defined, there is merit in **following a structured approach for setting it up**. This calls for developing a comprehensive end-to-end process flow to obtain a complete and efficient functional solution supporting statistical production. The approach comprises, firstly, a thorough review of the requirements, followed by the design, set-up and maintenance of a technical solution that integrates the different processes involved when extracting, transforming and loading the data in the final repository. The four main stages to be considered in this context are data collection; data processing; data quality management including data editing; and data dissemination.

5.1 Data collection

Experience shows that there are mainly **three different approaches** when collecting the data feeding into a SBSDB, namely: (i) clearing house approach, (ii) centralised approach, and (iii) network approach.[24] The **first approach** relies on the centralisation of the reported information by a clearing institution, which has the responsibility of ensuring the quality of the data; in this approach, however, the central institution may take significant time to achieve satisfactory data quality, and this can also lead to the collection of inconsistent and/or not harmonised data. In the **second, centralised approach**, a single organisation acquires all pertinent data (including from commercial providers), conducts quality controls, harmonises the data, and subsequently grants access to authorised users; this approach may be quite appealing for data users but can represent an important burden for the central agency, especially if it is tasked to perform all data quality controls. In the **third, network approach**, a coordinating agency will gather the information transferred from its partners (eg domestic statistical agencies, or national statistical agencies/central banks in the case of an international project as led by the ECB in the European context) in a bottom-up way. Effective coordination can be achieved by setting up bilateral contracts with participating organisations, not least to clarify the respective responsibilities for monitoring the quality and coverage of the data collected. This bottom-up compilation will then be complemented by data from commercial sources in a top-down way, with the objective to cover missing information and to set up the most complete and consistent database.

In any case, a **SBSDB is fundamentally multi-mode and multi-source**: the data are collected through push (by indirectly pushing the code information to the data) or pull access methods (by accessing it directly)[25] from various entities, including central banks, government agencies, international sources, commercial data providers, clearing houses, securities exchanges. This collection can be done in an automated fashion, typically through a web-user interface, but the use of unique identifiers is crucial. In their absence, non-standardised identifiers (names, residence etc) may be used by deploying suitable name-matching algorithms to avoid the risk of double-counting.[26] Another important consideration is that data should be collected from reporting entities in agreed fixed, defined file formats (eg CSV, XML, SDMX). This choice can have important implications. In particular, schema-based files (for instance, based on the SDMX standard) should ideally be preferred as their pre-defined structure allows to implement automatic, metadata-driven checks, in contrast to schema-free files.

5.2 Data processing

Once collected, **the processing of security-level data involves several steps**. **The starting point is to correctly label the granular information collected**, which requires the precise and unique identification of the financial instruments and entities (issuers, and holders when the SBSDB includes information on holdings). In practice, the instrument identifier (eg ISIN) guarantees the unique identification of the securities, while the entity identifier (eg LEI) allows entities to be identified at the issuer and parent level. Apart from the importance for organising the SBSDB, the use of (common) identifiers is also a key element for an effective matching of the various micro databases involving

[24] See Israël, J-M (2002): "The Centralised Securities Database", *IFC Bulletin*, no 13.

[25] See the discussion on so-called "distributed protocols" in OECD Statistics and Data Directorate (2018): "Which Strategies for NSOs in the Digital Era? Towards 'Smart Data' Strategies", SDD/CSSP(2018)7, August 2018.

[26] A popular name-matching algorithm is the string metric based on *Jaro-Winkler distance*; others include *Levenshtein distance* and *n-grams*. For a detailed discussion on the steps involved in deploying such an algorithm in the context of the ESCB's CSDB, please see Pérez, A , J Huerga, F Mayerlen and J Micheler (2016): "Unique identifiers in micro-data management - the Centralised Securities Database (CSDB) experience", *IFC Bulletin*, no 43. This paper also discusses the grouping algorithm - which creates a unique representation of a single entity and may be deployed when the security attributes for the same issuer identifier are reported by multiple data sources to avoid inconsistencies.

different stakeholders (eg reporting agents, creditors, debtors, securities issuers, investors, protection providers, etc) and instruments (eg securities, loans, derivatives, etc). It is also key to deal with information that is often spread across national borders (eg holding of domestic securities by foreign investors). Such linking exercises are essential to form comprehensive analyses on financial developments and associated risk exposures.[27]

Data processing and validation are the next steps to obtain a "clean" database. They entail three sub processes: (i) accounting for missing data, (ii) data revision due to validation failure, and (iii) data revision due to the receipt of new information.

Firstly, **missing data** can be easily spotted through automated checks. Several approaches can be used, depending on the type of data considered.[28] The first and easiest case is when the data source is **competitive** (ie multiple sources can provide information for the same record): as a result, alternative information is selected to remove inconsistencies and fill the gaps. A second, opposite scenario is when the source is **monopolistic** (ie reliable data can be obtained only exclusively). In this case, there is a need to set up a specific methodology to fill missing information, typically by leveraging the other attributes available for the specific security considered. Appropriate aggregate indicators, as observed for a given type of instrument by country, currency, industry or sector, may also be used for accounting for missing data. For instance, in the absence of price information for a given debt security, the ECB estimates it by considering all its reference attributes available, such as coupon information, currency of denomination, residual maturity, and credit standing of the issuer; as for equity, one can develop a pricing model based on individual past rates of return and the current values of market indices to infer missing values. Moreover, information from similar securities may also be used to approximate missing data. However, there may be scenarios when there are no observed information and proper estimations cannot be conducted – in such cases, the missing attribute may simply be substituted by default values defined by the database manager on an ad hoc basis.

The second sub process is to **correct the information collected in case of validation failures**. This is crucial to ensure the highest quality possible in the compilation process,[29] in the spirit of the Fundamental Principles of Official Statistics.[30] In practice, as noted in the *Handbook on Securities Statistics*, the process consists of automated and meticulous controls ("validation checks") incorporated in the data quality management system to identify outliers within specific statistical categories. Outliers may not only represent simple reporting mistakes, but also misclassifications that require further investigation and possible adjustments to the statistical exercises.[31] This detection can be done in multiple ways. One is to compare the received information with existing or past data; in case of large deviations, the data processing may be stopped with a feedback mechanism to go back to the reporting entity and ask for revised input. Some other basic controls comprise the comparison of total reported

---

[27] See a discussion of the benefits of LEI for monitoring systemic risk at Laurent, F, B Kaminska, C Aparicio, B Dworak, L Garcia, D Durant and F Pistelli (2021): "The benefits of the Legal Entity Identifier for monitoring systemic risk", *ESRB Occasional Paper Series*, no 18.

[28] For a discussion, please see Kleibl, J, and J Micheler (2022): "Data quality management frameworks for large multi-source micro databases", *10th European Conference on Quality in Official Statistics*, Vilnius, Lithuania.

[29] See Križman, I and B Tissot (2022): "Data governance frameworks for official statistics and the integration of alternative sources", *Statistical Journal of the IAOS*, vol 38, no 3, pp 947–55.

[30] Cf United Nations (UN) (2013): "Fundamental Principles of Official Statistics", resolution adopted by the Economic and Social Council, E/RES/2013/21, 28 October.

[31] As a recent important example faced by central banks, one may want to adjust aggregation categories to better take into account the impact of fintech on financial instruments and/or entities; see IFC (2020): "Towards monitoring financial innovation in central bank statistics", *IFC Report*, no 12.

positions with outstanding amounts (for debt securities) and market capitalisation (for equity securities), as well as with the previously reported positions and transactions (stock/flows analyses). Compilers may also try to compare their input data with the statistics collected by other agencies and authorities based on availability.

Turning to the third step, one needs to deal with the fact that **data can be subject to important and continuous revisions**, for example to integrate new information. The answer is usually to follow the same approach as during the validation phase. As for the dissemination phase, however, the revised data should ideally be accompanied by transparent explanatory notes framed in line with the broader revision policy set up by the compiling organisation. However, a key issue here relates to feasibility: it may be almost impossible in practice to provide a full account of the revisions in the case of such a large and complex micro database. The way to go is thus typically to focus only on the granular revisions that are above a certain magnitude and/or on their overall impact on the macro-level aggregates of interest. Lastly, in the case of routine revisions, the provisional nature of data may be notified to users in a systematic way.

5.3 Data editing and quality management

The challenges posed by the application of various quality checks on the vast amount of granular data points with multiple attributes stored in a SBSDB **puts a premium on setting up an overall data quality management (DQM) framework**[32] **for handling micro data at the organisation level**. For instance, an interactive DQM platform (iDQM) manages the ESCB's CSDB data, covering all the stages of its production chain, from input to output.

Ideally, a DQM framework for operating and assuring the quality of large multi-source micro data systems **should comprise the following elements**:

1. **Responsibilities** of the reporters and reporting entities;
2. Level at which **quality checks** will be applied: checks may be conducted on input versus output data, on individual records versus derived aggregates, etc;
3. **Attributes** on which quality checks will be applied;
4. **Distinction between attributes** for internal vetting versus attributes on which external competent agencies need to be involved;
5. **Frequency** of the quality checks to be conducted;
6. Frequency of data submission, as well as **deadlines** set-up to ensure punctuality and timeliness;
7. **Types of statistical checks** to be conducted. In this context, an increasing number of central banks and government agencies have been deploying advanced statistical techniques to improve their DQM system; this experience shows that tools based on artificial intelligence (AI) / machine learning (ML) can be usefully leveraged to improve data quality.[33]

---

[32] Many central banks in the euro area follow the Guideline of the European Central Bank of 19 May 2022 on the Centralised Securities Database and the production of securities issues statistics and repealing Guideline ECB/2012/21 and Guideline (EU) 2021/834 (ECB/2022/25).

[33] See, for a general overview, Araujo, D, S Doerr, L Gambacorta and B Tissot (2024): "Artificial intelligence in central banking", *BIS Bulletin*, no 84; and, for a specific application, Gottron, T, G Kanellos, and J Micheler (2022): "Introducing explainable

In practice, the development and operation of a DQM framework to deal with large-scale multi-source micro data systems can benefit from following a structured approach. As argued by Kleibl and Micheler,[34] the following **five core principles can be usefully considered** in this context:

1. **Vertical integration**: all technical and operational components involved in data processing should be integrated and owned by a single institution.
2. **Finalisation before formalisation**: the formalisation of the DQM framework should follow its technical and practical implementation, as many potential operational and technical issues cannot be well anticipated but will only materialise during the development process.
3. **Unified data processing and performance-driven data modelling**: the set-up of a standardised data processing should be accompanied by a data model tuned to maximise performance of the respective compilation steps.
4. **Integrated and prioritised DQM workflows**: such workflows allow for centrally identifying and resolving data quality issues in an efficient way, allowing data quality managers to invest their limited resources most effectively, especially in situations where it is not possible to address all raised data quality issues.
5. **Effective collaboration and clear ownership**: the framework should be based on effective collaboration among data quality managers and clear ownership of responsibilities.

5.4 Data dissemination

Once the processing and quality checks are completed, **a single data warehouse typically contains all the statistical products**, at both micro and more aggregated levels. At the micro-level, one should be able to access all the characteristics of interest for the securities, such as their yield to maturity, accrued interests, duration etc (if needed, this information can be derived by using standard algorithms). At the more aggregated level, the system should allow for producing customised as well as more ad hoc, non-conventional aggregates based on micro-level inputs. The standard practice is to keep a regular "golden copy" of the data warehouse, saved in a separate storage area.[35]

The next step is to **make (part of) the data warehouse information available to the relevant end-users**. However, before proceeding to dissemination, it is crucial to clarify its degree of confidentiality. If it is marked as public, it can be made available without any restrictions. But if a specific part is flagged as protected, then access will be limited, which is often the case for micro-level data that are confidential and/or market-sensitive. This information could however be made public after adequate aggregation (so that individual values cannot be identified), or by using anonymisation techniques. In any case, it is essential to have adequate confidentiality flags to apply appropriate access restrictions in the data warehouse and ensure that output data is disseminated only to authorised users.

supervised machine learning into interactive feedback loops for statistical production systems", *IFC Bulletin*, no 59.

[34] See footnote 28.

[35] It is however critical to carefully evaluate the pros and cons of establishing such a golden copy, as regards the implications in terms of size/storage, transmission mode, replaceability etc.

6. Data governance

The "quality" of a SBSDB calls for having a strong data governance in place. In practice, the concept of data governance refers to the various organisational features put in place for dealing with the data and cover all the related principles, policies and procedures, structures, roles and responsibilities.[36]

Based on experience, **a number of best practices** can be followed by central banks and government agencies to strengthen their data governance frameworks and support the end-to-end SBSDB data lifecycle. In particular, the following points stand out:

- Maintaining **data quality** across the entire data lifecycle is at the core of a SBSDB. For this, the conduct of regular audits to identify inconsistencies and take corrective actions is strongly recommended.
- As the development and the maintenance of a SBSDB involve engaging with multiple stakeholders, a **coordinating organisation** must be identified; this central agency will establish the required data flow channels and enter into a legal agreement with related parties.
- Developers of the SBSDB have underscored the importance of defining clear concepts and creating adequate **metadata** in a transparent way.[37] This puts a premium on setting up comprehensive data catalogues with information on all data inputs and related confidentiality labels.
- Due to nature of the work involved, **multiple teams** within the coordinating organisation need to be involved in the development and maintenance of the SBSDB; for instance, the departments in charge of IT, statistics or financial markets. Therefore, the proper identification of the different stakeholders, their roles and relationships is essential.
- In the event of **data revision**, adequate audit trails need to be maintained. This calls for proper documentation of the adjustments made and methodologies followed to deal with missing data before the revision.
- There are growing demands for developing **new, customised aggregates** based on micro-level SBSDB inputs. This needs to be examined thoroughly by the concerned parties, and the required changes should be considered in a calibrated fashion and with careful costs/benefit analysis. It also puts a premium on following sound change management processes.
- When the data are pulled from different sources, **proper legal frameworks** detailing the exchange processes need to be set up.
- There is value in having a specific module in the system that **monitors the timely reporting** of information to and from external entities and organising the sending out of automatic reminders.
- On the **administrative side**, data intensive projects like a SBSDB require continuous capacity building for managing the database, ensuring its effective utilisation and liaising with users.

---

[36] A more precise definition of data governance would be: "*A system of decision rights and accountabilities for the management of the availability, usability, integrity and security of the data and information to enable coherent implementation and co-ordination of data stewardship activities as well as increase the capacity (technical or otherwise) to better control the data value chain, and the resulting regulations, policies and frameworks that provide enforcement. This includes the systems within an enterprise, organisation or government that define who has authority and control over data assets and how those data assets may be used, as well as the people, processes, and technologies required to manage and protect data assets*"; cf UNECE (2024): "Data Stewardship and the Role of National Statistical Offices in the New Data Ecosystem", ECE/CES/STAT/2023/4.

[37] See de Matos, JC (2015): "Reaping the Benefits of Using Integrated Micro-Data for Statistical Purposes... and Beyond", *Statistika: Statistics & Economy Journal*, 95(3).