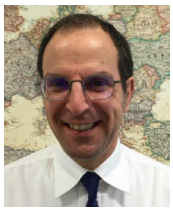
Incorporating microdata into macro policy decision-making

Received (in revised form): 9th September, 2021



Jean-Marc Israël
Former Head of Division, European Central Bank, Germany

Jean-Marc Israël worked for more than 21 years at the European Central Bank (ECB) as head of division. From 2017 to early 2020, he was Head of the Analytical Credit and Master Data (AMA) Division, also (co-)chairing the Working Group on AnaCredit and the Register of Institutions and Affiliates Data (RIAD) Hub Network, both reporting to the ESCB Statistics Committee. Earlier, for more than a decade, he headed the Monetary and Financial Statistics Division at the ECB and chaired the related Working Group, which also assists the ESCB Statistics Committee. The Division and Working Group dealt with the preparation and monitoring of legal texts and monitoring of relevant developments in the environment, data quality assurance and analysis of a wide range of monetary and financial statistics. These statistics, including those on banks' assets, liabilities or retail interest rates, on insurance and pension funds and on the money market statistical reporting, support economic, monetary and macroprudential policies and microprudential supervision. Earlier, Mr Israël was Head of the ECB Division on BoP and other external statistics. By the end of his career at the Banque de France, he had worked in four different departments: IT, External Affairs on Balance of Payments (BoP) statistics and as Adviser in the Monetary and Financial Statistics Directorate. He participated in, or chaired, various other working groups and task forces such as the BIS/Committee on the Global Financial System (CGFS) task force on credit risk transfers or the Contact Group on Data between the ECB, European Systemic Risk Board and European Supervisory Authorities. Mr Israël co-authored three ECB Occasional Papers and a paper on reporting schemes for banks in 'la Revue Banque' of March 2019.

E-mail: israljeanmarc.adv@gmail.com

Bruno Tissot
Head of Statistics and Research Support, Bank for International Settlements, Switzerland

Bruno Tissot is the Head of Statistics and Research Support at the BIS and Head of the Secretariat of the Irving Fisher Committee on Central Bank Statistics (IFC). He has been at the BIS since 2001, as Senior Economist and Secretary to the Markets Committee of Central Banks in the Monetary and Economic Department and then as the Adviser to the General Manager and Secretary to the BIS Executive Committee. Between 1994 and 2001 he worked for the French Ministry of Finance. He is currently Head of BIS Statistics and Research Support, having graduated from École Polytechnique (Paris) and the French Statistical Office, INSEE.

Bank for International Settlements, Centralbahnplatz 2, CH-4002 Basel, Switzerland
Tel: 0041 61 280 8294; E-mail: bruno.tissot@bis.org

Abstract The ongoing expansion of granular data sets calls for broadening official statistical frameworks to benefit from this information in order to assist the assembling of macroeconomic aggregates and/or facilitate the linkage between micro- and macro-level statistics. It can also provide important analytical benefits and effectively support central bank policies, with a greater ability to 'zoom in' on particular areas of interest and assess the distribution of macroeconomic aggregates within reporting populations. There are, however, important challenges associated with dealing with these data sets, for instance with regard to their quality, confidentiality and the manner of accessing them. Indeed, the task of integrating (granular and) microfinancial information in macro

frameworks has proved more complex than initially thought, and many challenges have yet to be addressed. This paper discusses an important issue: how to make use of granular information from private sources which are not part of the official statistics framework.

KEYWORDS: official statistics, granular data, policymaking, central banking, data integration

INTRODUCTION

Increasingly vast and complex amounts of granular data is becoming available to public authorities. This is particularly the case for central banks in their role as statistics compilers, a function that has clearly expanded in recent decades, reflecting the growing importance of finance in the economy (see[1] for the US situation). Another trigger was the launch of various and important microdata collections, especially in the context of the G20-endorsed Data Gaps Initiative (DGI) in response to the great financial crisis (GFC). In particular, the first phase of the DGI (2009–2015) stressed the importance of complementing macro-level aggregates with distributional information (cf [2], Recommendation no 15). The second phase (2016–2021) called for collecting more granular data to 'help straddle the divide between micro and macro analysis'.[3] Such an emphasis on the importance of microdata has been relatively new in the area of official statistics. For instance, the last version of the Manual for the System of National Accounts (SNA) devotes only a two-page section (out of almost 700 pages) to 'The use of microdata for macroeconomic accounting', stating that 'there would be considerable analytical advantages in having micro databases that are fully compatible with the corresponding macroeconomic accounts' but that this task 'may be difficult, if not impossible, to achieve'.[4]

Today, the importance of micro and granular information is fully recognised, not only by those market supervisors tasked with monitoring individual institutions, but also, more broadly, by all the public authorities in charge of macro policies in the economic and financial sphere. Yet microdata and granular information relate to two close but different concepts — an important distinction as the kinds of analyses will differ depending on the type of layer (micro or granular) they are run on.

As regards microdata, the traditional approach is to consider that it is 'an observation data collected on an individual object — statistical unit'.[5] In the area of economics and finance, this will comprise

> data on individual reporting units or specific transactions/instruments, which in most cases allow the identification of individual entities and are therefore considered confidential. In addition, publicly available data on individual reporting units are considered non-confidential although they can still be subject to data-sharing limitations due to commercial property rights.[6]

Financial microdata will, for instance, typically refer to (confidential or not) institution-level information such as balance sheets or P&L accounts that reflect the activities reported by individual players like banks or non-financial corporations — covering, for instance, information from central balance sheet data offices.[7]

Turning to granular data, it can be defined as comprising 'disaggregated data and micro data'.[8] Disaggregated data is 'data below the level of aggregated data and with a higher likelihood of identifying individual reporting units than in the aggregated data'.[9] So granular data covers less aggregated data than traditional statistics plus microdata, including when these are anonymised for dissemination. In the

financial sphere, the concept of granularity will, for instance, typically refer to data on individual transactions, for example, on loans or securities, which are an input to the production of aggregated information by compiling institutions, and can also be an area for analyses and research on their own.

In contrast, macro data is 'observation data gained by a purposeful aggregation of statistical micro data conforming to statistical methodology', according to the Organisation for Economic Co-operation and Development (OECD) glossary,[10] while aggregated data is defined as 'data aggregates that have a low likelihood of identification of individual reporting units, such as those found in traditional datasets'.[11]

An important reason for the higher interest in micro and granular information is that financial stress experienced at the level of individual entities, transactions or instruments can quickly reverberate across the entire financial system and beyond. Macro-level analyses should therefore be complemented with a micro-level or, in certain cases, granular approach — 'that is, we need to see the forest as well as the trees within it'.[12] More precisely, the high granularity of certain data sets can help: (i) 'zoom in' and get idiosyncratic information that is important from a system-wide perspective; (ii) give a better sense of the distribution of economic aggregates; (iii) assemble detailed information bricks

Table 1: The potential of linking micro and macro data

Area	Main objective	Examples
Micro-level knowledge	• Ability to 'zoom in'	→ Discovery of idiosyncratic micro information that can be hidden in macro aggregates → Detection of systemically important institutions, role of specific financial instruments, etc → Identification of specific relationships in the functioning of the financial system, for example, role of central counterparties
Distribution information	• Assess the distribution of aggregated variables	→ Distinction between corporates by activities (eg fintech), size (eg small and mid-size enterprises [SMEs]), foreign ownership (eg multinational enterprise [MNEs]) → Assessment of households' inequalities (eg revenues, wealth)
Macro compilation	• Enhance the quality of the compilation of 'macro' statistics	→ Possibility of assembling detailed information bricks to compile more precise macro estimates → Use and aggregation of granular databases, for example, loan-by-loan/security-by-security to compile from-whom to-whom (FWTW) System of National Accounts (SNA) tables → Identification of below macro-level activities (eg fintech services that are not properly covered by the SNA framework)
Policy support	• Develop evidence-based policies	→ Ex ante design of policies based on sub-samples of the population of interest, for example, quantitative impact studies (QIS) → Monitoring/understanding of feedback effects and unintended consequences of current policies → Ex post evaluation of policy effectiveness based on predictive analytics (eg propensity score matching techniques for comparing groups of similar characteristics subject (or not) to policy interventions)
Macroeconomic knowledge	• New approaches in making use of economic and financial data	→ Ability to reassemble and compile economic indicators from different perspectives (eg compilation of nationality-based statistics as a complement to the SNA residency-based approach) → (Micro) panel data analysis/bi-temporal analysis

to compile more precise 'macro' estimates; (iv) facilitate the design and evaluation of evidence-based policies, especially as regards financial sector reforms; and (v) trigger new, innovative ways of looking at economic and financial phenomena (see Table 1[13]).

After more than a decade since the GFC, what have been the results of the various and ambitious initiatives undertaken to collect granular information on the financial system? Central banks and financial supervisors have been at the forefront of this journey, for instance to ensure greater consistency between new micro-level data sets and traditional macro aggregates, adapt statistical frameworks to the rapidly evolving financial system, and exploit micro, firm-level data sets for financial stability work.[14–16] For instance, many jurisdictions around the globe now have at their disposal very large and granular loan-by-loan databases, representing the bulk of what is considered 'Big Data' by central banks.[17,18] Information on derivatives transactions reported by trade repositories (TRs) is another example of the detailed data sets that are increasingly of interest to central banks.[19]

Yet the task of integrating (granular and) micro financial information in macro frameworks has proved more complex than initially thought. It remains hindered by the limited availability of reliable and timely reports at the level of individual institutions, especially in the least regulated corners of the financial system. Moreover, the lack of international harmonisation continues to challenge the collection of comparable, entity-by-entity data among corporations, not least across jurisdictions. Furthermore, the underlying financial statistical infrastructure is still incomplete, reflecting the slow development of global identifiers, standards for exchanging information and data-sharing arrangements — despite notable progress since the GFC as regards the Legal Entity Identifier (LEI),[20] the Statistical Data and Metadata eXchange standard (SDMX)[21] and the actual international sharing of granular information on global financial institutions.[22]

The foregoing difficulties were particularly evident when COVID-19 struck, for instance during the related turmoil that occurred in financial markets at the beginning of 2020.[23] The pandemic also showed that a wealth of available microdata might not be fully exploited. Yet the response of official statisticians in various countries showed that 'quick wins' could be achieved by making better use of the information already collected[24]; for instance, to complement conventional survey statistics that became disrupted, to provide detailed statistics on target populations with richer insights on their distribution, or to shed light on the unexpected consequences of the pandemic (eg assessment of mobility patterns and consumer habits[25]).

Moreover, a balance should be struck between the growing interest in using granular information from alternative sources and the fundamental principles that govern the production of appropriate and reliable official statistics and guarantee that certain professional and scientific standards are adhered to.[26] For instance, it is important to ensure that the microdata involved is unbiased and produced transparently with adequate metadata information. It is also important to ensure that the provision of this kind of information is actually made independent of the data providers' business interests. A third requirement is to favour benchmarking exercises, especially against other types of data, to ensure accuracy and appropriate context. A final point is to certify that the integrity and confidentiality of the data is respected and that the information is used honestly.[27–29]

The above considerations clearly underscore the need to improve the integration of microdata into macro frameworks. This has become an important issue on central banks' agenda, as reflected in the Irving Fisher Committee on Central Bank Statistics' (IFC's) ongoing work on micro-level information — especially in finding appropriate sources, developing new methodological concepts and techniques, compiling policy-relevant indicators and making use of them, and taking advantage of rapid improvements in technology (the 'Big Data revolution').[30–33] An important objective

in this respect has been to highlight existing best practices and potential opportunities, especially to support policymaking, as well as to take stock of the challenges to be addressed as a priority. The starting point is that there are clear limitations in analysing economies based solely on the supply of macro statistics. In contrast, a growing and lasting demand for more granular insights has emerged from a wide range of users — for instance, to monitor risks in the financial and non-financial sectors, analyse interconnectedness and assess international spillovers.

Addressing this supply–demand information gap puts a premium on finding effective ways to drill down from aggregates to more micro information. This can offer a wide range of analytical opportunities, by: (i) reducing the risk of overseeing phenomena that may not be noticed at the macro level; (ii) better supporting the design of macro policies, especially in the areas of monetary and financial stability; and (iii) allowing for more accurate measurement of the impact of authorities' actions. In turn, such monitoring would also enable some fine-tuning wherever and whenever appropriate. Moreover, making better use of granular (including micro) data sets should be favourable for reporting agents, by lowering data collection burden with more stable and resilient requirements and reducing the need for them to compile aggregated data. Harmonising and streamlining data reporting, however, is essential to reap those benefits and corresponds, in fact, to a strong demand expressed by the financial industry.

Looking ahead, in order to make concrete progress attention should be devoted to four main tasks: building up new microdata collection frameworks (Section 'New frameworks for collecting micro data'); accessing and making use of more granular sources of information (Section 'Accessing and using micro data sets'); developing new and adequate analytical tools (Section 'New analytics for working with micro data'); and bridging the gap between micro- and macro-level statistics (Section 'Bridging the gap between macro- and micro-level statistics').

NEW FRAMEWORKS FOR COLLECTING MICRODATA

A first, obvious issue for producers of official statistics is how to organise themselves to collect the increasing amounts of granular information that are potentially available. This is particularly important for central banks and supervisory authorities, which are confronted, on the one hand, with increased reporting fatigue among respondents and, on the other hand, with a steady rise in data needs to conduct evidence-based macro policies and monitor the financial sector.

Various countries' experiences suggest that this trade-off can be best addressed by integrating reporting requirements and standardising data in generic 'cubes', that is, structures that allow for the storing, reporting and analysis of data points with multiple dimensions, so as to cover the wide range of reporting exercises and analytical exercises. Such an approach can also allow new information needs to be addressed with more agility and less effort as they arise.

According to Eurofiling, a forum initiative started in 2005 to improve collaboration on and awareness of European regulatory reporting among regulators, supervisors and entities from both the public and the private sectors, a key lesson of the qualitative case studies for recent international projects is that three main factors are necessary for success.[34] The first is effective collaboration between the multiple parties involved. To ensure high-quality inputs in data collections, central bank statisticians should, in particular, understand how the various reporters are organised, how they can cooperate, what should be done so that they remain engaged over time, and how data platforms should be set up to support the provision of continuous feedback. The second main issue is the demand for comprehensive data methodologies and models. Since

statistical compilation chains involve many stakeholders, statistical definitions should be well documented, compatible and applied consistently for the variety of purposes considered and with sufficient agility (for instance, to reconcile the legal and business approaches supporting data collections). A third issue is to develop adequate platforms and standards for data compilation. This calls for the development of comprehensive technical interfaces (for instance between firms' internal business data warehouses and their submissions for official statistics reporting), shared standards (ideally at the international level since the financial system is global) and common identifiers to correctly identify financial agents as well as their transactions and products.

Certainly, a number of important initiatives are already being developed in several central banks to address these issues. In terms of methodologies, the Bank of Spain has built up a Credit Registry Dictionary and has worked on using a data point model, based on the XBRL (eXtensible Business Reporting Language) standard, to streamline its data collection. Singapore has developed an initiative to foster a common understanding of supervisory and statistical reporting requirements in the financial industry. Turning to Europe, the ECB has designed a single data dictionary, the Banks' Integrated Reporting Dictionary (BIRD), to foster cooperation with the banking industry in the field of regulatory reporting, alleviate the reporting burden and improve data quality. This has been complemented with the provision of a large set of precise definitions that are operational and can even help beyond regulatory reporting — for instance, to support the compilation of the Analytical Credit Database (AnaCredit)[35] — and the use of a single multidimensional data model (SMCube) based on standard identifiers such as the LEI and ISIN (International Securities Identification Number), allowing in turn the same microdata to be used for multiple purposes. Similar efforts have been conducted by the European Securities and Markets Authority (ESMA), which is the rule setter for securities and derivatives markets in Europe, for instance to design a reporting framework for derivatives transactions through TRs combining methodological definitions, the description of the related reporting obligations and the use of an ISO standard. The European Commission has also been working on enhancing financial data standardisation and has developed a fully-fledged 'European strategy for data', especially as regards the financial data space.[36]

In addition to those strategies aiming at streamlining data reporting processes, it may be useful to define a comprehensive data strategy for the institution in charge of data collection. This should encompass the related business cases, the projects envisaged (eg building up a data platform), as well as the related governance issues faced, including data sharing/access and collaboration exercises. The Bank of Thailand has found that such a strategy can lead to many benefits,[37] namely that information is a strategic asset and that one could make more active use of it. The central bank has therefore developed a multi-year plan with the goal of improving the stocktaking, analysis and sharing of the various economic and financial microdata potentially available in the country — especially through greater use of high-frequency data from various sources, including unstructured ones, for example, text. The primary focus has been on micro-level statistics, since their granularity can provide various insights for policy use, for instance to analyse monetary policy mechanisms or developments in property markets, as well as to improve the efficiency of the market operations conducted by the central bank. Certainly, this project has faced many challenges, especially regarding staff skills (to be developed through external recruitment or internal training), the selection of technology to handle vast and various types of data sets (eg structured

as well as unstructured), the development of adequate legal and data governance frameworks, and the need to comply with data protection rules.

On the other side of data collection processes, reporting institutions are also making active progress. This is particularly the case for those banks and other financial institutions that have faced increasing reporting requirements, especially in the wake of the GFC. The resulting burden has been challenging their traditional way of collecting data from internal reports, which were typically organised along separate business processes — with the risk of redundancies, inflexibilities and inconsistencies between different reporting exercises — and were rarely standardised (implying that any new collection exercise would lead to extra reporting costs).

The experience in Austria is that one way to overcome these issues and make data collection exercises more efficient on the reporting side is to better integrate the reporting process.[38] To do so, the central bank has developed a 'RegTech approach' (which basically refers to the provision of methodology, technology and processes to financial institutions to support regulatory monitoring, reporting and compliance). An important factor in this approach was the decision by the largest Austrian banks — in coordination with the central bank — to establish a central reporting platform (AUREP). The goal was to ensure that all the data to be reported by commercial banks are sourced from a general ledger — that is, a general 'basic cube' (input layer) with ad hoc validation rules to support data quality management (DQM). This allows storing and organising all the information in 'smart cubes' made of specific data selected/aggregated/transformed so that these are easily available for external reporting exercises with different types of aggregation level (eg consolidated, sectoral). This set-up can also be used effectively to support the banks' internal reporting.

Such an approach can help to improve data quality, achieve synergies and lower reporting costs and can represent a 'game changer' with an overall impact on the information organisation framework. In addition, it facilitates the use of advanced analytical techniques, as data can be accessed and retrieved more easily. Lastly, one important lesson is that the more granular the reporting requirements, the more effective the approach can be, since the underlying data is not transformed when stored and can thus be reused for different exercises — minimising the potential impact of additional enquiries by the authorities. Further progress could be obtained if regulators set out their reporting requirements in standardised languages, possibly machine-readable ones. This would allow retrieval (or 'pull') of the data from the reporters' own internal cubes, instead of asking them to send (or 'push') the data themselves — reducing, in turn, reporting costs and further increasing data quality and timeliness.

The focus on reducing reporting burden is also high on the agenda of the Reserve Bank of New Zealand, which has developed a 'collect once, use multiple times' approach similar to the one adopted in Austria.[39] One of the main goals has been to enhance the accuracy of the data collected from the financial industry, especially banks, and ensure that all regular data collections are relevant, fit for purpose and cost-effective. In particular, and to avoid misunderstandings in implementing the related instructions, the Reserve Bank ensures that reporting exercises are designed in close consultation with the data providers. Building relationships, providing clarity around data requests and spending time discussing with data providers is not costless but is proving extremely valuable in the long run. Moreover, any new specific data request — covering all prudential, macroprudential and statistical purposes — can be linked back to a single high-level balance sheet or central 'hub'. The experience is that this 'collect once but use it multiple times' approach

can be instrumental in reducing errors in the data sent, addressing new reporting needs with more agility and facilitating data analytics work.

ACCESSING AND USING MICRODATA SETS

Once the (micro) data sets are collected, the ensuing steps are to enable easy and effective access of this information (overcoming the challenges related to size and complexity) and to make sense of it. In other words, it is not sufficient to 'collect' the dots but, more essentially, to 'connect' them[40] — that is, central banks should perform relevant analyses that can support both the preparation and the monitoring of their policies. These include monetary policy, typically at the heart of central banks' missions, but also increasingly macro- (since 2011 for the ECB) as well as microprudential policies (since 2014 for the ECB in the area of banking supervision). The need for adequate information supporting these three different functions has clearly reinforced the interest in collecting granular data. As stated by the former president of the ECB, well established macro statistics will continue to provide the 'big picture', but 'we should also offer a magnifying glass'.[41]

The information provided by very rich granular data sets, however, can be overwhelming. Making it usable requires the support of ad hoc techniques so as to transform simple 'data points' into 'knowledge'.[42] An important factor for success is therefore to have adequate staff skills (eg data scientists) as well as appropriate IT tools (eg Big Data analytics[43]). Another important lesson from central banks' work with microdata is the importance of combining different approaches, requiring the setting up of multidisciplinary teams involving IT staff, subject matter statisticians and economists. One reason is the large variety of microdata sets that require specific competencies to be properly accessed and analysed; for instance, commercial data, administrative data and social media data sets. In addition, the interest of central banks' users has also expanded, so that projects dealing with granular data have become more frequent and serve increasingly multiple purposes — from the traditional research area to using microdata as an auxiliary input (and sometimes even as the core input) for the fine-tuning of macro policies and micro surveillance processes. Furthermore, statistical dissemination has become more complex, as an impressive and increasing wealth of information is made available to internal users and the research community. This has important consequences for the organisation of central banks' data work. One such consequence is that information architectures need to allow for combining multiple data sets, sharing data throughout the institution and supporting various tasks such as regular statistical compilation, ad hoc data exploration and experimental activities. Another is that close cooperation is required among the various stakeholders involved so that different sources of information can be used in a complementary way and/or for different purposes.

These issues can pose important challenges in practice, not least because of the resource constraints faced by public institutions. Hence, while the use of aggregated data sets has become routine for those authorities in charge of monitoring macroeconomic developments and implementing related policies, progress in extracting insightful developments from microdata sets has been much more limited. In particular, while one can relatively easily conduct standard macro aggregations out of individual data points, the identification of patterns of interest out of heterogeneous developments combining large granular data sets can be a much more complex task. This calls for using IT innovation more effectively so as to enable authorities to gather better information to support their policies, for example, in financial supervision ('SupTech'[44]).

One solution to these challenges is knowledge-sharing. To facilitate that, several central banks, together with a host of international organisations and statistical offices, have formed the International Network for Exchanging Experience on Statistical Handling of Granular Data (INEXDA), supported by the Bank for International Settlements (BIS).[45] One of the main objectives of this initiative was to identify best practices in facilitating access to, and sharing of, microdata sources, for instance to develop data access procedures for researchers, as well as to explore new analytical possibilities, such as analysing the effect of policies.[46] This review has highlighted the benefits of automating decisions in workflows by: (i) relying on effective 'metadata schema' that describe how administrative data sets can be managed and used from a legal, organisational or technical point of view; and (ii) having workflows that follow such metadata information deterministically. Lastly, the international statistical community, including central banks, is also actively engaged in public sharing of IT tools in the context of SDMX[47] as well as software codes through secure online software repositories (eg based on the free, open source system Git). This can greatly support the reuse of codes in open source languages such as Python and R for accessing and exploring microdata sets without requiring the actual exchange of the underlying data.

Yet an important lesson is that the devil is often in the details: the various pilots implemented by central banks have faced concrete implementation challenges — including the lack of standardisation of the information contained in the data sets and of the related workflows ('metadata reference'), the difficulties in setting up adequate facilities allowing external users to access this information, and the management of confidentiality risks when results drawn from the work on microdata have to be disseminated.

In fact, a number of successful national initiatives are worth highlighting. One relates to dealing with external research projects using microdata, an area in which the Deutsche Bundesbank has developed significant expertise in recent years. A notable lesson is the importance of enhancing user knowledge of the data sets in question by providing rich contextual information (or 'metadata'). This can be achieved through multiple steps forming a feedback loop, replacing previous one-directional information flows.[48] The first step is to compile 'data-centric metadata', with information on the use of the data and on the related administrative workflows; for instance, to know who has been using a specific type of data in the past and for what purpose. The second is to learn from this experience in an automated way, for example, by relying on ad hoc machine learning (ML) techniques; for instance, to inform a user working on a specific topic what data sets are also of interest for other researchers conducting similar work — providing insights such as 'people working in your area are most frequently interested in these fields'. And the third step is to actively engage with data users to incorporate their specific knowledge and findings and, in turn, improve the data sets' metadata information, so as to facilitate future work — for example, information related to the data viewed by the user, similar data of interest, what others have done with similar data ('recipes') and what alternative recipes may be.

Another interesting initiative has been to facilitate the working of teams with different competencies to make the most of the new data landscape. The objective is to have specialised staff, represented by different 'colours', joining forces and leading to the assembling of mixed business and technology skills in composite groups of 'purple people', that is, those specialised staff whose skills overlap. To ensure such an adequate blending of staff skills, the Banco de Portugal has developed a tailored internal training programme in partnership with a

university, noting that attracting external specialised skills like data scientists had proved to be an unsuccessful task.[49] Similarly, De Nederlandsche Bank has developed a 'Data and Technology' traineeship with the goal of developing 'datapreneurs'.

Another important point is collaboration. The new integrated data management initiative — undertaken by Banco de Portugal as part of its strategic planning aims — promotes the creation of added value by making users, statisticians, data scientists, IT staff, supervisors, etc. work together. This collective approach is reported to have enhanced the use of microdata in the institution, removed cultural/organisational obstacles and rationalised the various processes involved in terms of data collection/processing/sharing. It was also an important element supporting the bank's internal governance model, with data stewards allocated to all business areas and tasked with overseeing the quality and fitness for the purpose of the data assets.[50] Turning to De Nederlandsche Bank, the focus has been on promoting discussions between data scientists and supervisors and policy officers, with the organisation of joint projects in a 'hub and spoke' model.

The good news is that these efforts can rapidly pay off, with central banks increasingly able to develop informative use cases based on the microdata sets stored in their vaults.[51] Moreover, experience shows that this information can improve understanding of the economy's functioning — for example, to facilitate macroeconomic forecasting exercises that are based on the modelling of economic agents' behaviour (for a US illustration, see [52]). In Austria, for instance, efforts to 'bridge' micro results with developments in macro aggregates have supported the analysis of the financial behaviour of households. While macro statistics based on financial accounts had been able to document significant trends in recent years (the 'what'), such as observed shifts in deposits' maturity and financial instruments, more granularity was required to understand the factors at play behind these phenomena (the 'why'). In contrast, the micro-level information drawn from the Consumption and Financing Survey can help identify the types of households behind observed aggregated portfolio changes.[53] In particular, this approach highlighted that the shift from Austrian deposits with longer maturity to overnight deposits could be observed for nearly all types of households but that a tiny fraction of them (those with higher net wealth) were responsible for the aggregated shift from debt securities to investment fund shares. The work on enriching aggregated information for socio-economic characteristics has also emphasised the need to ensure methodological and scope consistency between the different types of data.

Another successful example relates to ECB work on granular data on credit and credit risk. Instead of developing new ad hoc surveys that may lead to inconsistent pictures, the idea has been to pull together all the granular information of potential interest in one data set (collected from AnaCredit) that can support different analyses and policy uses.[54] For instance, such a rich data set can enable analysis of supply and demand credit factors depending on market conditions, small and medium-sized enterprises' specificities, the transmission channel of monetary policy, sectoral risks in the central bank's asset management operations, etc. This wide range of analyses can thus assist various functions, for example, monetary and macro- and microprudential policies as well as emerging policy needs related to environmental, social and governance factors. Moreover, the exercise provides value by allowing different data sets to be connected, with a holistic approach favouring consistency and interoperability across data domains. Furthermore, it was an opportunity to develop new or enhance existing statistics; for instance, data on credit proved to be more accurate when linked to a reliable business register.

To this end, the Register of Institutions and Affiliates Database (RIAD) was set up as the backbone of all granular statistics for the European System of Central Banks, with a unique master data set, a shared platform, high-quality checks and rules and a comprehensive data model including reference data on individual units and relationships among them.[55] Successfully combining different data sets, however, requires clear data definitions and methodologically sound reference information (metadata), highlighting in turn the importance of establishing a good dialogue with the financial industry (see Section 'New frameworks for collecting micro data').

NEW ANALYTICS FOR WORKING WITH MICRODATA

In view of their inherent complexity and size, granular data sets are difficult to work with, as they can require substantial time and IT capacity. This is indeed why the main applications on microdata pursued in the past have basically related to the extraction of aggregated indicators, typically using simple Structured Query Language (SQL)-based queries designed for 'traditional' relational systems. Yet increased computing power capacity and new analytical tools such as ML techniques have clearly changed the picture, as it has become easier to work directly on the underlying data points to perform various purposes, from pure statistical production tasks to more in-depth and/or real-time complex analyses. In turn, the insights gained in understanding, for instance, the functioning of complex financial systems, assessing potential vulnerabilities and analysing contagion effects can help to effectively prepare and monitor policy actions.

Several challenges, however, need to be addressed. The first is data access: one has to identify adequate sources, set up appropriate sharing arrangements and implement data governance procedures, not least as regards the ownership of, and access to, sensitive information.[56] A second issue is whether to opt for on-premises or cloud-based solutions. The latter are increasingly popular, not least because of their scalability and flexibility. In addition, while they can entail security risks, most notably leakages of sensitive data that may have high reputational costs for central banks,[57] the recent experience with the Data Science Hub set up by De Nederlandsche Bank shows that ways can be found to save confidential data in the cloud. A third issue is the 'black box' syndrome: one has to have a good understanding of the techniques involved in order to ensure that analyses are not only accurate but also 'interpretable' — so that specific explanatory causes or factors can be identified and communicated as policy-actionable 'causal stories'. This again emphasises the importance of the human factor.

Despite these challenges, one important avenue explored by central banks in making use of new analytical tools relates to their role as producer of official statistics. In particular, this is used to enhance DQM processes — performed, in general, through both automated checks (to verify predetermined relationships) and plausibility checks (by relying on 'acceptance regions') — for example, by setting up imputation techniques, detecting outliers and duplicates. Yet, and in contrast to traditional macro statistics, the checking of large granular data sets often characterised by heterogeneous reporting patterns can be cumbersome, complex and almost impossible manually.

Automated techniques can greatly facilitate these tasks, as demonstrated by the experience of the Bank of Italy in conducting ML-based quality checks on granular banking data.[58] The approach, based on a supervised learning algorithm, was to detect abnormal patterns in payment services data reported by banks. The results, cross-checked with reporting agents, show that this can help detect 'new outliers' (ie those that had not been initially identified by existing DQM processes) more precisely and with

reasonably high accuracy. Moreover, the set-up of dynamic acceptance thresholds and periodical training of the algorithms allow for an automatic update as new data comes in, reducing analysts' involvement.

Another important application is to make sense of the vast amount of textual information to extract summary indicators, for instance economic sentiment. This kind of approach has been implemented at the Central Bank of the Republic of Turkey to predict economic growth based on subjective information such as emotion, opinion and attitude.[59] The input information was extracted from published statements based on a lexical approach and the use of ML techniques like natural language processing (NLP) for the automated processing and analysis of large amounts of text — with the set-up of a term matrix to generate a vectorial representation of the document of interest and the subsequent extraction of a summary sentiment indicator.

A third important area relates to the analysis of financial networks with the identification of their topological characteristics, such as the likelihood of connection between specific nodes ('connectivity') or neighbours ('clustering'), the distribution of the shortest paths among nodes ('betweenness centrality'), and the relative importance of links and nodes in the system ('average degree'). Micro-level information can be particularly suitable for forming a comprehensive map of an ecosystem and its interconnections, for instance to analyse the financial system of China.[60] Moreover, and as experimented by the Data Science Hub set up by De Nederlandsche Bank, it is an important building block for conducting stress test exercises.[61] These typically involve designing severe circumstances that have yet to materialise and the flexible combination of granular data from different sources to perform the assessment of such hypothetical scenarios. For instance, individual banks' data on stocks (balance sheets) and flows (transactions) can help identify troubled firms, sectors or markets, allowing the propagation of potential shocks to be traced through the financial sector.

The Central Bank of Malaysia has followed a similar approach, with a more global perspective so as to analyse international linkages and cross-border contagion risk given the growing regionalisation of domestic banks and the significant presence of foreign banks in Malaysia.[62] The work was based on granular data provided by banks that was used to develop stress-tests to: (i) assess banks' vulnerability to external shocks at the macro and micro levels; (ii) estimate contagion paths across banks (ie how the shocks propagate themselves); and (iii) identify potential systemic linkages and cross-border domino effects between internationally active banks and Malaysian banks. The results show that the solvency contagion risk of Malaysian banks to external shocks has declined in recent years, but that potential sources of contagion remain. The approach also suggests that a micro-based network analysis approach, combined with counterfactual simulation, could usefully complement macroprudential surveillance tools and support the development of effective arrangements for crisis containment, management and resolution in case unexpected shocks happen. Lastly, the modelling highlights the benefits of cooperation in terms of information-sharing and surveillance to effectively address cross-border contagion risks.

BRIDGING THE GAP BETWEEN MACRO- AND MICRO-LEVEL STATISTICS

There has been a growing interest among central banks and financial supervisors in directly conducting analyses in a very granular way, especially to better prepare the actions of public authorities and help measure their effectiveness, as documented previously. Yet an important point of interest for policymakers is

how to link developments on the individual institutional or household level to the more general picture. Granular data sets can surely not be a substitute for aggregated statistics, but their input can add value for analysing macroeconomic issues, for instance the functioning of the complex financial system, the evolution of inequalities or sectoral issues.

In order to better understand the financial system, bridging the gap between micro- and macro-level statistics can be instrumental in assessing financial stability risks, an important issue for central banks.[63] The reason is that a financial crisis is, by definition, a system-wide, macro event that can be caused by very specific, micro factors, for instance at the level of financial institutions that have systemic importance. The challenge is, therefore, to extract signals that can have more general implications from the wealth of information on idiosyncratic firms. Research conducted at the BIS on predicting and preventing financial crises suggests that a good starting point is to assess the risk of individual bank distress by using bank loan data from credit registries to identify credit demand and supply factors.[64] From there, one can draw useful insights into the general effects of policies, including their respective contributions. For instance, the use of countercyclical macroprudential tools may address financial imbalances in a targeted way, while monetary policy instruments can affect the economy in a more indistinct way because of their general impact on credit or debt service ratios. Moreover, the granularity of the data can also be helpful in qualifying the results observed at the global level, for example, to assess the special challenges faced by emerging market economies.[65] One example relates to the Latin American region, where the effectiveness of financial stability policy actions is reported to be stronger when monetary and macroprudential policies are synchronised.

Turning to distributional issues, the current macro framework presents notable shortcomings. Certainly, the 2008 SNA clearly states the importance of considering the skewed distribution of income and wealth across households. But it recognises that getting this information is 'not straightforward and not a standard part of the SNA' (2008 SNA, no 24.69). One illustration of this point is the work conducted by the Board of Governors of the US Federal Reserve System to develop distributional financial accounts (DFAs), by using microdata from the Survey of Consumer Finances to distribute the aggregate household balance sheet information available from the Financial Accounts.[66] Another initiative to complement macro aggregates with micro-level information has been to enhance the measurement and geographical distribution of aggregate US housing wealth, a major component of total household wealth and also an important variable to consider in quantitative macroeconomic models and analyses.[67] Granular 'Big Data' information on housing values generated by ML-based valuation models from Zillow, a private real estate and data analytics firm, was aggregated with additional data from the US Census to estimate and make public long monthly series of aggregate housing wealth.[68] Moreover, additional detailed geographical breakdowns were also publicly released. The method has proved to be robust and accurate, providing much insight into the specificities observed at the local level and in a timely manner.

A final issue of interest is the better assessment of developments that are neither at the micro nor at the macro level, for instance by economic sector. In Europe, work has been conducted at Eurostat to improve understanding of the factors driving sectoral productivity developments, based on a similar approach developed by the University of Groningen in The Netherlands.[69] Taking below macro-level information into account has helped produce estimates of labour productivity, supporting a wide range of analyses. A major issue was the adjustment of sectoral measures for purchasing power parities (PPPs), which

are indicators of price-level differences across regions or countries — knowing that, traditionally, PPPs are not estimated for sectors but from the expenditure side of GDP. The International Comparison Program was set up to calculate such price-level indices for a wide range of goods and services, to be used as inputs for various studies such as real GDP comparisons or the measurement of poverty rates.

CONCLUSION

An impressive expansion of micro (granular) data sets has been made available since the GFC of 2007–2009. An adequate integration of this information has the potential to lead to a significant improvement in macro compilation exercises. To make this happen, however, it is important to broaden official statistical frameworks to make use of the growing availability of granular data to assist the assembling of macroeconomic aggregates and/or facilitate the linkage between micro- and macro-level statistics.

In addition, the 'microdata revolution' can bring important analytical benefits and effectively support central bank policies, with a greater ability to 'zoom in' on particular areas of interest and to assess the distribution of macroeconomic aggregates within reporting populations. One significant example is that granular data and microdata already assist in assessing how finance can contribute to the greening of economies.[70] The initiative also proved to be a welcome opportunity to highlight the potential of new types of granular information in 'unusual' circumstances. Indeed, and as observed subsequently with the sudden outbreak of the COVID-19 pandemic in early 2020, microdata sources can be a valuable complement to the 'traditional' official macro statistics available in times of stress, especially to alleviate compilation disruptions, assess pressure points and facilitate the implementation of targeted policy measures.

There are, however, also important challenges associated with dealing with these data sets, for instance as regards their quality, confidentiality and manner of accessing them. Indeed, the task of integrating (granular and) micro financial information in macro frameworks has proved more complex than was initially expected. To ensure concrete progress in the future, attention should focus on the entire information value chain (Table 2). In this context, the main action points should be to:

- build effective microdata collection frameworks based on a comprehensive data strategy helping to contain reporting burden;
- access and make use of more granular sources of information, with the need to overcome the challenges related to their size and complexity with a view to transforming simple data points into knowledge;
- promote the exchange of experience, for example, as regards access to microdata sets and external research projects, the development of diversified staff skills and the combination of different types of data sets;
- develop new and adequate analytical tools, for instance to enhance data quality assurance processes, extract summary indicators from a wealth of data points, and develop ML/text mining/network analysis approaches to maximise the potential of granular (including micro-level) information; and
- bridge the gap between micro- and macro-level statistical exercises as a means of enhancing the understanding of how the financial system functions and interacts with the economy, assessing distributional issues and facilitating sectoral analyses.

In supporting the above tasks, there is a growing interest in making use of granular information from private sources that are not part of the official statistical offering. What is unclear, however, is how data producers

Table 2: From micro to macro: The information value chain

Key elements	Main goals	Related issues
Information strategy	• High-level principles for data compilers/producers	→ Data as a strategic asset to be better shared and used in the organisation → 'Collect once, use multiple times' goal → High-level resources commitment (eg IT capacity, staff, budget)
	• High-level principles for data reporters	→ Internal information organisation (eg general ledger, 'smart cubes')
	• Coordination data compilers/reporters	→ Guidance on methodology, technology and processes involved (eg RegTech) → Identification of best practices (eg metadata, workflows, IT tools)
Data governance	• Building confidence in data management and use (eg data/confidentiality protection)	→ Organisational features for dealing with data → Cover all involved principles, structures, policies and procedures (eg quality assurance processes) → Ensure adequate staff skills (eg training)/new talents (eg data scientists)
	• Clarification of roles and responsibilities	→ Multidisciplinary teams with clear governance responsibilities (eg role of data stewards)
Data access	• Access rights	→ Data-sharing agreements → Copyright issues
	• Possibility to ingest data from multiple sources	→ Data storage requirements → Connection of data sets (eg identifiers) → Data knowledge (eg role of data owners)
Data collection	• Common reporting process	→ Access to and collection of raw data → Use of comprehensive data models/methodologies → Collaboration with stakeholders on data requirements → Ingestion of different data formats/structured versus unstructured data sets
	• Ability to collect a variety of data	→ Interest for complementary types of data (eg for crises versus normal times)
	• Data loading in a central place	→ Adequate platform and information transfer processes (eg technical interfaces, shared standards, identifiers, registers) → Grouping of individual data submissions
Data manipulation	• Support of different data management tasks, for example, data exploration versus regular statistical production	→ Automatised data cleaning (eg duplicates, outliers) → Data elaboration/transformation, business intelligence (BI) tools → Data integration (eg combination of sources)/aggregation
	• Micro–macro linkages	→ cf Table 1
Knowledge discovery	• Apply statistical techniques	→ 'Traditional' statistical tools → More sophisticated techniques, for example, econometrics, principal component analysis
	• Deal with Big Data	→ Big Data 'advanced analytics': ML/AI/NLP tools
	• Extract patterns in the data	→ Data mining (eg clustering/classification/use of textual information/network analysis)

(*Continued*)

Key elements	Main goals	Related issues
Communication/dissemination	• Present data to the users	→ Summary dashboards → Metadata information → Putting data in perspective (eg 'stories' versus black box calculations)
	• Policy support	→ Internal presentations (eg policy meetings) → Extraction of critical information, for example, suptech
	• External communication	→ Press releases/statistical publications → Development of economic/financial literacy → Information on usage of the data → External researchers' facilities

located outside the national statistical system should feature vis-à-vis the fundamental principles that govern the production of appropriate and reliable official statistics and adhere to certain professional and scientific standards. Helpfully, a number of private firms have already adopted dedicated and transparent mission statements and principles to address these concerns, and such initiatives may deserve to be strongly encouraged by the official (international) statistical community.[71]

Lastly, there are specific communication challenges, as well as potential legal difficulties, for policymaking institutions such as central banks when (confidential) granular analytical insights are used as the foundation for their decisions.

Authors' note

This article draws on the various presentations made on the occasion of the IFC Satellite Seminar on "Post-crisis data landscape: micro data for the macro world", co-organised with the Central Bank of Malaysia (CBM) and the European Central Bank (ECB) in Kuala Lumpur, Malaysia in 2019. The proceedings were published in the IFC Bulletin series. The views expressed here are those of the authors and do not necessarily reflect those of the BIS, the CBM, the ECB, the IFC or any of those institutions represented at the seminar. We thank Barend de Beer, Neil Humphries, Robert Kirchner, Michael Kock, Lim Sheng Ling, Ong Li Ming, Eric Nielsen, Laurent Olislager, Steffi Schuster and Johannes Turner for helpful comments and suggestions.

Notes and references

1. Greenwood, R., Scharfstein, D. (Spring 2013) 'The growth of finance', *Journal of Economic Perspectives*, Vol. 27, No. 2, pp. 3–28.
2. IMF Staff, FSB Secretariat. (2009) 'The financial crisis and information gaps: Report to G20 Finance Ministers and Central Bank Governors', Financial Stability Board and International Monetary Fund, October 29.
3. IMF Staff, FSB Secretariat. (2015) 'The financial crisis and information gaps', Sixth Progress Report Implementation of the G20 Data Gaps Initiative, Financial Stability Board and International Monetary Fund, September.
4. European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations and World Bank. (2009) 'System of National Accounts 2008', EC, IMF, OECD, UN and WB, New York, pp. 9–10.
5. OECD Glossary of Statistical Terms. (n.d.) 'Glossary of statistical terms', available at: https://stats.oecd.org/glossary/ (accessed 29th September, 2021).
6. For a description of relevant experience by countries and regional and international organisations that illustrates the efforts undertaken in the various related areas covering the use of common statistical identifiers; the exchange of experience on statistical work with granular data and improvements in transparency; the balancing of confidentiality and user needs; the linking of different data sets; the provision of data at the international level; the ways of improving data sharing of granular data; and the collection of data only once, see Inter-Agency Group on Economic and Financial Statistics. (2017) 'Update on the Data Gaps Initiative and the outcome of the Workshop on Data Sharing', IAG, March 14.
7. Irving Fisher Committee on Central Bank Statistics. (2017) 'Uses of Central Balance Sheet Data Offices' information', Proceedings of the IFC-ECCBSO-CBRT Conference, Özdere-İzmir, Turkey, September 26, 2016, *IFC Bulletin*, No. 45, October.
8. *Ibid.*, ref. 6 above.
9. *Ibid.*, ref. 6 above.
10. *Ibid.*, ref. 5 above
11. *Ibid.*, ref. 6 above.
12. Borio, C. (2013) 'The Great Financial Crisis: Setting priorities for new statistics', *Journal of Banking Regulation*, Vol. 14, pp. 306–317. Also published in BIS Working Papers, No. 408, April.

13. Tissot, B. (2018) 'Providing comparable information to assess global financial stability risks', *Eurostat Statistical Reports* KS-FT-18-001, January, European Union, Paris.
14. *Ibid.*, ref. 6 above.
15. See Irving Fisher Committee on Central Bank Statistics. (2017) 'Data needs and Statistics compilation for macro-prudential analysis', Proceedings of the IFC–National Bank of Belgium Workshop, Brussels, May 18–19, *IFC Bulletin*, No. 46, December.
16. *Ibid.*, ref. 7 above
17. Schubert, A. (2016) 'AnaCredit: Banking with (pretty) big data', Central Banking, 28th October.
18. Irving Fisher Committee on Central Bank Statistics. (2017) 'Big data', Proceedings of the IFC Satellite Seminar on Big Data at the ISI Regional Statistic Conference 2017, Bali, Indonesia, September 2017, IFC Bulletin, No. 44, September.
19. Irving Fisher Committee on Central Bank Statistics. (2018) 'Central banks and trade repositories derivatives data', *IFC Report*, No. 7, October.
20. Legal Entity Identifier Regulatory Oversight Committee. (2016) 'Collecting data on direct and ultimate parents of legal entities in the Global LEI System — Phase 1', LEI ROC, March 10.
21. Irving Fisher Committee on Central Bank Statistics. (2016) 'Central banks use of the SDMX standard', *IFC Report*, No. 4, March.
22. Bese Goksu, E., Tissot, B. (2018) 'Monitoring systemic institutions for the analysis of micro–macro linkages and network effects', *Journal of Mathematics and Statistical Science*, Vol. 4, pp. 129–136.
23. Financial Stability Board. (2020) 'Holistic review of the March market turmoil', November 17.
24. Ducharme, L. M., Tebrake, J., Zhan, Z. (2020) 'Keeping economic data flowing during COVID-19', *IMF Blog*, May 26.
25. De Beer, B., Tissot, B. (2020) 'Implications of Covid-19 for official statistics: A central banking perspective', Irving Fisher Committee on Central Bank Statistics, *IFC Working Papers*, No. 20, November.
26. United Nations. (2013) 'Resolution adopted by the economic and social council on 24 July 2013', E/RES/2013/21, October 28.
27. Irving Fisher Committee on Central Bank Statistics. (2021) 'Issues in Data Governance', *IFC Bulletin*, No. 54, July.
28. See, for example, the Research Mission Statement and related principles published by the US online real estate company Zillow, whose data is now being used by the Board of Governors of the US Federal Reserve System for official statistical purposes.
29. Hume McIntosh, S. (2018) 'Measuring aggregate housing wealth: New insights from automated valuation model', IFC-BNM-ECB Satellite Seminar on Post-Crisis Data Landscape: Micro Data for the Macro World, Kuala Lumpur, Malaysia, August 16, 2019.
30. Irving Fisher Committee on Central Bank Statistics. (2016) 'Combining micro and macro statistical data for financial stability analysis', Proceedings of the IFC Warsaw Workshop, December 14–15, 2015, *IFC Bulletin*, No. 41, May.
31. *Ibid.*, ref. 7 above.
32. Irving Fisher Committee on Central Bank Statistics. (2020) 'Computing platforms for big data analytics and artificial intelligence', *IFC Report*, No. 11, April.
33. Irving Fisher Committee on Central Bank Statistics. (2021) 'Use of big data sources and applications at central banks', *IFC Report*, No. 13, February.
34. Piechocki, M. (2021) 'Paradigms for modern statistical data frameworks and systems', Proceedings of the IFC-CBM-ECB Satellite Seminar, Kuala Lumpur, Malaysia, August 16, 2019, *IFC Bulletin*, No. 53, April.
35. European Central Bank: Eurosystem. (2019) 'AnaCredit Reporting Manual: Part I — General Methodology', 2nd ed., November, ECB, Frankfurt, Germany.
36. European Commission. (2020) 'A European strategy for data', Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, COM (2020) 66 final, pp. 30–31.
37. Chansang, P., Bank of Thailand. (2021) 'Data management in the data evolution era at Bank of Thailand', IFC Satellite Seminar on Post-Crisis Data Landscape: Micro Data for the Macro World, Kuala Lumpur, Malaysia, *IFC Bulletin*, No. 53.03, August 16, 2019.
38. Turner, J. (2018) 'OeNB's innovative reporting data model as RegTech/SupTech solution?', *ONB Eurosystem*, June 7.
39. Schuster, S., Humphries, N. (2019) 'Collect once, use multiple times — the Reserve Bank of New Zealand's data collection approach', *Reserve Bank of New Zealand*, August 16.
40. Caruana, J. (2017) 'International financial crises: New understandings, new data', speech at the National Bank of Belgium, Brussels, February 6.
41. Draghi, M. (2016) 'Central bank statistics: Moving beyond the aggregates', speech at the Eighth Statistics Conference, Frankfurt am Main, July 6.
42. Drozdova, A. (2017) 'Modern informational technologies for data analysis: From business analytics to data visualization', Eighth IFC Conference on Statistical Implications of the New Financial Landscape, Basel, Switzerland, September 8–9.
43. Wibisono, O., Ari, H., Widjanarti, A., Zulen, A., Tissot, B. (2019) 'The use of big data analytics and artificial intelligence in central banking', Proceedings of the IFC–Bank Indonesia International Workshop and Seminar on Big Data, Bali, Indonesia, July 23–26.
44. Broeders, D., Prenio, J. (2018) 'Innovative technology in financial supervision (suptech) — the experience of early users', *FSI Insights on Policy Implementation*, No. 9, July.
45. Bender, S., Hirsch, C., Kirchner, R., Bover, O., Ortega, M., D'Alessio, G., Teles Dias, L., Guimarães, P., Lacroix, R., Lyon, M., Witt, E. (2018) 'INEXDA — the Granular Data Network', *IFC Working Paper*, No. 18, October.
46. Bender, S., Deutsche Bundesbank. (2021) 'Best practise recommendations on facilitating access to microdata: Outcomes from the INEXDA working group on data access', IFC Satellite Seminar on Post-Crisis Data Landscape: Micro Data for the Macro World,

Kuala Lumpur, Malaysia, *IFC Bulletin*, No. 53.06, August 16, 2019.
47. *Ibid*., ref. 21 above.
48. Hirsch, C., Deutsche Bundesbank. (2021) 'Enhancing user knowledge by providing (richer) context to microdata', IFC Satellite Seminar on Post-Crisis Data Landscape: Micro Data for the Macro World, Kuala Lumpur, Malaysia, *IFC Bulletin*, No 53.07, August 16, 2019.
49. Teles Dias, L. (2021) 'Post-crisis skills landscape: The emergence of "Purple People" ', IFC Satellite Seminar on Post-Crisis Data Landscape: Micro Data for the Macro World, Kuala Lumpur, Malaysia, *IFC Bulletin*, No 53.08, August 16, 2019.
50. Moreno, M. (2021) 'Data Governance: an orchestra of people, processes and technology', IFC High-Level Meeting on Data governance, Tunis, Tunisia, IFC Bulletin, N 54.07, November 22, 2019.
51. La Cava, G. (2015) 'The development of databases linking micro and macro data — an Australian perspective', IFC Satellite Meeting at the ISI Regional Statistics Conference on Is the Household Sector in Asia Overleveraged: What Do the Data Say, Kuala Lumpur, Malaysia, *IFC Bulletin*, No 40, November 15, 2014.
52. Holmquist, E., Hume McIntosh, S. (2020) 'Policy uses of financial accounts: The U.S. Fed's perspective', IFC Workshop on the Use of Financial Accounts, Istanbul, Turkey, IFC Bulletin, No 51.16, March 18–20, 2019.
53. Turner, J. (2021) 'Financial behaviour of households under the lenses of distributional aspects: The OeNB's experiences', IFC Satellite Seminar on Post-Crisis Data Landscape: Micro Data for the Macro World', Kuala Lumpur, Malaysia, *IFC Bulletin*, No 53.09, August 16, 2019.
54. Israël, J.-M., Damia, V., Bonci, R., Watfe, G. (2017) 'The analytical credit dataset: A magnifying glass for analysing credit in the euro area', *ECB Occasional Paper Series*, No. 187, April.
55. Israël, J.-M. (2021) 'Linking micro datasets to better service policy-making and analyses', IFC Satellite Seminar on Post-Crisis Data Landscape: Micro Data for the Macro World, Kuala Lumpur, Malaysia, August 16, 2019.
56. *Ibid*., ref. 27 above.
57. *Ibid*., ref. 32 above.
58. Zambuto, F., Buzzi, M. R., Costanzo, G., Lucido, M. d., Ganga, B. L., Maddaloni, P., Papale, F., Svezia, E. (2020) 'Questioni di Economia e Finanza — Quality checks on granular banking data: An experimental approach based on machine learning', No. 547, March.
59. Akisoglu, B. (2021) 'Predicting economic growth using machine learning techniques and sentiment analysis', IFC-BNM-ECB, August 16, 2019.
60. Ehlers, T., Kong, S., Zhu, F. (2018) 'Mapping shadow banking in China: Structure and dynamics', *BIS Working Papers*, No. 701, February.
61. Ullersma, C., van Lelyveld, I. (2021) 'Granular data offer new opportunities for stress testing', in: Farmer, D., Schuermann, T., Kleinnijenhuis, A., Wetzer, T. (eds.), 'Handbook of Financial Stress Testing', Cambridge University Press, Cambridge.
62. Sababathy, H., Sheng Ling, L. (2021) 'Cross-border contagion risk to the Malaysian banking system', IFC Satellite Seminar on Post-Crisis Data Landscape: Micro Data for the Macro World, Kuala Lumpur, Malaysia, *IFC Bulletin*, No 53.14, August 16, 2019.
63. Crockett, A. (2000) 'Marrying the micro- and macro-prudential dimensions of financial stability', speech to the 11th International Conference of Banking Supervisors, September 20–21, Basel, Switzerland.
64. Hofmann, B. (2021) 'Predicting and preventing financial crisis — where do we stand?', IFC Satellite Seminar on Post-Crisis Data Landscape: Micro Data for the Macro World, Kuala Lumpur, Malaysia, *IFC Bulletin*, No. 53.17, August 16, 2019.
65. Bank for International Settlements. (2019) 'Chapter II: Monetary policy frameworks in EMEs: Inflation targeting, the exchange rate and financial stability', *BIS Annual Economic Report 2019*, BIS, Basel, Switzerland.
66. Batty, M., Briggs, J., Pence, K., Smith, P., Volz, A. (2019) 'The distributional financial accounts of the United States', Finance and Economics Discussion Series 2019-017, Board of Governors of the Federal Reserve System, Washington, DC, August 30.
67. *Ibid*., ref. 29 above.
68. Gallin, J., Molloy, R., Nielsen, E., Smith, P., Sommer, K. (2021) 'Measuring aggregate housing wealth: New insights from machine learning', *Journal of Housing Economics*, Vol. 51, p. 101734.
69. Olislager, L. (2021) 'Key indicators for sectoral productivity analysis', IFC Satellite Seminar on Post-Crisis Data Landscape: Micro Data for the Macro World, Kuala Lumpur, Malaysia, *IFC Bulletin*, No. 53.15, August 16, 2019.
70. See, for example, the OECD's 'Microdata Analysis for Environmental Finance and Investment (MEFI) initiative'.
71. *Ibid*., ref. 27 above.