Irving Fisher Committee on Central Bank Statistics

IFC Report

No 13

Use of big data sources and applications at central banks

2020 survey conducted by the Irving Fisher Committee on Central Bank Statistics (IFC)

February 2021

BANK FOR INTERNATIONAL SETTLEMENTS

Contributors to the IFC report[1]

BIS, IFC	José María Serena Bruno Tissot
BIS	Sebastian Doerr Leonardo Gambacorta

This publication is available on the BIS website (www.bis.org).

ISSN 1991-7511 (online)

ISBN 978-92-9259-436-7 (online)

1 Respectively, Economist (Jose.Serena@bis.org); Head of the Secretariat of the Irving Fisher Committee on Central Bank Statistics (IFC) and Head of Statistics & Research Support, BIS (Bruno.Tissot@bis.org); Economist (Sebastian.Doerr@bis.org); and Head of the Innovation and the Digital Economy Unit (Leonardo.Gambacorta@bis.org).

The views expressed are those of the authors and do not necessarily reflect the views of the IFC, its members or the BIS.

We thank Giulio Cornelli, Ade Nurmeme Mohd Noor, Kuniko Moriya and Li Ming Ong for helpful comments and suggestions, and Haiwei Cao, Giulio Cornelli and Alexandra End for excellent research support.

Contents

Executive summary

Big data sources are developing fast, and applications for making use of this new information are flourishing in parallel. This primarily reflects the impact of digitisation, with the development of the "internet of things" as well as a greater ability to digitally process "traditional" information, such as text. It is also a consequence of the large databases that have been created as an "organic" by-product of the complex operations taking place in our modern societies. Additionally, vast amounts of data have emerged in the administrative, commercial and financial areas, an evolution spurred by the important data collection strategies undertaken after the Great Financial Crisis (GFC) of 2007–09 in order to address the information challenges posed by the development of finance.

Central banks are no exception to this general picture. They have shown an increasing interest in using big data in recent years, as already documented extensively by the Irving Fisher Committee on Central Bank Statistics (IFC) (IFC (2017), Nymand-Andersen (2016), Mehrhoff (2019), Tissot (2017)). Central bank big data-related work covers a variety of areas, including monetary policy and financial stability as well as research and the production of official statistics. However, in contrast to the rapid pace of innovation seen in the private sector, big data applications supporting central banks' operational work had initially been limited. This reflects a number of constraints, such as a lack of adequate resources as well as the intrinsic challenges associated with using big data sources to support public policy.

Looking ahead, will central banks catch up and radically transform the way they operate in order to fully reap the benefits of the information revolution? Or will their use of big data sources and applications progress only gradually due to the inherent specificities of their mandates and processes? To shed light on these issues, **in 2020 the IFC organised a dedicated survey on central banks' use of and interest in big data**, updating a previous one conducted five years earlier.[2] The survey focused on the following key questions: What constitutes big data for central banks, and how strong is central banks' interest in it? Have central banks been increasing their use of big data and, if so, what were the main applications developed? And finally, which constraints are central banks facing today and how can they be overcome?

The survey's main conclusions are the following:

- **Central banks have a comprehensive view of big data, which can comprise very different types of data sets.** First and foremost, it includes the large "non-traditional" (or unstructured) data often characterised by high volume, velocity and variety and that must be processed using innovative technologies. But for two thirds of respondents, big data also includes large "traditional" (ie well structured) data sets that are often "organic", in the sense that they are collected as a by-product of commercial (eg payment transactions), financial (eg tick-by-tick price quotes observed in financial markets) and administrative (eg files collected by public institutions) activities – these data are often referred to as "financial big data".

[2] Almost two thirds of the 92 IFC institutional members answered the survey; see Annex 1 for the survey questionnaire and Annex 2 for the list of participating institutions.

- **Central banks are increasingly using big data.** Around 80% of the responding central banks now use big data regularly; in contrast, only one third of 2015 respondents had indicated they were using any big data sources. Moreover, interest in the topic of big data at the senior policy level is currently rated "very important" in more than 60% of cases, compared with less than 10% in 2015. Interest in big data is especially strong among advanced economies (AEs) and is catching up in a significant number of emerging market economies (EMEs).
- **The range of big data sources exploited by central banks is diverse.** A key source for the private sector is the "internet of things", with for instance the applications developed by many central banks to scrape online portals for information in numerical (eg prices of goods sold on the web) or textual format (eg messages posted on social media). Yet another important source of information is text from printed materials processed using digital techniques. Last but not least, central banks are increasingly using financial big data sets collected in a more "traditional" way, such as balance sheet information available in credit registries, loan-by-loan and security-by-security databases, derivatives trades reported to trade repositories (TRs), and payment transactions.
- **Big data is effectively used to support central bank policies.** As regards central banks' monetary policy and financial stability mandates, newly available databases and techniques are increasingly mobilised to support economic analyses and nowcasting/forecasting exercises, construct real-time market signals and develop sentiment indicators derived from semi-structured data. This has proved particularly useful in times of heightened uncertainty or economic upheaval, as observed during the Covid-19 pandemic. A majority of central banks also report using big data for micro-level supervision and regulation (suptech and regtech), with an increasing focus on consumer protection; for instance, to assess misconduct, detect fraudulent transactions or combat money laundering.
- The survey also underscored the **need for adequate IT infrastructure and human capital.** Many central banks have undertaken important initiatives to develop big data platforms so as to facilitate the storage and processing of very large and complex data sets. But progress has varied, reflecting the high cost of such investments and the need to trade off various factors when pursuing these initiatives. Additionally, central banks need to hire and train staff, which is difficult due to the limited supply of adequately skilled candidates (eg data scientists).
- Apart from IT aspects, **there are many other challenges that central banks face**. These include the legal basis for using private information and the protection, ethics and privacy concerns this entails, and the "fairness" and accuracy of algorithms trained on preclassified and/or unrepresentative data sets. Data quality issues are also significant, since much of the new big data collected as a by-product of economic or social activities needs to be curated before proper statistical analysis can be conducted. This stands in contrast to traditional sources of official statistics that are designed for a specific purpose, eg surveys and censuses.
- Moreover, a key issue is to **ensure that predictions based on big data are not only accurate but also "interpretable" and representative**, as to carry out evidence-based policy central banks need to identify specific explanatory causes or factors. Furthermore, transparency regarding the information produced by big data providers is essential to ensuring that its quality can be checked and that public decisions can be made on a sound, clearly communicated basis. Lastly,

there are important legal constraints that reduce central banks' leeway when using private and confidential data.

- **Cooperation could facilitate central banks' use of big data**, in particular through collecting and showcasing successful projects and facilitating the sharing of experience, for example to avoid repeating others' mistakes when setting up an IT infrastructure, or by pooling resources together. In particular, developing technical discussions between institutions is seen as a powerful way to build the necessary skillset among staff and develop relevant IT tools and algorithms that are best suited to central banks' (idiosyncratic) needs.
- **International financial institutions can help foster such cooperation.** For instance, they can help develop in-house big data knowledge, reducing central banks' reliance on big data services providers, which can be expensive and entail significant legal and operational risks. They can also facilitate innovation by promoting technological solutions and initiatives to enhance the global statistical infrastructure. In addition, they can make their resources available internationally or develop joint cloud computing capabilities to reduce operational risk arising from dependence on specific providers in a highly concentrated market.

1. Introduction

Big data sources and related innovative information technologies such as artificial intelligence (AI) and machine learning (ML) are garnering more and more attention in the central bank community. This strong interest is nothing new, and was already emphasised by the previous survey conducted among IFC members in 2015 (IFC (2015)). However, **three key reinforcing developments** have taken place since then: large data sets have become increasingly available; new techniques for handling this information in a practical way have been made accessible; and central banks have been actively setting up the IT infrastructure necessary for dealing with this environment; in particular, new big data platforms that facilitate the storage and processing of very large data sets and high-performance computing (HPC) that allows for faster processing, in-depth statistical analysis and complex data simulations (IFC (2020b)).

As regards the first development, **vast amounts of data are increasingly being made available**, reflecting several factors. First, the so-called data revolution, pushing the "organic" increase in information collected as a by-product of other activities (Groves (2011)). Indeed, many processes leave a digital footprint, leading to the creation of vast amounts of records. The increase in digital payments (eg through credit cards or mobile devices) is a case in point. For one, every transaction is recorded with granular details on eg its amount, currency, beneficiary, purpose and location. Additionally, the use of technological innovation (fintech payments) is fostering financial inclusion (CPMI-WB (2020)) and the documentation of transactions that were previously conducted informally. Another growing source of data has been the widespread use of web-based tools such as internet searches.[3] However, the "internet of things" is much broader and basically includes the network of all connected digital

[3] For instance, the searches provided by Google Trends; see trends.google.com/trends/?geo=US.

devices such as mobile phones or physical sensors. As a result, a growing part of human activity is leaving a digital footprint that provides detailed information on personal activities, locations and habits. Mobility trends derived from smartphone location data (eg based on Apple Mobility Trends Reports[4]) are an example of this. A third key source has arisen out of a growing ability to process printed formats, especially by scanning pictures and text. This allows textual (unstructured) information to be accessed in a new way that is easier to handle. And a fourth source has been the collection of granular structured data sets that provide much richer insights on the distribution of the population of interest compared to conventional sample surveys. This is exemplified by the initiatives launched after the GFC to collect very large data sets on the financial system, often with high frequency (FSB-IMF (2009)).

Turning to the second factor behind the increase in central banks' interest, **various techniques have become accessible to practitioners** that facilitate the handling of large and complex data sets. Attention has, on the one hand, focused on relatively simple, user-friendly business intelligence tools (IFC (2019a)), which can facilitate data discovery and visualisation processes with various functions such as drill-down versus drill-across/through capabilities, dashboards and interactive query interfaces. On the other hand, more complex analyses can be performed by relying on "big data analytics" that can provide faster, more holistic and more connected insights (FSB (2017)). Many central banks have specifically developed applications using ML, a subfield of AI that relies on a sequence of actions for automated optimisation in order to solve a problem (Chakraborty and Joseph (2017)). A key advantage they have is the ability to design algorithms that automatically improve themselves in successive iterations as new observations come in (Doerr et al (2021)). These techniques can be relatively simple to apply, for instance in order to conduct basic analyses (eg text content assessment using string metrics), identify clusters of "similar" observations in a given data set, focus the analysis on a reduced number of dimensions, and identify relationships between elements in a system in order to analyse network effects. Yet the calculations involved can also be more complex, as is the case with unsupervised ML algorithms that require little human intervention and deep learning algorithms that can replicate the functioning of neural networks in order to detect patterns in data of various formats (eg unstructured data). Experience shows that central banks are increasingly using these more sophisticated approaches for a variety of purposes, ranging from data cleaning, nowcasting and network analysis to natural language processing for textual analysis (Wibisono et al (2019)).

Third, central banks have been actively working on **building up the IT infrastructure necessary to exploit big data**. The focus has been primarily on developing (i) platforms that can process very large data sets and deal with information of a semi-structured or unstructured nature (eg "data lakes"); (ii) HPC capabilities to support fast and complex statistical analysis/simulation; and (iii) the related human resources and skills (which require striking the right balance between data scientists, IT specialists, economists and mathematicians/statisticians). Certainly, progress has been limited in practice, since central banks have to deal with many concrete issues such as the selection of the related hardware, the choice between proprietary and open source technology, the decision between developing the solution in-house or in the cloud, and the type of information to be handled. In doing

[4] See https://covid19.apple.com/mobility.

so, they have to trade off various factors such as technology trends, system complexity, cost, performance, reliability, operating model and security needs.

To shed light on these issues, in 2020 the IFC decided to **update its 2015 survey on central banks' use of and interest in big data**. The purpose of this was to revisit central banks' experience and progress in terms of their use of large data sets and related new techniques. Specifically, the IFC aimed at covering four main issues. First, how central banks define big data, and the sources they draw upon. Second, whether they already use such information, the main applications involved, and the purpose of their work. Third, the challenges they face and that hinder a more widespread use of big data. And fourth, their future plans for using big data in terms of the specific projects considered and the role that central bank cooperation could play in this regard. As an annex, the survey also examined in greater detail how central banks use large payment data sets, an area characterised by rapid innovation.

This report outlines the **main conclusions of the IFC survey**. Section 2 summarises how central banks define big data, confirming that their approach usually encompasses both unstructured and large, structured data sets. Section 3 documents the type of interest that central banks have in using big data. Section 4 describes the main applications reported to support their current big data work. Section 5 reviews future prospects for big data usage at central banks, including expected challenges and the initiatives planned to address them. Section 6 discusses the specific projects envisaged by central banks in the near term and the benefits that international cooperation can bring.

2. What is big data for central banks?

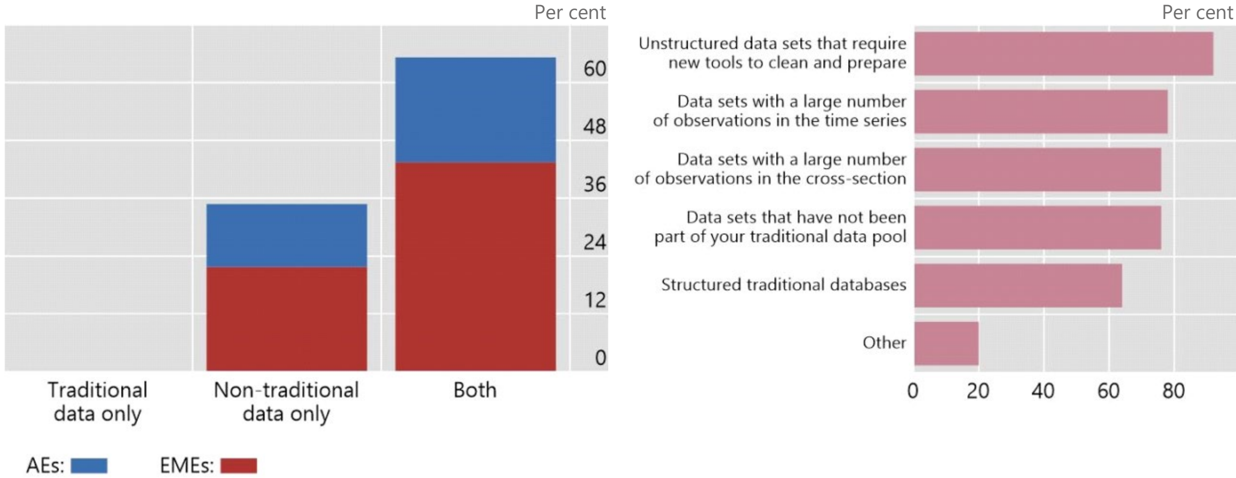
Big data is often defined by the so-called three "Vs" – that is, high volume (eg number of records and attributes), velocity (speed of data production eg tick data) and variety (eg structure and format of the data set) (Laney (2001)). The reality is more complex in practice, and **big data can include the information generated by several processes** such as social media, web-based activities, machine sensors, or financial, administrative or business operations. This comprehensive view of big data is confirmed by the survey results for both AEs and EMEs. Certainly, around one third of central banks believe that the concept of big data only includes non-traditional data (Graph 1, left-hand panel). But, for two thirds of the respondents, it should also include "traditional" yet large data sets. Data collected for administrative or regulatory/supervisory purposes and that are often referred to as "financial big data" (Cœuré (2017), Draghi (2018)) are a case in point; these data sets are often well structured by design.

According to the survey, a comprehensive definition of big data would cover all types of **data sets that require non-standard technologies in order to be analysed**. The reason for this is that traditional statistical techniques such as descriptive analysis, inductive statistics (eg econometrics) or non-parametric analysis face limits when applied to large data sets. The difficulties are even more obvious when dealing with unstructured data, for instance when these are stored as text or images; analysing them requires extracting information that can be turned into structured data, as done for instance with natural language processing algorithms to process human language in a numerical way.

Central bank definitions of big data and main sources Graph 1

How does your institution define big data?[1]

Big data sets in central banks[2]



[1] The graph reports the share of respondents that selected each respective answer to the question "How does your institution define big data?". Respondents could select multiple options. Specifically, 35% of respondents consider only non-traditional data as big data. Non-traditional data include "unstructured data sets that require new tools to clean and prepare", "data sets with a large number of observations in the time series", "data sets with a large number of observations in the cross-section", and "data sets that have not been part of your traditional pool". The remaining 65% additionally consider "structured traditional databases" as big data. [2] Percentage of central banks.

Sources: IFC big data survey (2020); authors' calculations.

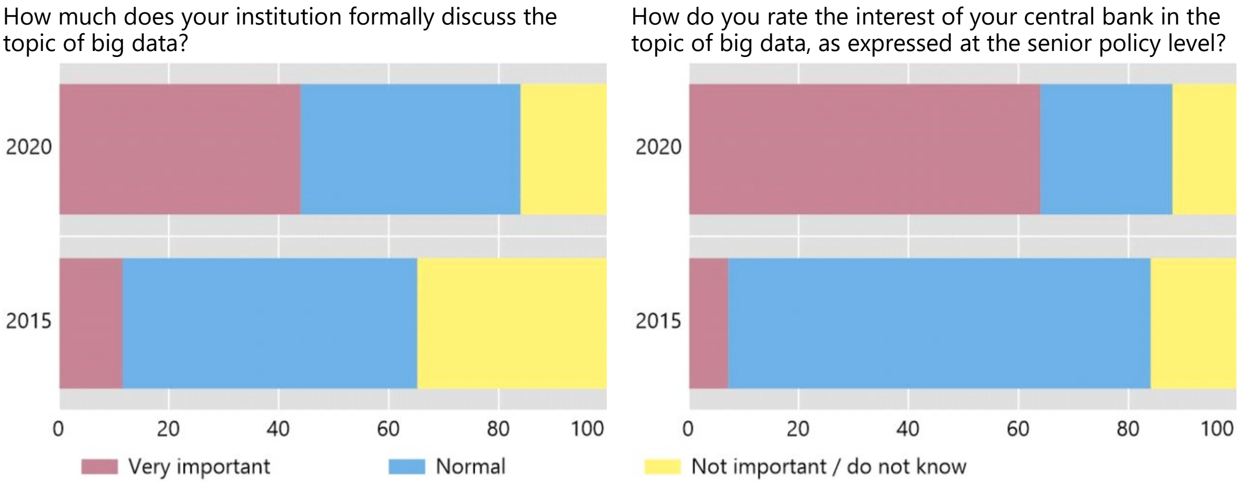
The above approach allows **central banks to differentiate between several distinct categories of big data sources**. The first one, mentioned by more than 90% of the respondents, comprises the **unstructured data sets**, such as text messages (eg social media), images scraped from the internet, and information sent by sensors and other connected devices (Graph 1, right-hand panel). This type of information – "the internet of things" – may not necessarily be large, but it is complex and cannot be easily managed using traditional statistical techniques tailored to numerical data sets. In particular, it requires specific tools in order to be cleaned and properly prepared, and, in many instances, one has to acquire these data from private providers, already aggregated and organised. Three important examples are worth mentioning. First, mobility reports, which provide aggregate commuting trends obtained through GPS and which were able to support the monitoring of households' access to workplaces and recreation areas when the Covid-19 pandemic struck in 2020 (De Beer and Tissot (2020)). The second example relates to internet searches, such as Google trends, that can be used to ascertain certain economic factors – for instance, expectations on labour market dynamics (Doerr and Gambacorta (2020a,b)). A third source of unstructured information for central banks is text in printed format, such as newspaper articles, firms' financial statements, official press releases etc.

A second category of big data, referred to by almost 80% of the respondents, relates to **data sets with a very large number of observations in the time series**. This typically includes large, structured financial big data sets. There are other types of data sets that are, in contrast, brand new, especially those that have been collected in the context of post-GFC regulatory reforms: a good example relates to individual derivatives transactions now being reported to TRs, which have led to the compilation of very large databases with multiple attributes and a high frequency.

Two other specific categories of structured data sets were also highlighted by more than three fourths of central banks. One relates to data that are not new but were not considered part of the traditional statistics analysed by central banks in the past. Payment transactions are a good example, since these data were primarily collected for market monitoring purposes; in recent years, central banks have increasingly tried to use them more effectively in order to conduct economic analyses as well. A second specific category relates to cross-sectional data sets, which provide observations for the entire population of interest, collected with multiple attributes at one point in time. Credit registries, which collect loan-by-loan data sets, are a case in point: their information is not really new, having been collected for many years, but available innovative IT tools and techniques are easing their analysis. In the past, such information was mainly collected through well defined panel surveys. Expanded IT capabilities now allow for the collection of much richer databases covering the entire population of interest (eg census exercises) at different points in time.

Interest in big data

In per cent of respondents | Graph 2



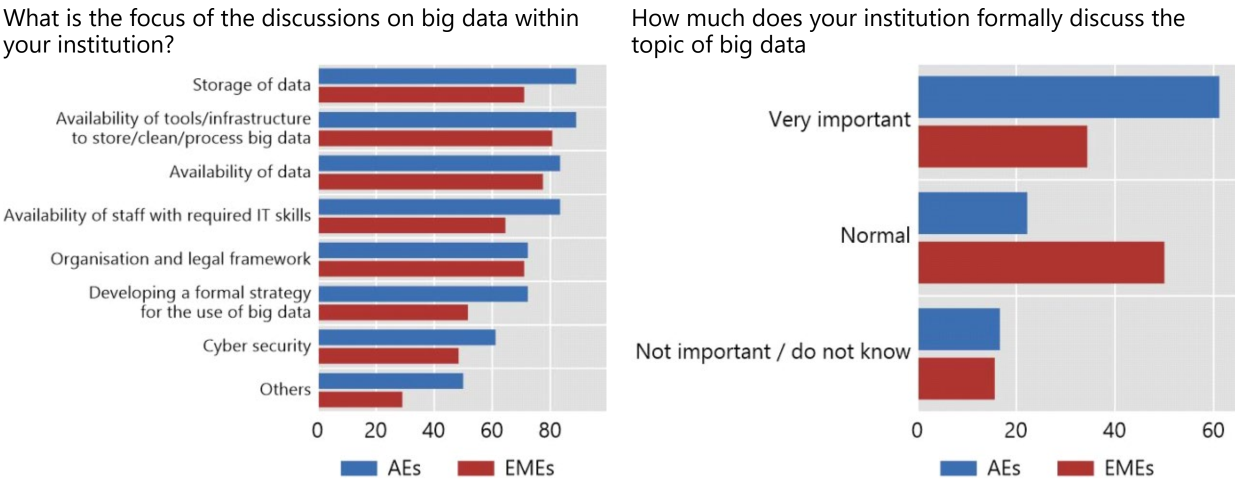
Sources: IFC big data survey (2020); authors' calculations.

3. Central banks' interest in big data

Central banks' interest in big data has significantly increased in recent years, in particular in comparison to the 2015 survey (Graph 2, left-hand panel), and it is now widespread. The topic of big data currently occupies a very important position in formal deliberations held at nearly 45% of the central banks surveyed (up from 12% in 2015). Consistent with that, only 15% of respondents report that this topic is still not discussed in their institution or is only lightly discussed (down from 35% in 2015). Interest has been rising even faster at a senior policy level (right-hand panel). Nearly 65% of the central banks surveyed, senior officials are reported to have expressed strong interest in big data, up from less than 10% in 2015.

Big data discussions across advanced and emerging market economies

In per cent of respondents | Graph 3



Sources: IFC big data survey (2020); authors' calculations.

The **focus of central banks' internal discussions** on big data is quite broad (Graph 3, left-hand panel). A vast majority of the respondents (from about 70% to 90% of them) report that these discussions deal with data storage issues, staff IT skills, the availability of IT infrastructure (hardware and software), data access, and organisational and legal topics. In contrast, the formulation of a data strategy as well as cyber security aspects appear slightly less prominently on their agendas.

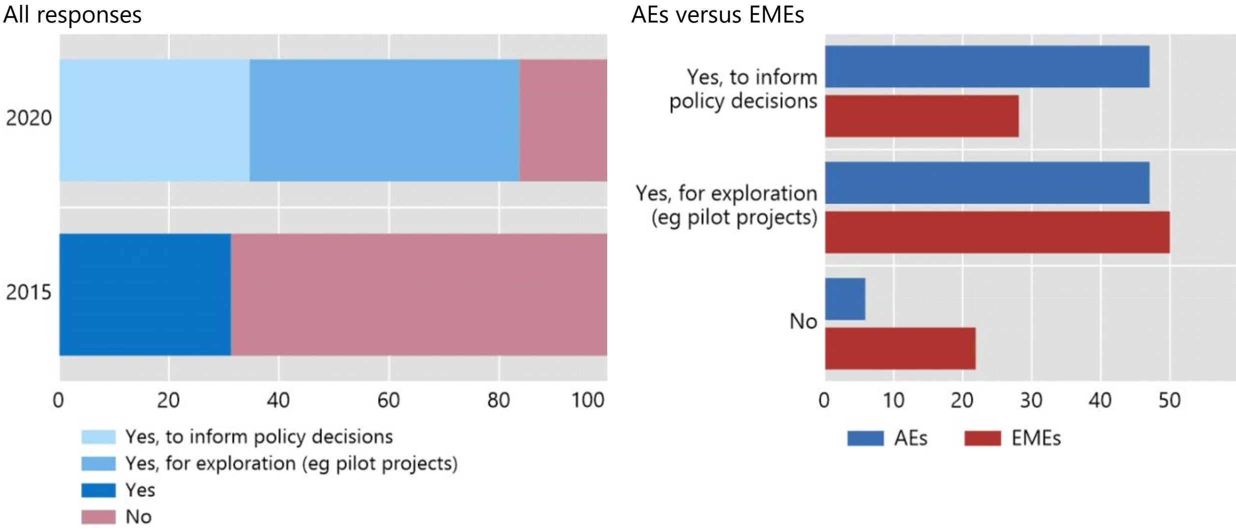
However, the picture is somewhat different **depending on level of economic development**. For instance, while big data discussions are rated "very important" by almost two thirds of AEs' central banks, they are rated "normal" by half of the respondents in EMEs (Graph 3, right-hand panel). Moreover, issues of interest differ slightly. In AEs, three topics garner the most attention: how to store data, human capital issues and infrastructure availability (left-hand panel). In contrast, discussions in EMEs focus primarily on infrastructure and data availability. These differences may reflect different stages of big data development in central banks: while many of those located in AEs are already actively exploring how to exploit big data, several of their counterparts in EMEs appear to still be struggling to obtain basic access to big data.

4. Central banks' work with big data

Use of big data by central banks has markedly increased relative to 2015 (Graph 4, left-hand panel). More than 80% of the central banks surveyed now use big data sources to support their work, up from 30% five years earlier. Certainly, half of the respondents use them for exploratory purposes only, eg conducting pilot projects. Nevertheless, big data work is already operational, informing policy in one third of the cases. Usage is more frequent across AEs (right-hand panel), where decisions are made based on big data sources in almost half of the jurisdictions. In contrast, one fifth of respondents in EMEs do not use this information at all.

Are you currently using any big data sources?

In per cent of respondents Graph 4



Sources: IFC big data survey (2020); authors' calculations.

Data sources supporting main central bank functions

When using big data, **central banks rely on a large variety of sources**, consistent with their comprehensive view of the concept of big data (Section 2). These data sources are shown in Graph 5, where the most frequently used ones are depicted with larger words. The first main category includes unstructured data sets: many central banks already extract text from newspapers through natural language processing – for example, to quantify qualitative factors (eg impact of sentiment and uncertainty about economic developments), and also make use of internet-based information (eg search queries).[5] A second, almost equally important category comprises financial big data sets that are quite complex to deal with because of their high granularity – for instance, credit registries or payment data collected on a transaction-level basis.

[5] See Sawaengsuksant (2019) for a discussion of related use cases.

Main big data sources

Word count on sources[1] Graph 5

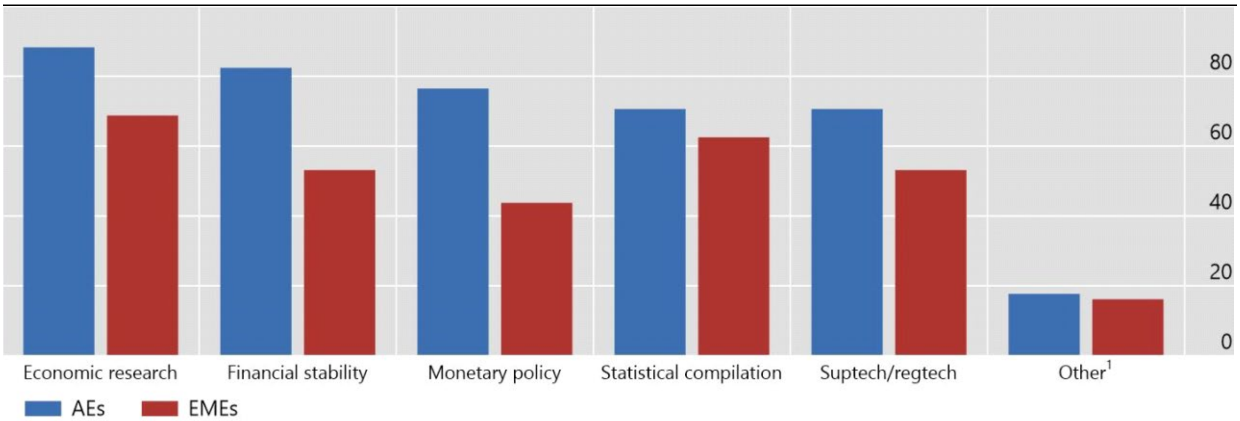
[1] Open answers are transformed by removing special characters, white spaces or stop words (such as “the”, “a” or “we”). A text-mining algorithm then counts the frequency of individual words. Words mentioned more frequently appear larger.

Sources: IFC big data survey (2020); authors’ calculations.

Turning to the **concrete projects being undertaken** to work with big data, the survey shows that they are already supporting most of central banks’ main functions (Graph 6).[6]

For what general purposes does your institution use big data?

In per cent of respondents Graph 6



[1] Includes monitoring crypto-assets, cyber security, impact and network analysis.

Sources: IFC big data survey (2020); authors’ calculations.

[6] See IFC (2019b,c) for detailed presentations of big data applications in central banking.

Specifically, around three fourths of reporting central banks use big data as an input for economic research. About 60% also exploit such information to support financial stability and suptech- (to supervise financial institutions) and regtech-related (to facilitate compliance with regulatory requirements) microfinancial work as well as statistical compilation. Perhaps more surprisingly, big data usage is slightly lower for the traditional central bank mandate of monetary stability (mentioned by about half of the respondents).

In general, big data usage cases are somewhat more frequent among respondents in AEs (95%), compared with in EMEs (80%). This is particularly the case with projects supporting central banks' two key policy areas, eg monetary and financial stability. In contrast, the use of big data for statistical compilation purposes is reported to be equivalent between AEs and EMEs.

Four major types of applications

The big data projects undertaken by central banks involve **four main types of applications**: natural language processing; nowcasting exercises; applications to extract economy-wide insights from granular financial big data sets; and suptech and regtech applications.

The first type relates to processing textual information through **natural language processing**. The goal is generally to collect qualitative text-based information and summarise it quantitatively. A good example is the computation of so-called economic policy uncertainty (EPU) indices to assess the degree of uncertainty faced by economic agents.[7] Such indices are basically constructed by setting up dictionaries that allow for the definition of specific terms that refer to uncertainty, and then searching them in the text considered (for instance in newspaper articles). These selected terms are then counted and aggregated in order to provide a synthetic index that reflects the degree of uncertainty displayed in the document of interest. Another example of central banks' applications of big data in order to process human language relates to policy evaluation. For instance, one can quantify the monetary policy stance that is communicated to the public via the publication of meeting minutes. Similarly, market expectations of interest rate decisions can be assessed by analysing market commentaries ahead of policy meetings (Andhika Zulen and Wibisono (2019)). Such exercises can be updated frequently, which is a key advantage compared to the more traditional surveys of market participants that are conducted by several central banks. The information collected on market expectations can also be particularly useful when future markets are not well developed or lack liquidity. In contrast, reported use of text data to inform financial stability policies has been relatively scarce so far, although it appears to be developing as well.[8]

The second main area relates to **nowcasting exercises** to produce high-frequency analyses of the economic situation in "real-time". Up to one third of the respondents mention that big data is used for this purpose, especially to provide inflation and economic growth estimates sooner and/or more frequently (Graph 7,

7 See Baker et al (2016) and www.policyuncertainty.com/.

8 Correa et al (2020) measure sentiment in financial stability reports using natural language processing. In doing so, they construct a dictionary of financial stability terms by examining more than 1,000 financial stability reports.

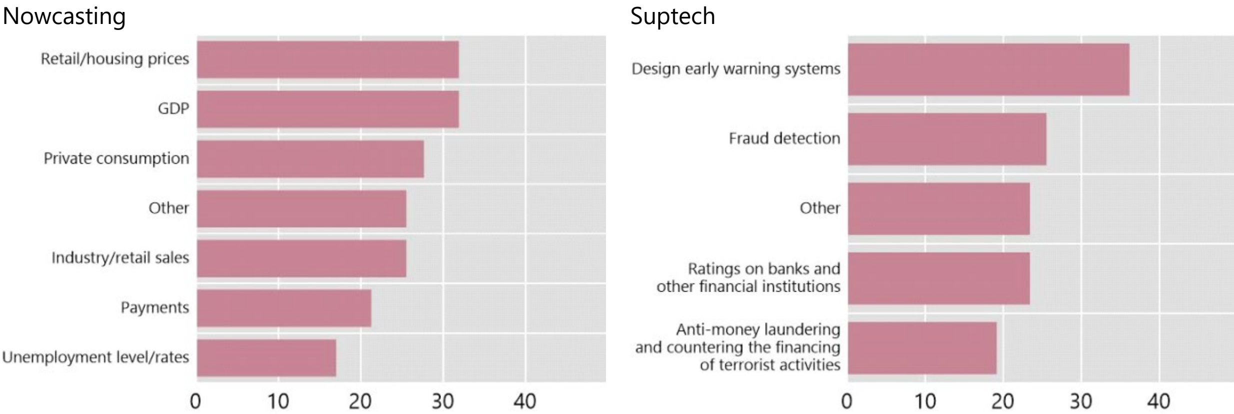
left panel). Their time horizon varies substantially, ranging from the compilation of weekly indices (eg production of weekly indicators of economic activity; Lewis et al (2020)) to advance estimates of upcoming monthly statistical releases (eg next monthly industrial production or CPI numbers) and GDP growth forecasts with a higher time lead – say, for the current quarter or the next one or two. Moreover, nowcasting models can help fill statistical gaps, eg when reference series do not exist or are available only at a low frequency or are suddenly disrupted, as during the Covid-19 pandemic. Real estate markets constitute a case in point, as data are often lacking in official statistics, while house prices can be sourced from the internet relatively easily (Oehler (2019)). Another example of possible use cases relates to the less developed economies, where the provision of intra-annual official statistics can be scarce.

A key ingredient of these big data-based nowcasting applications is the **simultaneous use of many information sources**, combining different frequencies and types. The models will typically start with the use of traditional "hard" economic indicators, such as electricity consumption. Some of these statistics may be scraped directly from the web, such as online retail sales and prices (an approach reported by 25% of the central banks to predict inflation dynamics). Other, less structured information can also be called in. Textual information, like newspaper articles, can be used to produce economic growth forecasts (Kalamara et al (2020)) and assess labour market conditions (Bailliu et al (2019)). So do internet search queries like Google Trends (Sawaengsuksant (2019)), for instance to nowcast unemployment (D'Amuri and Marcucci (2017)) or predict car sales, a prime component of private consumption (Nymand-Andersen and Pantelidis (2018)). Usually, these nowcasting exercises are frequently updated as new data come in, and various techniques – eg Lasso (Least Absolute Shrinkage and Selection Operator) – will be applied to select the combination of series which maximises the forecast at a given point in time (Richardson et al (2019)). One advantage is that this approach does not rely on specific relationships estimated ex ante (as is the case of bridge models used for "traditional" nowcasting exercises) and may be better suited to identifying turning points, especially at times of economic upheaval (INSEE (2020)).

For what specific purposes does your institution use big data?

In per cent of respondents

Graph 7



Sources: IFC big data survey (2020); authors' calculations.

A third category includes the various **applications developed by central banks to extract economy-wide insights from granular financial big data sets**, with the primary objective of supporting macro stabilisation policies. These applications basically aim to exploit the large and structured micro data sets collected for regulatory and statistical purposes. For example, TRs' trade data can be used in a number of different ways, such as network analysis to identify interconnections between market participants as well as systemic risk. Another example relates to credit registries, which can be used for detailed credit quality assessments – eg estimates of default probabilities or loss-given-default (Petropoulos et al (2019)). And a last example relates to data on individual payment transactions (Gil et al (2018)), which have the advantage of being available quickly and often, not least because of the surge in digital transactions seen in recent years (Box 1). These data can shed interesting light on the functioning of the financial system, helping to monitor real-time payments settlements. They are also increasingly used to support nowcasting exercises, as mentioned by around one fifth of the respondents (Graph 7, left-hand panel).

The fourth main category comprises the wide range of **suptech and regtech applications** to support micro-supervisory policies. This can cover multiple tasks, as documented by Broeders and Prenio (2018), di Castri et al (2019), Coelho et al (2019) and Financial Stability Board (2020). In general, many of these applications focus on micro-level risk assessment. For instance, firm-level information gathered from financial statements or newspapers can be used to support early warning exercises or enhance credit scoring (mentioned by one third and one fifth of the respondents, respectively; Graph 7, right-hand panel). Another important area relates to fraud detection (one fourth of the cases) – for instance, by screening credit contracts for suspicious terms and conditions in order to enhance consumer protection. Lastly, almost one fifth of central banks deploy big data algorithms for anti-money laundering/combating the financing of terrorism (AML/CFT) purposes – for instance, when analysing payment transactions to identify suspicious patterns.

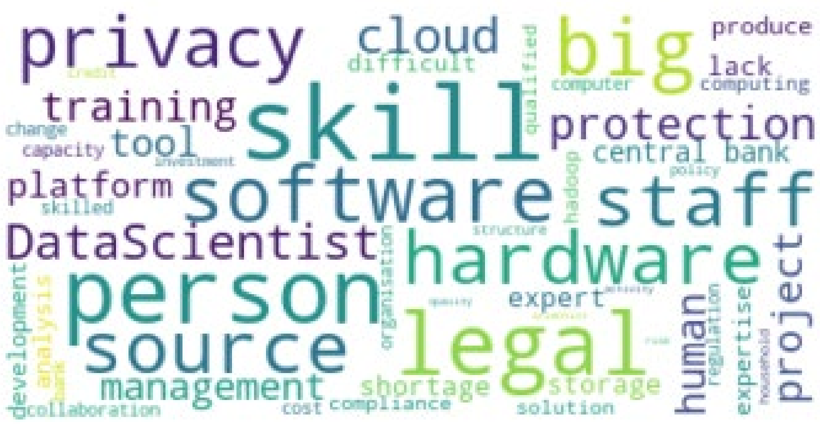
5. Challenges

A key insight from the survey is that **several challenges have been hampering a more widespread use of big data** by central banks. These challenges are shown in Graph 8, where the most frequently mentioned ones are depicted with larger words. Four main binding constraints can be highlighted: the need for an adequate IT infrastructure; heavy data processing work; human capital bottlenecks; and reputational, legal and ethical considerations.

First, a **reliable and high-powered IT infrastructure** is needed to exploit big data, requiring large investments in hardware and software. Many central banks are already exploring how to set up big data platforms (IFC (2020b)), but progress has been rather slow so far, for several reasons. For one, central banks have to make do with legacy equipment: their data storage facilities and the hardware and software they use to compile traditional statistics may not be useful for handling big data sets. But setting up a new infrastructure can be highly costly, and central banks need to mobilise the necessary resources without compromising the production of traditional statistics. For instance, a key issue for central banks is whether to stick to the Structured Query Language (SQL) used to maintain their historical data processes based on a "traditional" relational system, or to instead switch to a brand new platform that offers the possibility of storing non-structured data.[9] The fragmentation of solutions for handling big data constitutes an additional constraint, reinforced by the fact that the technological landscape remains heterogeneous and fast-changing. An important aspect is that central banks have to manage the proprietary tools developed by commercial vendors and also be able to use the growing number of applications developed in the open source community.[10] The institution risks ending up with a very heterogeneous IT infrastructure composed of various technologies that have different life cycles and that are not fully compatible. Another key 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, but they can entail security risks, most notably that leakages of confidential data may have high reputational costs for central banks.

Big data challenges

Word count on challenges | Graph 8



[1] Open answers are transformed by removing special characters, white spaces or stop words (such as "the", "a" or "we"). A text-mining algorithm then counts the frequency of individual words. Words mentioned more frequently appear larger.

Sources: IFC big data survey (2020); authors' calculations.

Second, and apart from the IT challenges discussed above, specific **data-processing problems** constrain the use of big data by central banks. This is particularly the case for the large and granular financial big data sets collected for administrative and regulatory purposes. In practice, these often reveal important quality issues such as multiple outliers, missing observations, duplicates and other anomalies. Substantial cleaning is thus needed before this information can be

[9] A growing number of central banks are choosing to invest in so-called "not only" SQL (NonSQL) databases, which can also handle non-structured data; see IFC (2020b).

[10] The decision often involves a trade-off, as open source software does not entail licence fees, but does entail other costs (eg the need to develop in-house support or buy it from an external company).

effectively used to support policy work – and keep the data lakes that are being set up from ending up as "data swamps".[11] The process of merging databases and matching individual observations across domains may also be prone to error and resource intensive. A key lesson from the survey is that such problems can be handled effectively through dedicated big data techniques. For one, reporting through regtech applications can substantially enhance the quality of the data submitted by reporting agents (eg commercial banks for supervisory monitoring exercises), for example by setting up automatic checks. Additionally, various tools can help automatically clean big data sets as well as matching databases, reducing the need for manual intervention. In particular, a significant number of central banks report using ML technologies for outlier detection and correction based on "imputation techniques" (Kwon (2019)) as well as for probabilistic record linkage ("fuzzy matching") (Pinto et al (2019)). This can be achieved using simple algorithms designed to incorporate experts' judgment in a "supervised" way – see for instance the ECB process set up to deal with TRs' data (Perez-Duarte and Skrzypczynski (2019)). More sophisticated, "unsupervised" algorithms are also being explored in this endeavour (Benatti (2019)).

Third, central banks need to **build up human capital** in order to exploit big data. Setting up and maintaining big data platforms requires a specific type of skillset, combining statistical, IT, and analytical/mathematical aspects. Yet the supply of "data scientist" capabilities is scarce and in high demand (Cœuré (2020)), especially in the private sector. One solution is for central banks to train existing staff, but learning the new techniques that are needed can require significant time and effort. For instance, while most of central banks' traditional work has relied on statistical inference based on a few software packages (eg Stata, EViews, Matlab), big data sets have to be handled using other techniques (eg sophisticated ML algorithms) and software (eg Python, R). In addition, the experience of central banks shows that these skill adjustments take place not only at the operational level, eg the statisticians in charge of using advanced tools; those analysing the output of complex models must also have a good understanding of new techniques in order to ensure that big data predictions are not only accurate but also representative and "interpretable" – so that specific explanatory causes or factors can be identified and communicated for policy use. Another issue is attracting and retaining talent, esp. in the face of intense competition from the private sector. This may also call for a review of existing public compensation schemes, career systems and internal hierarchical organisations in central banks.

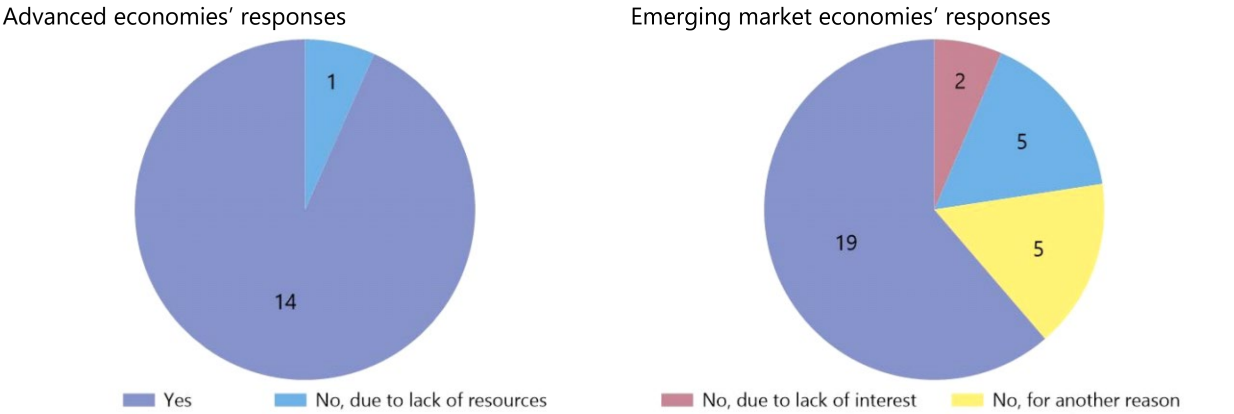
Fourth, important **reputational, legal and ethical aspects** can limit the use of big data sources by central banks (Tissot (2019)). Reputational aspects may particularly hinder the use of information sourced from the internet when little is known about its accuracy. For instance, internet-based indicators such as search queries and messages on social media may not be representative of the real economy – not everybody is on Twitter, or only a subset of the CPI basket prices can be scraped from the web. Similarly, ML algorithms trained on preclassified and/or potentially biased data sets may be of little help when new data arrive. As a result, the relationship that seems to exist between unstructured data and a certain

11 TRs' data (ie derivatives transactions) constitute a case in point (IFC (2018)). Reporting requirements are particularly complex and demanding, so the two counterparties in a trade may report it very differently, leading to double-counting and other quality issues.

phenomenon may unexpectedly deteriorate when additional information arrives (eg the incorporation of new, "out-of-sample" information).[12] These issues are particularly important for central banks, since inaccurate information could lead them to make inadequate policy decisions and in turn undermine their reputation and credibility. As a result, and in contrast to what can be seen in the private sector, big data information is still used very cautiously. For instance, central banks observing a surge in inflation using online prices may prefer to stick to well established price surveys when making monetary policy decisions. Another related issue is that sound professional and scientific standards should be adhered to before new information can legitimately be incorporated into public policy processes, in line with the Fundamental Principles for the production of appropriate and reliable official statistics.[13] A last issue are legal and ethical aspects, especially because they can hinder the use of big data compiled by commercial providers eg through social media and electronic devices. Central banks may not have access to the data collected by the companies who own them, or they may be reluctant to access these private data if they are collected from citizens who value privacy protection. This can also clash with the rising pressures faced by central banks to publish their data and develop public open data access.

Is your institution planning to start any big data-related projects in 2020/21?

Number Graph 9



Sources: IFC big data survey (2020); authors' calculations.

6. Looking forward: the benefits of cooperation

The survey confirmed that **big data usage is bound to increase** despite the many challenges faced by central banks. The vast majority of respondents (more than 70%) were planning to conduct a big data project in 2020/21. This is particularly the case

[12] The failure of Google Flu Trends (GFT) provides a good example of these perils, as it was initially intended to provide estimates of influenza activity based on Google Search queries but was discontinued in the mid-2010s (Lazer et al (2014)).

[13] Such as quality assurance processes and methodological transparency; see https://unstats.un.org/unsd/dnss/gp/FP-Rev2013-E.pdf.

for AEs, of which almost all respondents (90%; Graph 9, left-hand panel) were willing to do so. By contrast, this was the case for only 60% of central banks in EMEs; the primary reason was not any absence of interest, mentioned by only 6% of EME respondents, but rather a lack of resources (16%; Graph 9, right-hand panel).

A second important insight from the survey is that **central banks are willing to join forces to reap the benefits of big data**. Indeed, half of them reported an interest in collaborating on one or more specific projects, with three types of cooperation envisaged. First, by sharing knowledge among those institutions that have developed specific expertise that can be used again in other jurisdictions. The topics involved can be quite diverse, including general big data techniques (eg data visualisation, network analysis, ML tools), more general information management issues (eg development of open source coding, data sharing protocols, encryption and anonymisation techniques for using confidential data) as well as certain applications that are more specific to the central bank community (eg suptech and regtech areas). Second, by using big data to work on global issues such as international spillovers, global value chains and cross-border payments; by definition, such work depends on sufficient international cooperation, for instance to share information across countries. Third, by developing joint exploratory projects to benefit from economies of scale and collectively share (limited) financial and human resources. Such envisaged projects cover, in particular, the use of complex language processing technologies and ML algorithms, for instance to exploit payments data (Box 1).

Box 1

Big payments data

Digital payment settlement transactions[(1)] have experienced rapid growth, pushed by the information revolution and financial innovation (IFC (2020a), Bech and Hancock (2020)). The footprint that these transactions leave includes a wealth of details, including the type of instrument, payment purpose or the counterparties involved, which can be easily collected very frequently. As a result, payments data now constitute one of the major big data sources available to central banks.

This new information brings vast analytical opportunities with it, as underlined by the survey; in particular, two thirds of reporting central banks think that all types of payments data can be useful for their work (Graph A1, left-hand panel). Many are already exploiting them for analytical purposes, eg to analyse private consumption (Gil et al (2018)). They also use big data techniques to identify fraudulent transactions as well as systemic risks arising from interconnectedness among financial market participants.

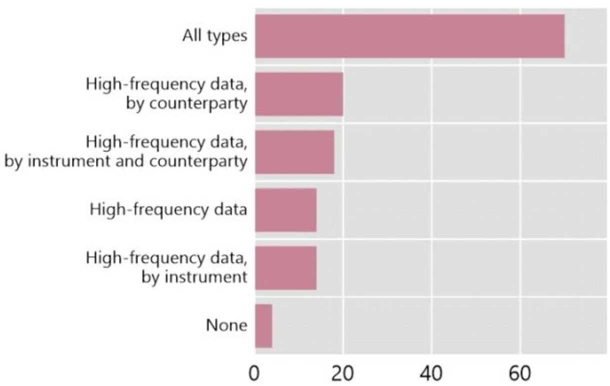
Central banks also value international cooperation to exploit payment data through joint projects. Close to 90% of the respondents are willing to contribute to pilot studies using this information for a variety of purposes (Graph A1, right-hand panel). Preference for nowcasting exercises is particularly strong (40% of the cases reported interest), followed by projects aimed at enhancing surveillance (30%).

Payments data: interest of central banks

Graph A1

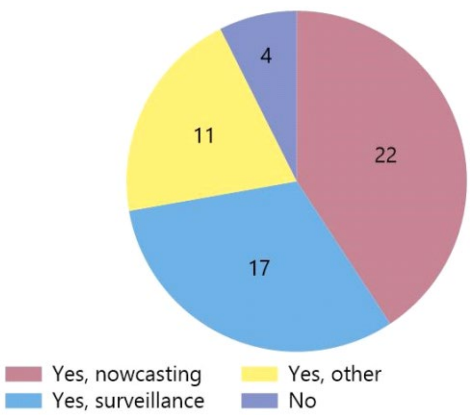
Which types of payments data are useful for your institution?

Per cent



Would your institution be willing to contribute to a pilot study on the use of payments data?

Number



Source: IFC big data survey (2020).

(1) Such transactions refer to the actual movement of money, often among financial institutions, to close pending transactions. Digital payments are also invading retail payments, but cash remains important in several countries (Boar and Szemere (2020)).

The reasons for cooperation expressed by central banks are quite diverse. One is that most central banks face similar problems, either when setting up big data infrastructures, building up human capital or developing the right algorithms.[14] Exchanges of views can therefore help to identify best practices and avoid the so-called "NIH (not invented here) syndrome" when different institutions seek to address problems in parallel (IFC (2020b)). This can be particularly helpful for those that are still in the incipient stages of the big data journey and have to make strategic IT decisions, for instance when choosing between cloud computing and on-premises solutions, deciding to rely on open source vs proprietary products, or designing training programmes to hire and retain talent. A second benefit of cooperation is that early adopters of big data work can develop technical assistance programmes based on their experience; a large and growing number of central banks have indeed set up such programmes as part of their international outreach. And a third benefit is that cooperation can facilitate data sharing across borders and thereby enhance data availability to support policy (Buch (2019)). This has been, for instance, a key objective pursued in the context of the G20-endorsed Data Gaps Initiative, especially its second phase (Recommendation # II.20 "Promotion of data sharing by G-20 economies";

[14] For example, natural language processing applications to gauge sentiment, supervised/unsupervised learning algorithms to clean data, and nowcasting models to support macroeconomic policies.

FSB- IMF (2015)) as well as that of OECD projects on enhancing access to and sharing of data (EASD).[15]

International financial institutions can greatly support these cooperative approaches. They can facilitate innovation by promoting technological solutions for harmonising data standards and processes among jurisdictions, for instance with the development of the Statistical Data and Metadata exchange (SDMX) standard, developed by the members of the Inter-Agency Group on Economic and Financial Statistics (IFC (2016)), and the work of the International Network for Exchanging Experience on Statistical Handling of Granular Data (INEXDA) (IFC (2019d)). Knowledge can be shared through regular meetings and the sharing of experience and data from pilot projects, possibly organised under the umbrella of the BIS and the Basel-based committees, including the IFC. Turning to more specific projects, the BIS Innovation Hub has been established in order to identify and develop insights into critical trends in financial technology relevant to central banks, explore the development of public goods to enhance the functioning of the global financial system, and serve as a focal point for a network of central bank experts on innovation. Such a network could undoubtedly play an important role in facilitating international cooperation to exploit big data sources and techniques.

15 See "Data governance: enhancing access to and sharing of data".

References

Andhika Zulen, A and O Wibisono (2019): "Measuring stakeholders' expectation on central bank's policy rate", *IFC Bulletin*, no 49, January.

Bailliu, J, X Han, M Kruger, Y Liu and S Thanabalasingam (2019): "Can media and text analytics provide insights into labour market conditions in China?", *IFC Bulletin*, no 49, January.

Baker, S, N Bloom and S Davis (2016): "Measuring economic policy uncertainty", *Quarterly Journal of Economics*, vol 131, no 4, pp 1593–636.

Bech, M and J Hancock (2020): "Innovations in payments", *BIS Quarterly Review*, March, pp 21–36.

Benatti, N (2019): "A machine learning approach to outlier detection and imputation of missing data", *IFC Bulletin*, no 49, January.

Boar, C and R Szemere (2020): "Payments go (even more) digital", *Analysis of the 2019 Red Book Statistics*, November.

Broeders, D and J Prenio (2018): "Innovative technology in financial supervision (suptech) – the experience of early users", *FSI Insights*, no 9, July.

Buch, C (2019): "Building pathways for policy making with big data", welcoming remarks, International Seminar on Big Data, Bali, 26 July.

Chakraborty, C and A Joseph (2017): "Machine learning at central banks", Bank of England, *Working Papers*, no 674.

Coelho, R, M De Simoni and J Prenio (2019): "Suptech applications for anti-money laundering", *FSI Insights,* no 18.

Cœuré, B (2017): "Policy analysis with big data", speech at the conference on Economic and Financial Regulation in the Era of Big Data, Paris, 24 November.

——— (2020): "Leveraging technology to support supervision: challenges and collaborative solutions", speech at the Financial Statement event series, Peterson Institute, 19 August.

Committee on Payments and Market Infrastructures and World Bank Group (CPMI-WB) (2020): "Payment aspects of financial inclusion in the fintech era", *CPMI Papers*, no 191.

Correa, R, K Garud, J Londono and N Mislang (2020): "Sentiment in central banks' financial stability reports", *Review of Finance*, April.

D'Amuri, F and J Marcucci (2017): "The predictive power of Google searches in forecasting unemployment", *International Journal of Forecasting*, vol 33, no 4, pp 801–16.

De Beer, B and B Tissot (2020): "Implications of Covid-19 for official statistics: a central banking perspective", *IFC Working Papers*, no 20, November.

di Castri, S, S Hohl, A Kulenkampff and J Prenio (2019): "The suptech generations", *FSI Insights on policy implementation*, no 19, October.

Doerr, S and L Gambacorta (2020a): "Identifying regions at risk with Google Trends: the impact of Covid-19 on US labour markets", *BIS Bulletin,* no 8, April.

——— (2020b): "Covid-19 and regional employment in Europe", *BIS Bulletin*, no 16, May.

Doerr, S, L Gambacorta and J M Serena (2021): "Big data and machine learning in central banking", *BIS Working Papers*, forthcoming.

Draghi, M (2018): "Welcome remarks at the third annual conference of the ESRB", September.

Financial Stability Board (FSB) (2017): *Artificial intelligence and machine learning in financial services – market developments and financial stability implications*, November.

——— (2020): *The use of supervisory and regulatory technology by authorities and regulated institutions: market developments and financial stability implications*, October.

Financial Stability Board and International Monetary Fund (FSB-IMF) (2009): *The financial crisis and information gaps*.

——— (2015): *The financial crisis and information gaps – sixth implementation progress report of the G20 Data Gaps Initiative*.

Gil, M, J Pérez, A J Sánchez and A Urtasun (2018): "Nowcasting private consumption: traditional indicators, uncertainty measures, credit cards and some internet data", *IFC Bulletin*, no 50.

Groves, R (2011): "'Designed data' and 'organic data", in the Director's Blog of the US Census Bureau, www.census.gov/newsroom/blogs/director/2011/05/designed-dataand-organic-data.html.

Institut national de la statistique et des études économiques (INSEE) (2020): "High-frequency' data are especially useful for economic forecasting in periods of devastating crisis", *Point de Conjoncture*, June, pp 29–34.

Irving Fisher Committee (IFC) (2015): "Central banks' use of and interest in 'big data", *IFC Report*, no 3, October.

——— (2016): "Central banks' use of the SDMX standard", *IFC Report*, no 4, March.

——— (2017): "Big data", *IFC Bulletin*, no 44, March.

——— (2018): "IFC report on central banks and trade repositories derivatives data", *IFC Report*, no 7, October.

——— (2019a): "Business intelligence systems and central bank statistics", *IFC Report,* no 8, October.

——— (2019b): "Are post-crisis statistical initiatives completed?", *IFC Bulletin*, no 49, December.

——— (2019c): "The use of big data analytics and artificial intelligence in central banking", *IFC Bulletin*, no 50, December.

——— (2019d): "INEXDA – the Granular Data Network", *IFC Working Paper*, no 18, October.

——— (2020a): "Central banks and fintech data issues", *IFC Report*, no 10, February.

——— (2020b): "Computing platforms for big data analytics and artificial intelligence", *IFC Report*, no 11, April.

Kalamara, E, A Turrell, C Redl, G Kapetanios and S Kapadia (2020): "Making text count: economic forecasting using newspaper text", Bank of England, *Staff Working Papers*, no 865.

Kwon, B (2019): "Imputation for missing data through artificial intelligence", *IFC Bulletin*, no 49.

Laney, D (2001): "3D data management: controlling data volume, velocity, and variety", META Group (now Gartner).

Lazer, D, R Kennedy, G King, and A Vespignani (2014): "The parable of Google Flu: traps in big data analysis", *Science*, vol 343 (6176), pp 1203–05, 14 March.

Lewis, D, K Mertens and J Stock (2020): "Monitoring real activity in real time: the weekly economic index", Federal Reserve Bank of New York, *Liberty Street Economics*, 30 March.

Mehrhoff, J (2019): "Demystifying big data in official statistics – it's not rocket science!", *IFC Bulletin,* no 49.

Nymand-Andersen, P (2016): "Big data: the hunt for timely insights and decision certainty", *IFC Working Paper*, no 14.

Nymand-Andersen, P and E Pantelidis (2018): "Google econometrics: nowcasting euro area car sales and big data quality requirements", *ECB Statistics Paper Series*, no 30, November.

Oehler, S (2019): "Developments in the residential mortgage market in Germany – what can Google data tell us?", *IFC Bulletin,* no 49.

Pérez-Duarte, S and G Skrzypczynski (2019): "Two is company, three's a crowd: automated pairing and matching of two-sided reporting in EMIR derivatives' data", *IFC Bulletin*, no 49.

Petropoulos, A, V Siakoulis, E Stavroulakis and A Klamargias (2019): "A robust machine learning approach for credit risk analysis of large loan-level datasets using deep learning and extreme gradient boosting", *IFC Bulletin,* no 49.

Pinto, A, J Neves and T Pereira (2019): "Keeping track of MNEs through business group databases: the experience of Banco de Portugal", *IFC Bulletin*, no 49.

Richardson, A, T van Florenstein Mulder and T Vehbi (2019): "Nowcasting New Zealand GDP using machine learning algorithms", *IFC Bulletin*, no 50.

Sawaengsuksant, P (2019): "Standardised approach in developing economic indicators using internet searching applications", *IFC Bulletin*, no 50.

Tissot, B (2017): "Big data and central banking", *IFC Bulletin*, no 44.

———— (2019): "Financial big data and policy work: opportunities and challenges", *Eurostat Statistical Working Papers*, Eurostat, KS-TC-19-001-EN-N.

Wibisono, O, H Ari, A Widjanarti, A Zulen and B Tissot (2019): "Using big data analytics and artificial intelligence: a central banking perspective", in "Data Analytics", *Capco Institute Journal of Financial Transformation*, 50th edition, pp 70–83.

Annex 1: Survey on central banks' use of big data

1. How does your institution define big data (please select all that apply)?

☐ Data sets with a large number of observations in the cross-section (data points with multiple attributes, eg trade repository data)

☐ Data sets with a large number of observations in the time series (eg high-frequency)

☐ Structured traditional databases (eg credit registry data)

☐ Unstructured data sets that require new tools to clean and prepare (eg web scraping, textual analysis)

☐ Data sets that have not been part of your traditional data pool (eg commercial data on web searches) or data by commercial data providers

☐ Other (please specify below)

2. On a scale from 1 (not discussed) to 5 (extensively), how much does your institution formally discuss the topic of big data?

☐ 1 (not discussed, go to question 4)

☐ 2

☐ 3

☐ 4

☐ 5 (extensively)

3. What is the focus of the discussions on big data within your institution (please select all that apply)?

☐ Availability of data

☐ Storage of data

☐ Organisation and legal framework

☐ Cyber security

☐ Availability of staff with required IT skills

☐ Availability of tools/infrastructure to store/clean/process big data

☐ Developing a formal strategy for the use of big data (please specify below)

☐ Other (please specify below)

4. On a scale from 1 (very low interest) to 5 (very high level of interest), how do you rate the interest of your central bank in the topic of big data, as expressed at the senior policy level?

☐ 1 (very low interest)

☐ 2

☐ 3

☐ 4

☐ 5 (very high level of interest)

5. **Are you currently using any big data sources?**

☐ Yes

  ☐ For exploration (pilot projects)

  ☐ To inform policy decisions on a regular basis

☐ No (go to question 9)

6. **If you answered yes to question 5, please indicate:**

(i) the source of the big data that you use (eg Google internet search data / large administrative data sets)

(ii) the provider of these data (eg Google corporation / central bank data)

(iii) how you rate the usefulness of these data on a scale of 1 (not useful) to 5 (very useful)

(iv) what challenges you are facing with respect to these data sources (eg data access, analysis, usage, quality, privacy)

(i) Source	(ii) Provider	(iii) Usefulness	(iv) Challenges

7. **For what purposes does your institution use big data?**

A. General purpose (please select all that apply)

☐ Monetary policy

☐ Statistical compilation

☐ Economic research

☐ Financial stability

☐ Suptech (the use of innovative technology by supervisory authorities for the purpose of supervision)

  ☐ In microprudential supervision/oversight

  ☐ In macroprudential supervision/oversight

☐ Developing and evaluating the use of regtech (the use of innovative technology by financial institutions for compliance and to meet regulatory requirements)

☐ Other (please specify below)

B. Specific purpose (please select all that apply)

☐ Nowcasting

  ☐ Unemployment level/rates

  ☐ Industry/retail sales

  ☐ Retail/housing prices

- ☐ Private consumption
- ☐ GDP
- ☐ Payments
- ☐ Other (please specify below)

☐ Suptech

- ☐ Design early warning systems
- ☐ Ratings on banks and other financial institutions
- ☐ Fraud detection
- ☐ Anti-money laundering and countering the financing of terrorist activities
- ☐ Other (please specify below)

☐ Construction of web-based confidence/uncertainty indicators

☐ Measurement of the impact of information demand (eg web searches) on specific economic variables

☐ Platforms to facilitate data exchange

- ☐ Within your institution
- ☐ Outside your institution
- ☐ Other (please specify below)

8. Please describe (i) the main projects currently under way in your institution using big data, (ii) the source of big data used for and (iii) the purpose of the projects and (iv) the platform and application used.

(i) Project	(ii) Source	(iii) Purpose of the project	(iv) Platforms/applications

9. Is your institution planning to start any big data-related projects in 2020/21?

☐ Yes

☐ No, due to lack of resources (go to question 12)

☐ No, due to lack of interest (go to question 12)

☐ No, for another reason (go to question 12)

10. Please describe (i) the projects planned for 2020/21, (ii) sources of big data used for and (iii) purpose of these new projects, as well as the (iv) platform and application used.

(i) Project	(ii) Source	(iii) Purpose of the project	(iv) Platforms/applications

11. What projects would your organisation be interested in collaborating on with other central banks, supervision authorities or statistical agencies?

(i) Project	(ii) Description

12. Which business areas use payments data in your institution?

Business areas

☐ No business areas (go to question 14)

13. Which types of payments data are useful for your institution (please select all that apply)?

☐ All types

☐ High-frequency data

☐ High-frequency data, by instrument

☐ High-frequency data, by counterparty

☐ High-frequency data, by instrument and counterparty

☐ None (go to question 15)

14. In which area(s) do you think payments data are particularly useful? Please specify:

15. Would your institution be willing to contribute to a pilot study on the use of payments data (please select all that apply)?

- ☐ Yes
  - ☐ Nowcasting
  - ☐ Surveillance
  - ☐ Other (please specify below)
- ☐ No (please specify reasons below)

16. Please describe some of the potential resource implications (eg constraints, expenditures on sourcing and maintaining infrastructure) for your central bank in the regular production and/or analysis of big data.

IT (eg hardware and software)	
Human capital (eg computer skills)	
Legal issues (eg in collecting data)	
Other	

Annex 2: List of members that responded to the survey

1. Angola
2. Austria
3. Belgium
4. Brazil
5. Canada
6. CEMLA
7. Chile
8. Croatia
9. Cyprus
10. Czech Republic
11. Denmark
12. ECB
13. Estonia
14. Finland
15. France
16. Germany
17. Greece
18. Hungary
19. India
20. Indonesia
21. Israel
22. Italy
23. Japan
24. Latvia
25. Lebanon
26. Lithuania
27. Macao SAR
28. Malaysia
29. Malta
30. Mauritius
31. Mexico
32. Montenegro
33. Morocco
34. Netherlands
35. North Macedonia, Republic of
36. Norway
37. Peru
38. Philippines
39. Portugal
40. Romania
41. Russia
42. Saudi Arabia
43. Serbia
44. Slovakia
45. Slovenia
46. South Africa
47. Spain
48. Switzerland
49. Thailand
50. Turkey
51. United Kingdom
52. United States