

Irving Fisher Committee on Central Bank Statistics

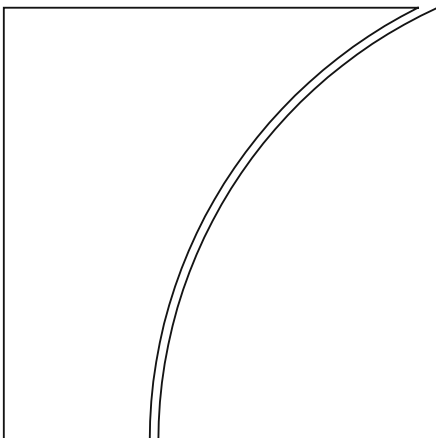
IFC Report

No 3

Central banks' use of and interest in "big data"

2015 Survey conducted by the Irving Fisher
Committee on Central Bank Statistics (IFC)

October 2015



BANK FOR INTERNATIONAL SETTLEMENTS

Contributors to the IFC report¹

BIS

Central Bank of the Republic of Turkey

European Central Bank

Bruno Tissot (IFC Secretariat)

Timur Hülagü

Per Nymand-Andersen

Laura Comino Suarez

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

© *Bank for International Settlements 2015. All rights reserved. Brief excerpts may be reproduced or translated provided the source is stated.*

ISBN 978-92-9197-283-8 (online)

ISBN 978-92-9197-282-1 (print)

¹ The views expressed in this document reflect those of the contributors and are not necessarily the views of the institutions they represent.

Contents

1. Executive summary	1
2. Background: the “big data” concept.....	3
3. Survey findings	6
Annex 1: IFC Questionnaire on central banks’ use of and interest in “big data”	19
Annex 2: Big data – references within the literature (replicated)	22
Annex 3: List of participating central banks and respondents	24

1. Executive summary

This report presents the results of the 2015 IFC survey of central banks on the use of and interest in “big data”.²

The aim of this survey was twofold:

- (i) To take stock of central banking experience in the use of big data; and
- (ii) To explore central banks’ interest in this topic with a view to defining a roadmap for further action.

The online survey took place in early 2015. The vast majority (69) of IFC member central banks responded, representing a response rate of 83%.

The main conclusions of the survey are the following:

Conclusion 1: *There is strong interest in big data in the central banking community, in particular at senior policy level.* Around two thirds of central banks are formally discussing or reviewing this topic internally.

Conclusion 2: *Central banks actual involvement in the use of big data is currently limited.* In contrast to their interest in the topic, only one third of the central banks surveyed are already using big data sources regularly or have started pilot initiatives. The related big data projects consist primarily of using conventional structured data sets (relying for instance on “official and administrative” sources and on micro data reported by the banking industry). However, there also seems to be a significant amount of interest in using private big data sources, such as Google search data, commercial data vendors’ data sets, mobile positioning data, news media etc.

Conclusion 3: *Big data can be useful for conducting central bank policies.* However, “conventional structured data sources” appear to be better known and more effectively mobilised than “new” big data sources. This might reflect the accumulated experience of central banks in working with large data sets from public sources or micro supervisory data, as well as the challenges posed by using new, private big data sources.

Conclusion 4: *Big data are perceived as a potentially effective tool in supporting macroeconomic and financial stability analyses.* Most central banks surveyed expect a growing use of big data sources for macroeconomic and financial stability purposes, especially in the area of: economic forecasting (for economic indicators such as inflation, housing prices, unemployment, GDP, industrial production, retail sales, external sector developments, tourism activity), business cycle analysis (eg sentiment indicators, nowcasting techniques), and financial stability analysis (eg construction of risk indicators, assessment of investors’ behaviour, identification of credit and market risk, monitoring of capital flows, supervisory tasks). Moreover, it was felt that big data could also be used to enhance the quality of existing, “more conventional” statistics.

² In the case of questions regarding the results, please contact the IFC Secretariat (bruno.tissot@bis.org), Timur Hülagü, CBRT (timur.hulagu@tcmb.gov.tr), Per Nymand-Andersen, ECB (per.nymand@ecb.int) and Laura Comino Suarez, ECB (Laura.Comino_Suarez@ecb.europa.eu).

Conclusion 5: Big data may also create new information/research needs.

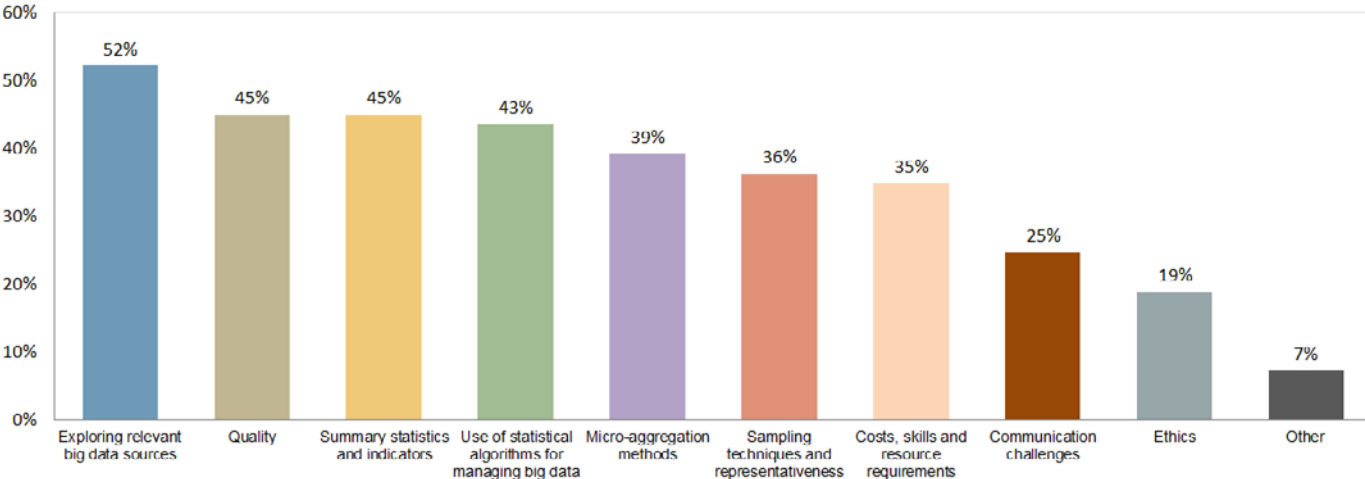
Interestingly, a significant proportion of the central banks surveyed expressed an interest not only in how big data information can help them to better analyse economic variables, but also in assessing the potential economy-wide influence of big data activities (eg web searches).

Conclusion 6: International cooperation can add value. More than 70% of the central banks would like to cooperate with their counterparts when investing in the area of big data. The way forward, as expressed by the vast majority of the respondents, is to set up a roadmap and define specific pilot projects. This step-by-step approach is perceived to be the most effective, particularly in clarifying the benefits of using big data sources, managing the associated challenges, and supporting the implementation of central bank policies.

Conclusion 7: Exploring big data is a complex, multifaceted task. Central banks have identified as many as nine priority areas to focus on (see Graph 1). These cover a wide range of topics, from purely statistical ones (eg sampling techniques) to economic analyses (eg summary indicators), administrative issues (eg resources) and public policy (eg communication).

Big data-related topics of central bank interest

Graph 1



Note: multiple responses possible.

Conclusion 8: Regular production of big data-based information will take time, especially because of resource issues.

A large majority of the respondents (ie close to 60%) does not consider that their central bank is ready to start regular statistical production using big data sources. A major concern relates to high costs in terms of investment in human capital and IT.

2. Background: the “big data” concept

In January 2015 the Irving Fisher Committee on Central Bank Statistics (“**IFC**”) launched an online survey on central banks’ use of and interest in big data, with the collaboration of the Bank for International Settlements (“**BIS**”), the Central Bank of the Republic of Turkey (“**CBRT**”) and the European Central Bank (“**ECB**”). The survey was answered by 69 IFC member central banks and monetary authorities worldwide, representing a response rate of 83%³ (see online questionnaire in Annex 1, and the list of participating central banks and contact details in Annex 3). This survey followed up the webinar organised for central banks in October 2014⁴ to present the outcome of the 2014 ECB workshop “*Using big data for forecasting and statistics*”, co-organised with the International Institute of Forecasters.⁵ Other international initiatives have also been launched in the recent past to better assess the handling of the big data revolution by the official community, including the central banks.⁶

The aim of the survey was to assess central banks’ experience and their interest in using big data. In particular, a key issue is whether and how new big data information related to financial and economic topics can help central banks to (i) better monitor the economic situation; (ii) enhance the effectiveness of their policy measures; and (iii) assess the impact of their actions within the financial system and the economy at large.

The scope of the survey was not to provide a unified definition of what “big data” actually is. In fact, this concept is not clearly defined and central banks appear to have a varying understanding and perception of what it entails. For instance, some central banks tend to consider granular “administrative data” (eg credit registers) or micro “financial information data” (eg security-by-security data sets) as an integral part of “big data”; others may have a more restricted concept, focusing mainly on web-based indicators.

Indeed, the literature provides little in the way of precise definitions for the term “big data”. What volume of data is needed before one can classify specific information as “big”? What are the specific characteristics, if any, of big-data data sets? How fast does the content of such data sets change over time? A further difficulty is that this concept is a fluid one. “Big data”, as defined 10 years, ago no longer seem “big” today, owing to the steady expansion of the daily volumes of data collected and of the variety of data, eg in terms of formats such as pictures or texts, frequency (daily data, tick data) and quality. One common view, however, is that big data can be broadly referred to as the four Vs: Volume, Variety, Velocity

³ Banco Central del Paraguay is not an IFC member but participated in the survey through the Centre for Latin American Monetary Studies (CEMLA), which is an IFC member.

⁴ See 2014 IFC Annual Report, www.bis.org/ifc/publ/ifc_ar2014.pdf.

⁵ www.ecb.europa.eu/events/conferences/html/20140407_workshop_on_using_big_data.en.html.

⁶ See the *Introductory discussion on big data* reported in the minutes of the 50th plenary meeting of the European Committee on Monetary, Financial and Balance of Payments statistics (CMFB) held on 2-3 July 2015 (available on www.cmfb.org/publications/minutes); cf also the workshop “*Big data: Building data strategies for central banks in light of the data revolution*” (www.riksbank.se/Documents/Forskning/Konferenser_seminarier/2015/program_workshop_big_data_150909.pdf) organised by the Sveriges Riksbank on 9 September 2015.

and Veracity (see Annex 2 for an overview of various definitions in the literature on “big data”).⁷ In any case, the big data concept encompasses a variety of large-scale, raw information which has to be combined and processed to make sense: from this perspective, one may wish to speak of “*smart data*” instead of “*big data*”.

As regards the actual types of big data, a good starting point is provided by the work conducted under the aegis of the United Nations Department of Economic and Social Affairs.⁸ Under this approach, big data can be classified into three types: (1) social networks (human-sourced information, such as blogs, videos, internet searches); (2) traditional business systems (process-mediated data, such as data produced in the context of commercial transactions, e-commerce, credit cards); and (3) internet of things (machine-generated data, such as data produced by pollution or traffic sensors, mobile phone locational information, and logs registered by computer systems). Feedback from central banks suggests that the third category should be split between “(3a) administrative systems” and “(3b) private business systems”: these two types of data sources have been used in different context over time and raise different types of issues.

At a less conceptual level, the survey highlighted some important characteristics of “big data” as perceived by the central bank community:⁹

- Big data are generally made of “granular” or “micro” bits of information that are usually produced by IT systems (eg web search engines, IT records of financial operations).
- Big data is very big – meaning large data sets, various different content, high frequency (ie the three Vs: high volume, high velocity, high variety).
- Big data can thus draw on multiple, potentially inconsistent sources (hence the importance of the fourth “V” of “veracity”).
- Big data are usually, but not always, a by-product: data are available because they result from business operations (eg credit card operations) or personal activities (eg web searches); in contrast, “standard” statistical data are usually compiled for specific purposes.
- As a corollary, big data are often available at low cost (with exceptions); what is really costly is to manage these data.
- Big data often apply to data sets that are “too large to be manipulated in a conventional way”; that is, big data comprise information that pose challenges to existing statistical systems.
- Big data are therefore generally felt to be difficult to manage in terms of IT resources and human skill mix.

⁷ P Nymand-Andersen, “Big data – the hunt for timely insights and decision certainty: Central banking reflections on the use of big data for policy purposes”, *IFC Working Paper*, 2015, forthcoming.

⁸ See Meeting of the Expert Group on International Statistical Classifications, “Classification of Types of Big Data”, United Nations Department of Economic and Social Affairs, ESA/STAT/AC.289/26, May 2015.

⁹ There are, of course, exceptions, eg mobile data, which can be costly to produce, and satellite images which are a product, not a by-product.

- Big data often need to be adequately stored, not least because of the size of the data sets and confidentiality issues (eg privacy protection, rights to access/modify personal information, anonymisation techniques).
- Big data may pose specific quality issues. In particular, there is a risk that big data can be misleading (what is their exact veracity?) or even be manipulated (raising ethical challenges).
- Big data sets usually need to be correctly filtered so as to extract appropriate intelligence (eg specific patterns) and inform decisions.

Looking ahead, the survey results suggest that the central bank community has an interest in launching specific pilot projects related to big data, which could cover the following specific topics within the four main categories defined above (ie social networks; administrative systems; private business systems; and internet of things):

- administrative data set (eg corporate balance sheet data);
- web search data set (eg web-based search indicators);
- commercial bank data set (eg credit card operations);
- financial market data (eg high frequency trading, bid-offer spreads).

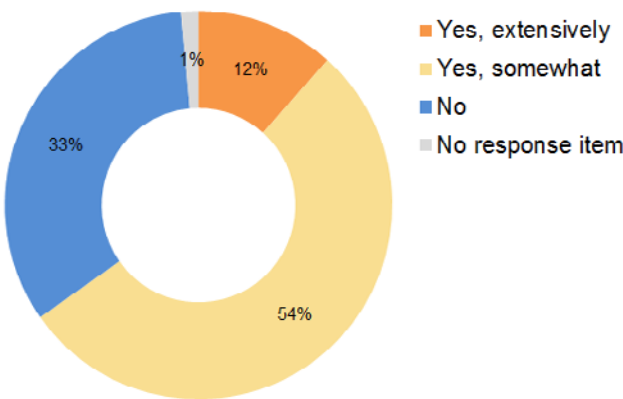
3. Survey findings¹⁰

3.1 The topic of big data is on the agenda of central banks

For two thirds of the respondents, the topic of big data has been extensively/somewhat discussed by their central bank, against one third of central banks for which this is not the case (Graph 2). This shows a noteworthy and growing awareness of this topic among the central bank community.

Is the topic of big data being formally discussed within your central bank?

Graph 2



Answers	Percentage	Count
Yes, extensively	12%	8
Yes, somewhat	54%	37
No	33%	23
No response	1%	1
Respondents		69

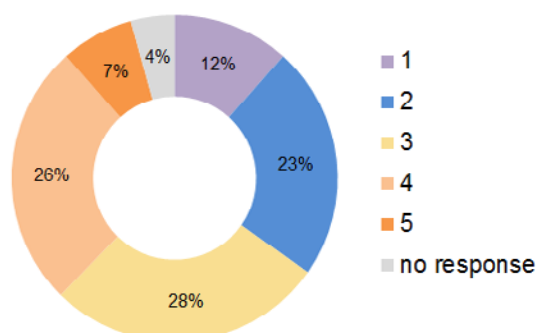
3.2 There is a significant interest in the topic of big data at senior policy level

Almost two thirds of the respondents ranked their central bank interest in the topic of big data as very high (5), high (4) or medium (3). About one third of the respondents ranked this interest as low (2) or very low (1) (Graph 3).

¹⁰ All the quantitative information presented in this document is derived from the IFC Big Data survey (2015).

How do you rate the interest of your central bank, as expressed at the senior policy level, in the topic of big data?

Graph 3



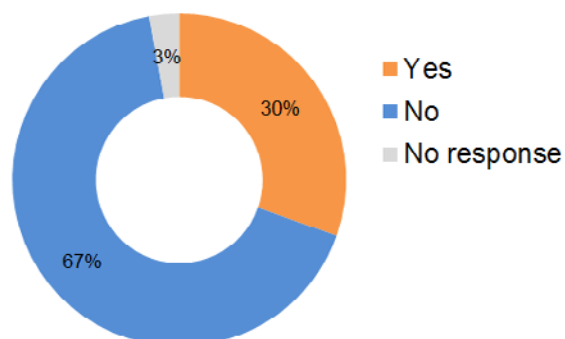
	Percentage	Count
1 (Very low)	12%	8
2	23%	16
3	28%	19
4	26%	18
5 (Very high)	7%	5
No response	4%	3
Respondents		69

3.3 Central banks are not yet using big data sources

Two thirds of central banks are not yet using big data sources, while less than one third are already doing so (Graph 4).

Are you already using big data sources?

Graph 4



	Percentage	Count
Yes	30%	21
No	67%	46
No response	3%	2
Respondents		69

3.4 The most useful big data sources are official sources and internal databases from central banks and national statistical offices

Respondents provided numerous examples for the potential use of big data (see Table 1).

Given the specific types of big data sources available, most answers emphasised first the usefulness of public and official sources of big data, which were rated as medium to very high. The advantages of these sources are numerous (quick

access, higher quality) although they pose specific challenges especially in terms of data access (ie restricted data-sharing practices¹¹) and confidentiality (ie the data collected for specific policy purposes may not be in the public domain at the individual level, for privacy/legal reasons). Another specificity is that these data sets are often numerical, ie they are not well designed for handling qualitative information (eg texts).

A second important source of big data for central banks is their internal (micro) databases or those of similar public authorities such as national statistical offices, foreign trade offices etc (when sharing arrangements exist). The main advantage is obviously central banks' familiarity with these data, which are usually perceived to be of good quality and easy to access and cross-check. But they also pose challenges in terms of classification, irregularities (eg outliers) and content extraction/visualisation.

A third reportedly useful source of big data is represented by the databases of financial institutions which can, at least partly, be accessed by central banks in the context of their public policy mandates.

Fourth, internet-based data are also deemed worthwhile, although to a lesser extent. In terms of advantages, they can be collected rapidly, made available for a very low cost, be updated frequently, and be analysed using standard technical methods. But several challenges were highlighted, such as timeliness (since the data have to be passed by the data providers to the users), operationality, cross-country consistency (eg different internet usage practices), coverage and representativeness (putting into question the apparent causalities displayed by these data and their predictive power), as well as communication to the public.

A fifth important category comprises big data provided by data vendors, which can be quite heterogeneous but are at least partly "formatted" for users.

Other data sources exist, such as university and supermarket records, but the related experience of central banks appears relatively limited. As regards mobile positioning data, one issue is the need to update the methodologies required to use this source of information. Similarly, media and social network sources are difficult to interpret and compared to economic or financial developments (eg how to determine whether messages are negative or positive?).

¹¹ See on this topic: IFC, *Data-sharing: issues and good practices*, Report to the BIS Governors prepared by IFC Task Force on Data Sharing, January 2015.

Examples of the use of big data sources

Table 1

Big data sources – main types	Examples
Official sources	Foreign trade/investment transactions Taxation files Central balance sheet offices Credit bureaus/registers Housing/mortgage registers Financial market supervisors Public financial statements Financial market activity indicators
Internal public databases	Central bank monetary & financial surveys National statistical offices Banking supervisory data Public balance sheet data
Internet-based data	Baidu Google Job portals Multiple other websites Web-scraping information providers ¹²
Databases of financial institutions	Resident credit institutions (loans, mortgages) Non-bank entities Derivatives trade repositories Investment fund holdings, custodian activities
Data vendors	Microsoft Analysis Services Commercial providers Settlement operations (including FX) Securities statistics Credit card operations
Mobile positioning data	Mobile operators Location-based service providers
Media and social networks	Ebsco, Eureka, Coosto, Facebook, You Tube
Supermarket records	Tesco (online price data set) POS (point of sale) data for price index

3.5 Conventional data versus new data sources

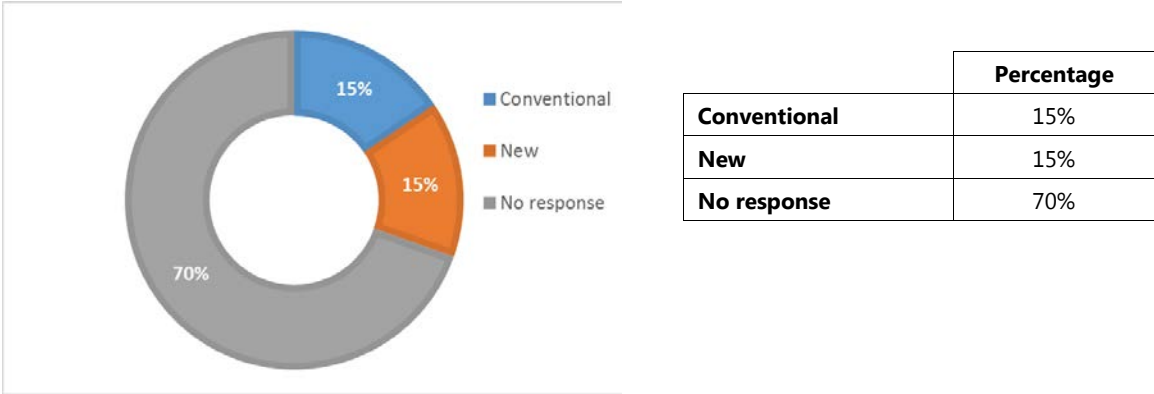
Conventional data sources for central banks usually include administrative records and reported data from the banking sector covering securities, derivatives, credit etc. New data sources refer more to web-, IT- or telecom-based indicators, such as internet job search or price data made available by various internet operators such as Google, news media firms and mobile operators.

¹² Computer software techniques for extracting information from websites.

As already stated, around 70% of central banks are not using any big data source. For those doing so, the survey showed that the big data sources are equally distributed between conventional and new types of source (Graph 5).

Conventional data sources versus new ones

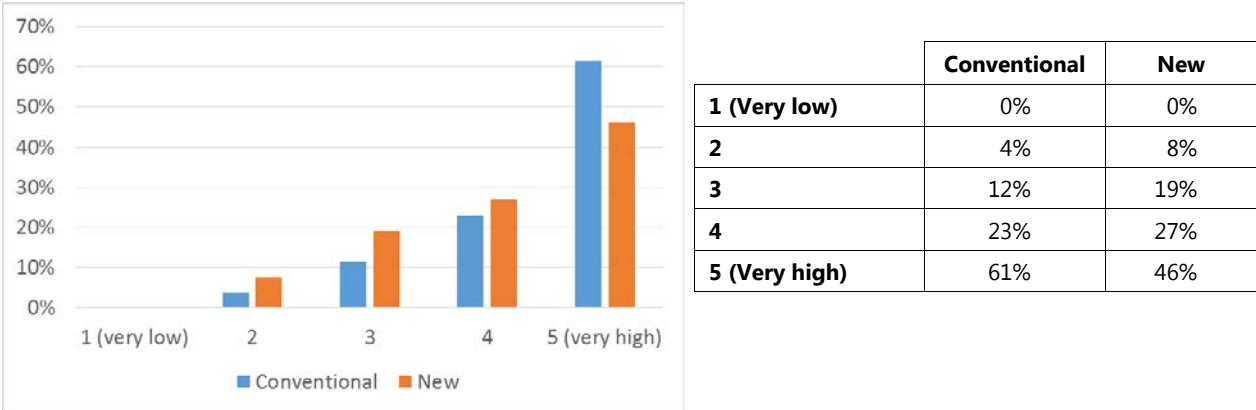
Graph 5



Both conventional and new data are generally reported as useful (Graph 6), with a slight preference expressed for conventional data. This result may be explained by the nature of central banks’ activities and the fact that they have been working with “conventional” data sources for a longer period of time. Many institutions appear to have started only recently to consider how new big data sets might be used for policy purposes.

Usefulness of big data by source type¹³

Graph 6



¹³ One respondent reply could not be classified and has been omitted.

3.6 Challenges posed by using big data

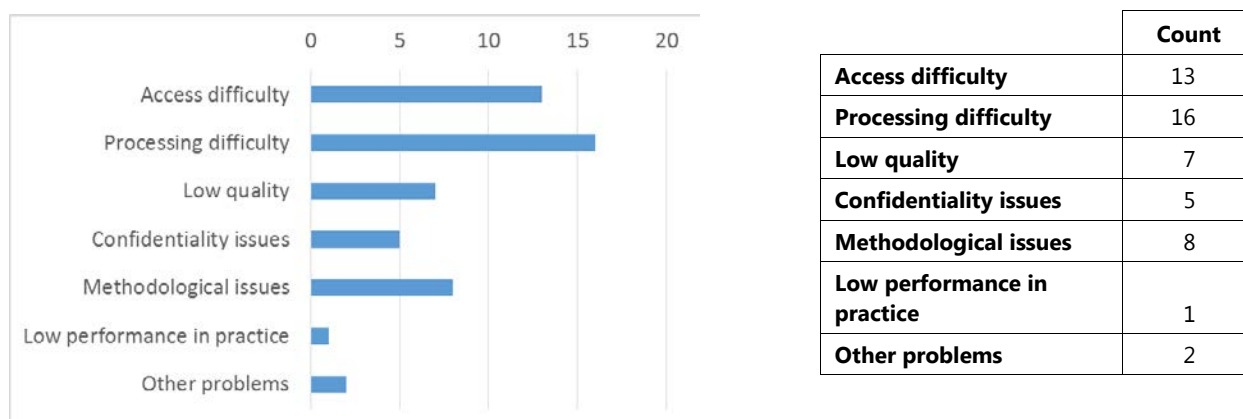
The main challenge to central banks in using big data relates to accessing and processing the data (Graph 7). Many countries reported that accessing (in a timely and rapid way), storing or analysing the data is a real difficulty. This also explains the limited degree of preparedness at central banks for big data processing, owing to limitations in human and IT resources (hardware and software).

Another major concern relates to data quality, for instance cross-country comparisons as well as consistency over time. Given the lack of general methodological guidance and/or best practice for using big data, the risk lies in following incorrect processes and in undermining the quality of central banks' work. This could potentially create reputational issues that would need to be carefully assessed by public authorities. From this perspective, the focus of attention, as inferred from the survey responses, seems to be on existing limitations regarding quality insurance and validation procedures (due to the high volume, frequency and growth of the data sets).

Lastly, concerns about confidentiality issues are also important, in that they limit the scope for accessing, sharing and disseminating data beyond a certain level of granularity.

Challenges of using big data

Graph 7



3.7 Central banks expect a growing use of big data for macroeconomic and financial stability purposes

Most surveyed central banks expect a growing use of big data sources for macroeconomic and financial stability purposes, especially (see Graph 8) in the area of:

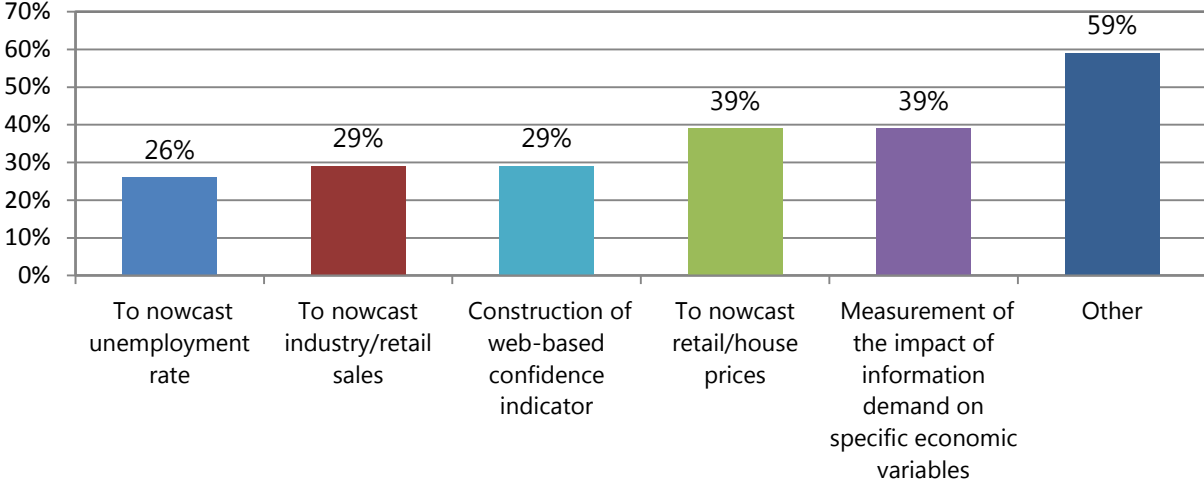
- economic forecasting: for economic indicators such as inflation, prices (for instance 39% of central banks expect to nowcast retail/house prices using big data), unemployment, GDP, industrial production, retail sales, tourism activity;
- business cycle analysis (eg sentiment indicators, nowcasting techniques); and
- financial stability analysis (eg construction of risk indicators).

Interestingly, 39% of the respondents mentioned an interest in measuring the potential influence of big data activities (eg web searches). Indeed, there is already significant literature on the economy-wide impact of internet operations, for example to capture how they can support productivity, price discovery and network effects (and thereby contributing to sizeable gains in terms of GDP).¹⁴

What kind of outcomes are you expecting as a result of exploring big data sets?

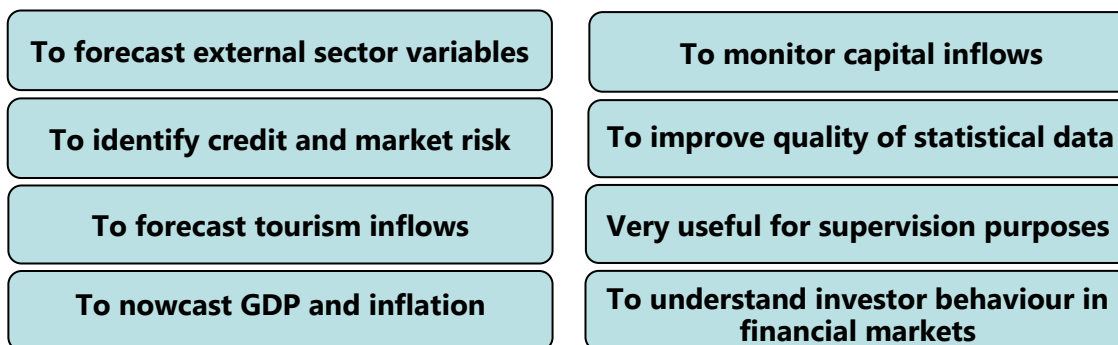
Graph 8

Note: multiple responses possible.



Lastly, the survey suggests that the range of potential applications for big data is quite wide: almost two thirds of the central banks indicated that they have other expectations in addition to the possibilities listed in the questionnaire (see Graph 9). This was particularly the case for financial stability purposes: big data are thought to be potentially useful for the identification of credit and market risk, the monitoring of capital flows, various supervisory tasks, and the understanding of financial market behaviour.

¹⁴ See J Bughin, L Corb, J Manyika, O Nottebohm, M Chui, B de Muller Barbat and R Said, *The impact of Internet technologies: Search*, McKinsey Global Institute, July 2011. Their estimates suggest that search contributed to between 1.2 and 0.5 percent of GDP in five countries studied.

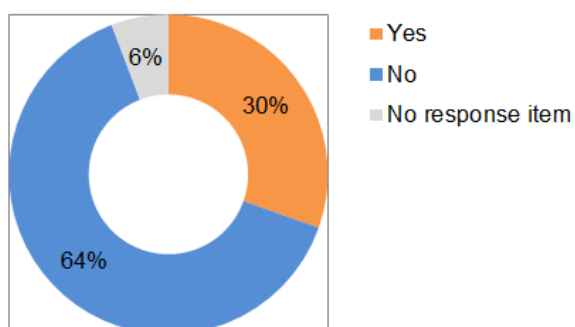


3.8 Only a limited number of central banks plan to start big data projects

Several central banks do not expect to start big data-related initiatives soon, although almost one third are exploring some pilot projects during 2015–16 (Graph 10).

Is your central bank planning to start any big data-related pilot projects in 2015–16?

Graph 10



	Percentage	Count
Yes	30%	21
No	64%	44
No response	6%	4
Respondents		69

3.9 Examples of pilot projects envisaged by central banks

Central banks are collecting data from various sources such as administrative records, search engines, reporting banks’ and financial institutions’ databases, online news, supermarkets – in order eg to nowcast unemployment, conduct research on price dynamics, improve the understanding of credit risk, market risk and financial operations, and analyse micro data to support central banks’ policy (see Table 2).

Examples of big data pilot projects envisaged in 2015–16

Table 2

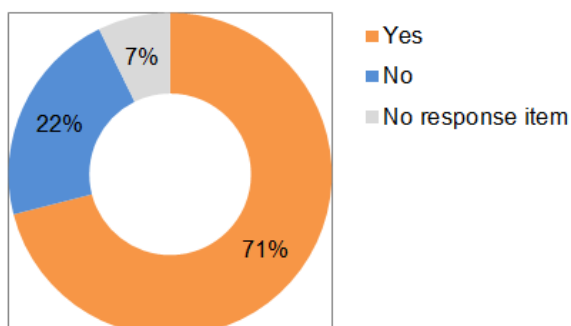
Big data source	Examples of projects
Internal database at central banks and national statistical offices	Microeconomic behaviour modelling of small and medium-sized enterprises (SMEs) Improved statistical quality controls Network analysis of the financial system Centralisation of all operations carried out by the central bank in a single repository
Official sources	Implementation of big data software Centralised institutional data warehouse to improve data access, visualisation, analysis, storage Analysis of various micro statistics (loans, derivatives, local government balance sheets) Transaction data for network analysis Tax information
Internet-based data	Construction of unemployment/employment indicators Build-up of economic sentiment & house price indices Improvement in the quality of web-based data Research on producer price dynamics Measurement of house prices Text data (eg signal analysis)
Database of financial institutions	Collection of granular credit data, measurement of credit risk Loan-by-loan data set Securities holdings statistics Mutual funds data Analyses of investor behaviour/expectations Analysis of financial markets' liquidity and patterns (FX, bond markets, equities)
Mobile data (eg international calls)	Enhancement of balance of payments statistics (tourism)
Media and social networks	Nowcasting of industrial production Analysis of central bank policy perceptions
Supermarket records	Early indicators of inflation
Surveys	Economic agent confidence indicators

3.10 Central banks are willing to cooperate with other IFC members and engage in the topic of big data

Almost three fourths of central banks would like to collaborate with other IFC members in the area of big data. Only 22% are not willing to be part of such an initiative (Graph 11).

Would you be willing to cooperate with other IFC members and engage your central bank in the area of big data?

Graph 11



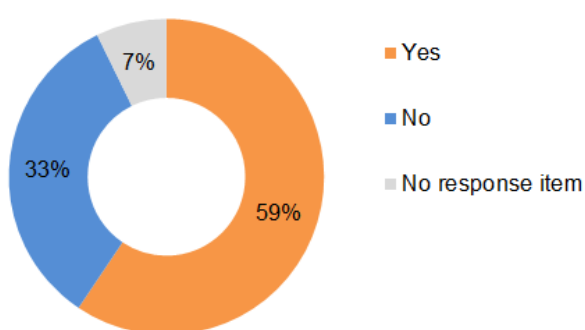
	Percentage	Count
Yes	71%	49
No	22%	15
No response	7%	5
Respondents		69

3.11 Central banks are willing to contribute to a roadmap on the potential use of big data

The majority of central banks would like to contribute to the establishment of a “roadmap” for the potential use of big data; only 33% answered negatively (Graph 12). The aim of such a roadmap could be to provide guidance for using big data for monetary and financial stability purposes, based on the experience accumulated in various countries, and examples of “best practice”.

Would your institution be willing to contribute to a roadmap for central banks on the potential use of big data?

Graph 12



	Percentage	Count
Yes	59%	41
No	33%	23
No response	7%	5
Respondents		69

3.12 Work should primarily focus on exploring the big data sources, their quality issues, the build-up of summary indicators, and the use of algorithms for managing big data

The survey showed that there is significant interest among central banks for a wide range of statistical topics as regards the use of big data.

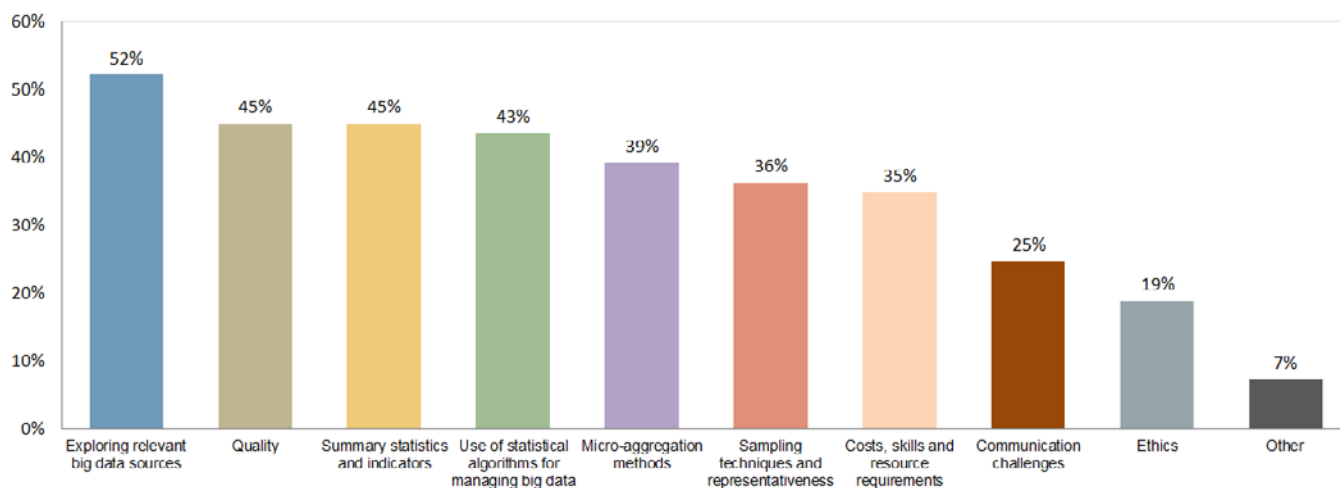
The four areas on which attention should be focused are reported to be the exploring of relevant big data sources, the analysis of big data quality, the production of summary statistics and indicators using big data sources, and the use of statistical algorithms for managing big data (Graph 13).

Significant interest was also expressed in proposals to work on micro-aggregation methods; sampling techniques and representativeness; big data-related costs and resource requirements; communication challenges; and, to a lesser extent, ethics-related issues.

In addition, survey participants underlined central banks' interest in confidentiality and data-sharing issues (eg integration of different sources of data, legal challenges); methodologies for exploiting and making sense of big data; the design of an analytical framework for big data analysis; and the need to consider public trust in big data information.

Which statistical topics would be of interest to you as part of the big data subject?

Graph 13



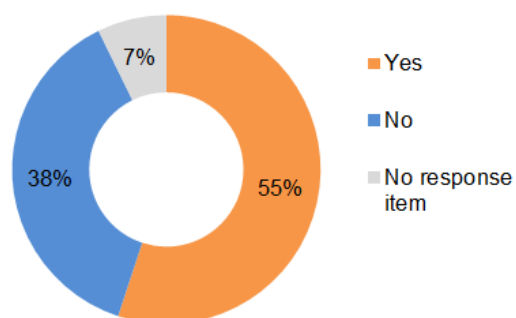
Note: multiple responses possible.

3.13 Central banks are willing to contribute to a pilot study on the use of selected big data sources

Some 38 central banks out of 69 would like to participate in a pilot study for the use of selected big data sources (Graph 14).

Would your central bank be willing to contribute to a pilot study on the use of selected big data sources?

Graph 14



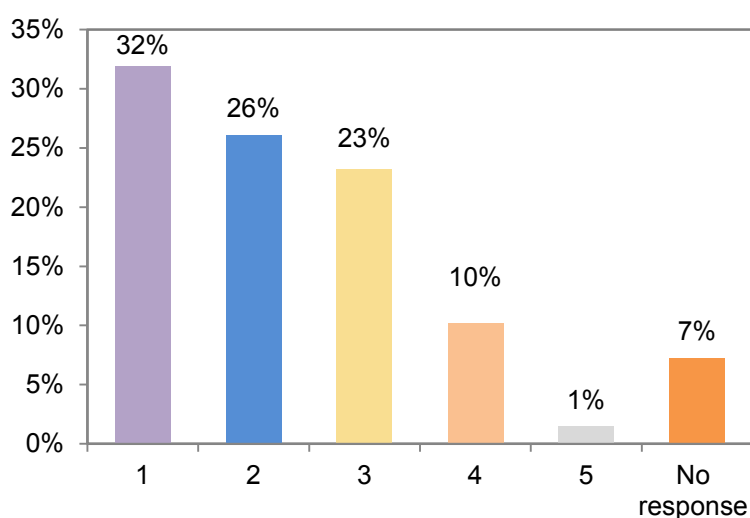
	Percentage	Count
Yes	55%	38
No	38%	26
No response	7%	5
Respondents		69

3.14 Most central banks are not ready to start the regular production and/or analysis of big data

Some 58% of the central banks did not consider themselves ready to start a regular production and analysis of big data (although 11% did, however, think otherwise: see Graph 15).

How would you rate the readiness of your central bank to start regular production and/or analysis of big data (1: low readiness/not ready to 5: high readiness)

Graph 15



	Percentage	Count
1	32%	22
2	26%	18
3	23%	16
4	10%	7
5	1%	1
No response	7%	5
Respondents		69

3.15 Big data projects would require significant IT and human resources

For many central banks, big data projects would require significant investment in IT infrastructure. The challenge is that resources are reported to be limited due to budget constraints and human capital limitations (eg the need to train staff and develop specialised expertise to be able to handle and analyse large data sources). Moreover, it was felt that the IT implications of big data are complex, as specific software and hardware are required to ensure the adequate collection, storage, analysis and reporting of such information.

Annex 1: IFC Questionnaire on central banks' use of and interest in "big data"

1. Is the topic of big data being formally discussed within your central bank?

- Yes, extensively
- Yes, somewhat
- No

2. How do you rate the interest of your central bank, as expressed at the senior policy level, for the topic of big data?

- 1 (very low interest)
- 2
- 3
- 4
- 5 (very high level of interest)

3. Are you already using any big data sources?

- Yes
- No (you will be directed to question 5)

4. Please indicate:

(i) the source of the big data that you use (e.g. Google internet search data),

(ii) the provider of these data (e.g. Google corporation)

(iii) how you rate the usefulness of these data on a scale of 1 to 5, where 1 means very low and 5 means very high, and

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

	(i) Source	(ii) Provider	(iii) Usefulness	(iv) Challenges
Source 1				
Source 2				
Source 3				
Source 4				
Source 5				

5. What kind of outcomes – if any – are you expecting as a result of exploring big data sets?

Please select all that apply

- a) Ability to nowcast unemployment level/rates
- b) Ability to nowcast industry/retail sales
- c) Ability to nowcast changes in retail/housing prices

- d) Construction of web-based confidence indicator
- e) Measurement of the impact of information demand (e.g. web searches) on specific economic variables
- f) Other (please specify below)

6. Is your central bank planning to start any big data-related pilot projects in 2015/16?

- Yes
- No (you will be directed to question 8)

7. Please describe (i) the source(s) for and (ii) the purpose of the project(s).

Example:

- (i) European Postal Federation; (ii) Early indicator of trade flows

Project	Source	Purpose of the project

8. Would you be willing to cooperate with other IFC members and engage your central bank in the topic of big data?

- Yes
- No

9. Would your institution be willing to contribute to a roadmap for central banks on the potential use of big data?

- Yes
- No (you will be directed to question 11)

10. Which statistical topics would be of interest to you as part of the big data subject?

Please select all that apply

- Exploring relevant big data sources
- Sampling techniques and representativeness
- Quality
- Summary statistics and indicators
- Use of statistical algorithms for managing big data
- Micro-aggregation methods
- Communication challenges
- Costs, skills and resource requirements
- Ethics
- Other (please specify below)

11. Would your central bank be willing to contribute to a pilot study on the use of selected big data sources?

- Yes
- No

12. How would you rate the readiness of your central bank to start regular production and/or analysis of big data?

- 1 (low readiness/not ready)
- 2
- 3
- 4
- 5 (high readiness)

13. Please describe some of the potential resource implications for your central bank in starting the regular production and/or analysis of big data, especially in terms of IT resources (both software and hardware) and human capital (e.g. related to IT, statistics and economic research).

Annex 2: Big data – references within the literature¹⁵

Year	Reference	Definition
2015	Oxford Dictionaries ¹⁶	<i>“Extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions: ‘much IT investment is going towards managing and maintaining big data.’”</i>
2014	Gartner ¹⁷	<i>“Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.”</i> This definition of big data is referred to as a 3V model .
2014	IBM ^{18, 19, 20}	IBM has extended the Garner’s definition by adding a fourth “V” to the 3V model called “Veracity”, as part of expressing the uncertainty and quality of data sources. <i>“Every day, we create 2.5 quintillion bytes of data — so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is big data.”</i>
2014	SAS ²¹	<i>“For most organizations, big data is the reality of doing business. It’s the proliferation of structured and unstructured data that floods your organization on a daily basis – and if managed well, it can deliver powerful insights.”</i>
2013	Fedtech Magazine ²²	<i>“New trends in IT are often thought of in terms of leading-edge technology solutions to significant enterprise challenges. In other words, organizations face challenges (for example, improving data center efficiency or providing IT services to remote workers) and IT solutions (such as cloud computing and mobile devices) address them. But one of the biggest trends in IT today, Big Data, is actually named for the challenge it represents, rather than the solution.</i> <i>At its core, Big Data means lots of data — so much data collected via so many evolving mechanisms that it can be overwhelming. It is increasingly easy for government agencies to store all that data. But Big Data as an IT strategy requires making sense of the data being collected — processing, analyzing and exploiting it for government, partner and constituent gain.</i>

¹⁵ From P Nymand-Andersen, “Big data – the hunt for timely insights and decision certainty: Central banking reflections on the use of big data for policy purposes”, *IFC Working Paper*, 2015, forthcoming.

¹⁶ Oxford Dictionaries: www.oxforddictionaries.com/definition/english/big-data.

¹⁷ Gartner: www.gartner.com/it-glossary/big-data/.

¹⁸ IBM: www-01.ibm.com/software/data/bigdata/what-is-big-data.html.

¹⁹ IBM: Big data, bigger outcomes (PDF).

²⁰ www-01.ibm.com/software/data/bigdata/what-is-big-data.html and the infographic can be found using the following link www.ibmbigdatahub.com/sites/default/files/infographic_file/4-Vs-of-big-data.jpg.

²¹ SAS: www.sas.com/en_us/insights/big-data.html.

²² Fedtech: www.fedtechmagazine.com/sites/default/files/122210-wp-big-data-df.pdf.

		<i>Big Data refers to digital information that is massive and varied, and that arrives in such waves that it requires advanced technology and best practices to sort, process, store and analyze. Organizations that do so effectively can use it to their advantage. Big Data is less about the terabytes than it is about the query tools and business intelligence software needed to make sense of the terabytes."</i>
2012	Microsoft ²³	<i>"The increasingly large and complex data that is now challenging traditional database systems YouTube videos, Facebook posts, credit card transactions, store inventory, your last grocery purchase. Trillions of pieces of information are being collected, stored, and analyzed almost daily with increasing speed. Big Data addresses one of the most critical issues facing business today: how to gain value from the growing reams of complex data."</i>
2012	Global Pulse ²⁴	<i>"'Big Data' is a popular phrase used to describe a massive volume of both structured and unstructured data that is so large that it's difficult to process with traditional database and software techniques. The characteristics which broadly distinguish Big Data are sometimes called the '3 V's': more volume, more variety and higher rates of velocity. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos posted online, transaction records of online purchases, and from cell phone GPS signals to name a few. This data is known as 'Big Data' because, as the term suggests, it is huge in both scope and power."</i>
2011	McKinsey ²⁵	<i>"'Big data' refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze. This definition is intentionally subjective and incorporates a moving definition of how big a dataset needs to be in order to be considered big data—i.e., we don't define big data in terms of being larger than a certain number of terabytes (thousands of gigabytes). We assume that, as technology advances over time, the size of datasets that qualify as big data will also increase. Also note that the definition can vary by sector, depending on what kinds of software tools are commonly available and what sizes of datasets are common in a particular industry. With those caveats, big data in many sectors today will range from a few dozen terabytes to multiple petabytes (thousands of terabytes)."</i>
2001	META Group ²⁶ (now Gartner)	<i>Gartner's paper does not use the term "big data" but rather reports on data management challenges: "While enterprises struggle to consolidate systems and collapse redundant databases to enable greater operational, analytical, and collaborative consistencies, changing economic conditions have made this job more difficult. E-commerce, in particular, has exploded data management challenges along three dimensions: volumes, velocity, and variety." This "3V model" has almost become the standard way of defining big data.</i>

²³ Microsoft: http://download.microsoft.com/download/.../Microsoft_Big_Data_Booklet.pdf.

²⁴ Global Pulse: Big Data for Development: Challenges & Opportunities.

²⁵ McKinsey: Big data: The next frontier for innovation, competition, and productivity (see full report as pdf).

²⁶ D Laney: "3D Data Management: Controlling Data Volume, Velocity and Variety", Gartner, 2001.

Annex 3: List of participating central banks and respondents

Central bank	Contact details
Bangko Sentral Ng Pilipinas	mrguerrero@bsp.gov.ph
Bank Indonesia	hendy_s@bi.go.id
Bank of Algeria	achari@bank-of-algeria.dz
Bank of Canada	mchitu@bank-banque-canada.ca
Bank of England	stephen.sabine@bankofengland.co.uk
Bank of Estonia	jaanus.kroon@eestipank.ee
Bank of Finland	laura.vajanne@bof.fi
Bank of France	jacques.fournier@banque-france.fr
Bank of Greece	ntsaveas@bankofgreece.gr
Bank of Israel	tzahif@boi.gov.il
Bank of Italy	riccardo.debonis@bancaditalia.it
Bank of Japan	masahiro.higo@boj.or.jp
Bank of Korea	shwanpark@bok.or.kr
Bank of Latvia	agris.caune@bank.lv
Bank of Lithuania	rliksa@lb.lt
Bank of Mauritius	vikram.punchoo@bom.mu
Bank of Mexico	podogher@banxico.org.mx
Bank of Morocco	m.taamouti@bkam.ma
Bank of Mozambique	felisberto.navalha@bancomoc.mz
Bank of Portugal	joao.cadete.matos@bportugal.pt
Bank of Slovenia	janez.fabijan@bsi.si
Bank of Spain	penalosa@bde.es
Bank of Thailand	pusadeeg@bot.or.th
Bank of the Republic – Colombia	cvarelba@banrep.gov.co
Board of Governors of the Federal Reserve System	charles.thomas@frb.gov
Central Bank of Armenia	martin.galstyan@cba.am
Central Bank of Bosnia and Herzegovina	ahadziomeragic@cbbh.ba
Central Bank of Chile	rvicuna@bcentral.cl
Central Bank of Cyprus	angeloskapatias@centralbank.gov.cy
Central Bank of Iceland	tomas.orn.kristinsson@sedlabanki.is
Central Bank of Ireland	joe.mcneill@centralbank.ie
Central Bank of Luxembourg	roland.nockels@bcl.lu
Central Bank of Malaysia	hctoh@bnm.gov.my
Central Bank of Malta	pulej@centralbankmalta.org
Central Bank of Nigeria	sidoguwa@cbn.gov.ng

Central Bank of Norway	vetle.hvidsten@norges-bank.no
Central Bank of Paraguay	ljara@bcp.gov.py
Central Bank of Suriname	sjahangir@cbvs.sr
Central Bank of the Islamic Republic of Iran	a.moghtadaei@cbi.ir
Central Bank of the Republic of Austria	johannes.turner@oenb.at
Central Bank of the Republic of Azerbaijan	samir_nasirov@cbar.az
Central Bank of the Republic of Turkey	gulbin.sahinbeyoglu@tcmb.gov.tr
Central Bank of the Russian Federation	pev@cbr.ru
Central Bank of Trinidad and Tobago	ahenry@central-bank.org.tt, dstoddard@central-bank.org.tt
Central Bank of Venezuela	rpuerta@bcv.org.ve
Central Reserve Bank of Peru	teresa.lamas@bcrp.gob.pe
Czech National Bank	vlastimil.vojacek@cnb.cz
Deutsche Bundesbank	robert.kirchner@bundesbank.de
European Central Bank	aurel.schubert@ecb.int
Federal Reserve Bank of New York	anthony.cirillo@ny.frb.org
Hong Kong Monetary Authority	rsyyuen@hkma.gov.hk
Magyar Nemzeti Bank	gerendasj@mnb.hu
Monetary Authority of Singapore	eileen_cm_tan@mas.gov.sg
National Bank of Belgium	rudi.acx@nbb.be
National Bank of Poland	eugeniusz.gatnar@nbp.pl
National Bank of Romania	marian.mustareata@bnro.ro
National Bank of Serbia	jelena.maravic@nbs.rs
National Bank of Slovakia	gregor.bajtay@nbs.sk
National Bank of the Republic of Belarus	a.galov@nbrb.by
National Bank of the Republic of Macedonia	angelovskaba@nbrm.mk
National Bank of Ukraine	polovnev@bank.gov.ua
Netherlands Bank	s.butzke@dnb.nl
People's Bank of China	rjianhong@pbc.gov.cn
Reserve Bank of India	opmall@rbi.org.in
Reserve Bank of New Zealand	rochelle.barrow@rbnz.govt.nz
South African Reserve Bank	johan.vandenheever@resbank.co.za
Sveriges Riksbank	gunnar.blomberg@riksbank.se
Swiss National Bank	guido.boller@snb.ch