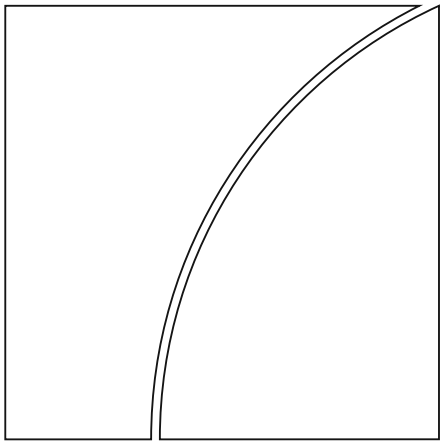
Irving Fisher Committee on Central Bank Statistics

IFC Report

No 9



Business intelligence systems and central bank statistics

October 2019

BANK FOR INTERNATIONAL SETTLEMENTS

Contributors to the IFC Report[1]

Central Bank of the Russian Federation (CBRF)	Dmitry Protsenko
	Maria Vilkova
Bank for International Settlements (BIS)	Edward Lambe
	Bruno Tissot (IFC Secretariat)

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

ISSN 1991-7511 (online)

ISBN 978-92-9259-295-0 (online)

[1] Respectively Economic Adviser, Central Bank of the Russian Federation (CBRF) (protsenkods@cbr.ru); Head of Aggregated Data Sector, CBRF (vilkovamYu@cbr.ru); Senior Business Analyst, Bank for International Settlements (BIS) (Edward.Lambe@bis.org); and Head of Statistics and Research Support, BIS, and Head of the Irving Fisher Committee on Central Bank Statistics (IFC) Secretariat (Bruno.Tissot@bis.org).

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

Contents

Contributors to the IFC Report 2

Business intelligence systems and central bank statistics 4

1. Executive summary 4

2. Background 5

3. The need for BI in statistical production systems 6

4. Types of BI systems used in central banks 7

5. Technology stack involved in statistical computing 9

   BI tools 9

   Statistical package / programming language 10

6. Data visualisation and publications 11

   Medium for publication/dissemination 12

   Static versus dynamic graphical content 13

   Software for generating graphical content 14

References 17

Appendix: List of countries that responded to the survey 19

Business intelligence systems and central bank statistics[2]

1. Executive summary

Central banks increasingly need to use business intelligence (BI) systems to collect, manage and analyse data in order to inform policy decisions. This report presents the results of a survey conducted by the BIS's Irving Fisher Committee on Central Bank Statistics (IFC) in 2019. Its main findings can be summarised as follows:

- **BI has become an integral element supporting the statistical function** in the vast majority of central banks, especially to analyse, present and integrate statistical data. Other use cases include data elaboration, big data analytics and statistical quality processes.
- Reflecting their various needs, the **range of BI tools central banks use is quite large**. The main categories are data visualisation instruments, design report instruments, developed analytic tools and online analytical processing (OLAP) technology, and analytic and strategic dashboards. In practice, central banks are using multiple BI tools that complement each other. The degree of satisfaction with the use of BI tools is quite high.
- The **related technology stack** comprises packaged solutions available "off the shelf", with a strong preference expressed for the business analytics service provided by Microsoft and SAS. In addition, central banks also rely on the extensive use of statistical packages/programming languages for conducting specific BI work: the most used are MATLAB, Stata and R – with a high demand for Python in data science and big data-related projects.
- **Visualisation** is an important issue when making sense of data, and is particularly key for central banks' communication. From this perspective, a large majority of central banks have developed an internet-based dissemination strategy; only one quarter of central banks rely exclusively on printed documents for dissemination/publication purposes.
- **Internet-based data dissemination** offers users the possibility to interact with central banks' publications in a dynamic/interactive way. As a result, two thirds of central banks use a mix of both static and dynamic graphical content when presenting information on their websites.
- JavaScript Graphing Libraries are the most popular means for **generating graphical content for web publication**. Other tools, such as the spreadsheet and business analytics services provided by Microsoft, are also frequently used to create graphical content. In addition, a significant number of central banks have developed internal, "customised" solutions for their graph production.

[2] This report benefited from comments by Giulio Cornelli and Barend de Beer. Giulio Cornelli's assistance with the preparation of the graphs is also gratefully acknowledged.

2. Background

In early 2019, the Central Bank of the Russian Federation (CBRF) and the BIS joined forces to organise a survey on the use of BI systems by central banks around the world, with a focus on their statistical function.[3] The objective was to facilitate the **exchange of experience in using BI systems**, with particular focus on the technology and instruments commonly involved in central banks' statistical production tasks (eg dissemination of economic statistics, publications) and contributions to decision-making processes (eg data analysis and visualisation exercises). The survey was answered by 35 IFC members (see Appendix).

The **concept of BI** is usually understood as encompassing the technology-driven methods and techniques (eg IT applications and practices) mobilised to collect, manage and analyse data in order to inform business decisions. BI systems are therefore used for multiple tasks in central banks, including collecting, storing, integrating, searching and querying, reporting, analysing, presenting, publishing and visualising data.[4] The ultimate objective is to make sense of the information that is relevant for the business considered and to facilitate evidence-based decision-making – thereby transforming data into information and then information into knowledge (Drozdova (2017)).

BI is not a new concept, and firms as well as public authorities like central banks have traditionally focused on getting the right "actionable" information to conduct their operations and base their decisions on. Yet **three major developments** have reinforced the need for BI systems in recent years:

- The surge in data available through digital innovation (the "big data revolution"; IFC (2017a)), combined with the greater demand for more granular sources of information on the financial system[5]
- The expansion of IT tools and computing power to manage these data (IFC (2019)
- The growing recognition of data's role in supporting (better) decisions[6]

The survey covered **four main areas**:

(i) The need for BI in statistical production systems

(ii) The types of BI systems currently used by central banks

[3] For an introduction to BI's role in the financial landscape generally, and in central banks' statistical units more specifically, see Drozdova (2017).

[4] For an example of how a central bank is approaching all these different data tasks in an integrated way, see Prokunina (2019).

[5] With the increasing emphasis on the importance of micro data in the context of the Data Gaps Initiative (DGI) endorsed by the G20 after the Great Financial Crisis (GFC) of 2007–09, especially its second phase (FSB-IMF (2009, 2015)). For a discussion of the (micro) data implications, see Tissot (2018).

[6] This has been particularly relevant for public authorities since the GFC, with greater attention devoted to developing central banks' intelligence activities by taking advantage of new methods and information sources (Nishimura (2012)) and to conducting evidence-based policies (Tissot (2017)).

(iii) The technology stack involved in this context, with regard to commercially available BI tools and the statistical packages / programming languages used by central banks

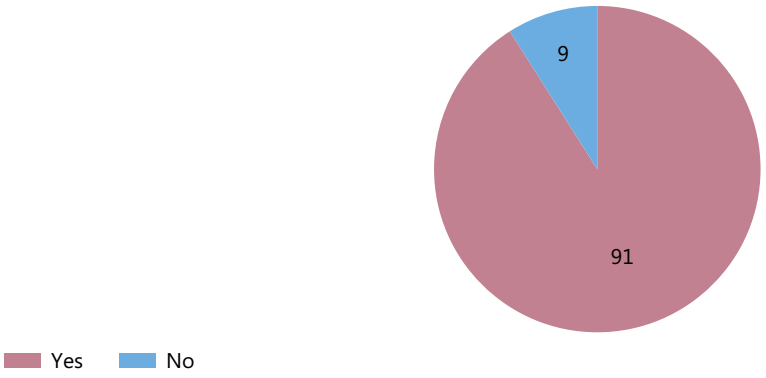
(iv) Issues related to data visualisation exercises and publications

3. The need for BI in statistical production systems

BI has become **a key element supporting a central banks' statistical function**. The vast majority of the survey respondents (91%; Graph 1) indicated that their organisations use BI, the aim being to give users easy ("intuitive") and direct ("self-service") interactive access to data and the related analytics tools, without requiring extended support from IT specialists.

Is business intelligence a part of your statistics production system?

Percentage of surveyed central banks — Graph 1



Source: IFC survey on BI systems in central bank statistics (2019).

The survey provided interesting information on central banks' objectives when using BI, with the following three main aspects (Graph 2):

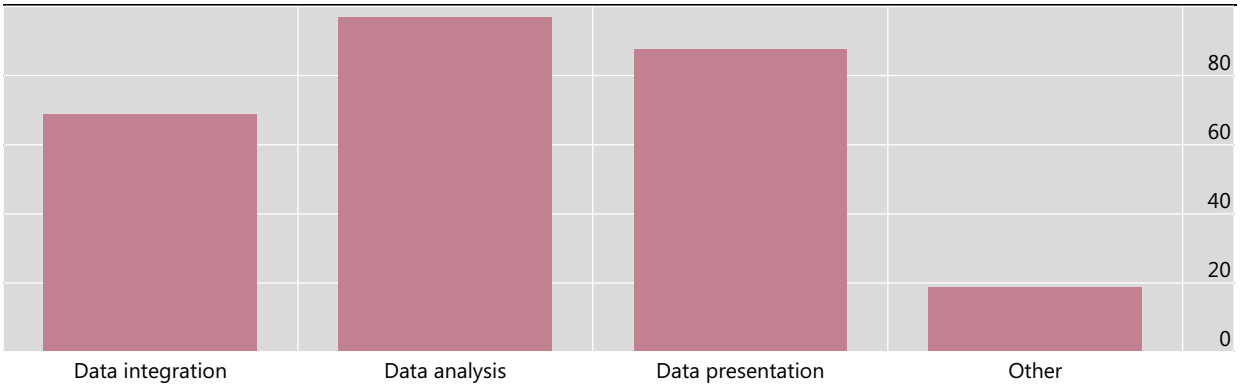
- A key objective is to support **data analysis**. Data analysis is an important element of the business process used by central banks to inform policy decisions. The analytical needs are various, from simple to more complex big data-type requests (advanced analytics).
- A second key objective, as stated by more than 80% of central banks, is **data presentation**. This can be for internal (eg preparation of policy meetings) or external (eg publication of statistics) purposes.
- A third objective, mentioned by two thirds of the respondents, is **data integration**, which aims at combining different data sets to produce a specific type of information.[7] There are many areas in central banking that rely on such

[7] For a presentation of the challenges posed by data integration and related solutions, see Müller (2019).

data integration exercises – from the compilation of financial accounts by drawing on multiple data sources, such as security-by-security databases, monetary statistics, etc,[8] to the production of early warning indicators compiled for policy purposes by drawing on multiple measures of risks.[9] Data integration tasks have been increasing in recent years in the context of the big data revolution and central banks' greater involvement in the collection and use of micro data sets. This has highlighted the need for a unified information system, with the development of comprehensive in-house data warehouses/marts/lakes to host the complex and rapidly increasing amount of data collected.[10]

For which function do you use BI tools?

Percentage of central banks using BI (multiple answers possible) Graph 2



Source: IFC survey on BI systems in central bank statistics (2019).

Apart from these three main categories, the survey pointed to many more avenues for using BI tools. Around 20% of central banks mentioned **other objectives** such as: data elaboration (eg transformation of raw data to yield additional insights), big data analytics (eg implementation of artificial intelligence (AI) and machine learning techniques) and "more traditional" statistical quality processes (eg data cleaning, identification of duplicates, detection of outliers).

4. Types of BI systems used in central banks

Reflecting their various needs, the range of types of BI systems central banks use is quite large. The survey showed that a vast majority of the respondents – between 60% and 80%; Graph 3 – use the following four main types:

- **Data visualisation instruments**, to provide a graphic representation of data

[8] See Tissot (2016).

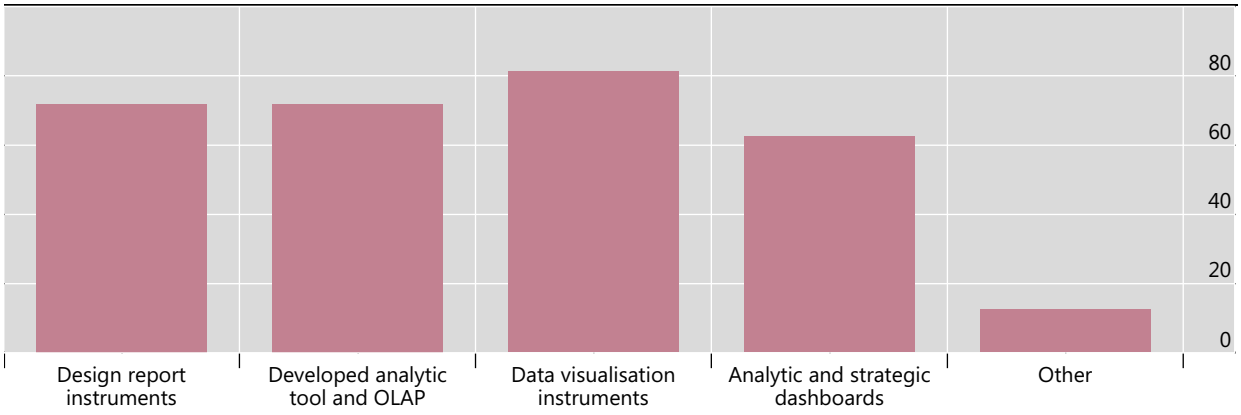
[9] See eg IFC (2017b) in the financial stability area.

[10] For instance, the Deutsche Bundesbank has set up an integrated micro data-based information and analysis system (IMIDIAS) to facilitate the handling of granular data used to support its activities; see Staab (2017). Similarly, the Bank of France has developed a data lake to build a coherent data framework, integrate new big data techniques, and enhance efficiency (Lacroix (2019)).

- **Design report instruments**, to generate the regular reporting of selected indicators
- **Developed analytic tools and online analytical processing (OLAP) technology**, to manage multidimensional databases (or "cubes" made of several "dimensions" that can be flexibly queried by the users to produce a customised view of the data they are interested in)
- **Analytic and strategic dashboards**, to provide a graphical interface and summary reporting tool for selected indicators that is interactive and can be updated from the underlying database in real time

Which type of BI tools do you use?

Percentage of central banks using BI (multiple answers possible) Graph 3



Source: IFC survey on BI systems in central bank statistics (2019).

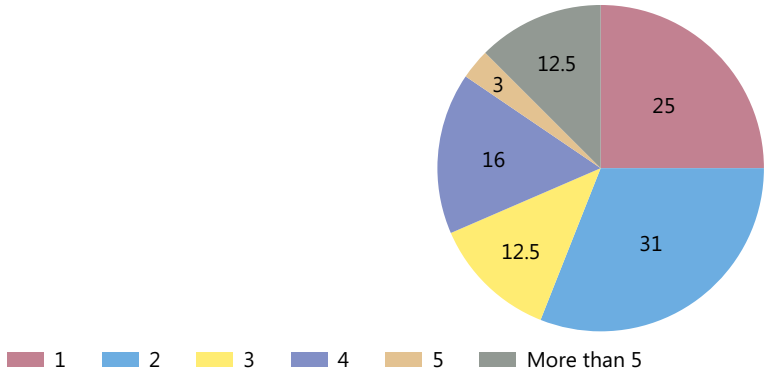
Nowadays, a wide variety of BI tools are available for data analysis, which offer many different features. As a result, one BI tool can be stronger for some functions, for instance data analysis, but weaker for others, for instance data visualisation and publication. Reflecting these respective advantages and limitations, and the fact that statistical production involves multiple steps[11] – eg data collection, storing, cleaning, analysis, presentation and final publication, several tools can be used in parallel.

Indeed, the survey showed that in practice central banks are **using multiple BI tools** that complement each other. As shown in Graph 4, only 25% of the respondents use one single tool. Almost one third use two BI tools, more than one quarter use three to four tools, and one sixth use five or more tools.

[11] As illustrated by the Generic Statistical Business Process Model (GSBPM), https://statswiki.unece.org/display/GSBPM/Generic+Statistical+Business+Process+Model; (UNECE (2019)).

How many BI tools are used in your organisation?

Percentage of central banks using BI Graph 4

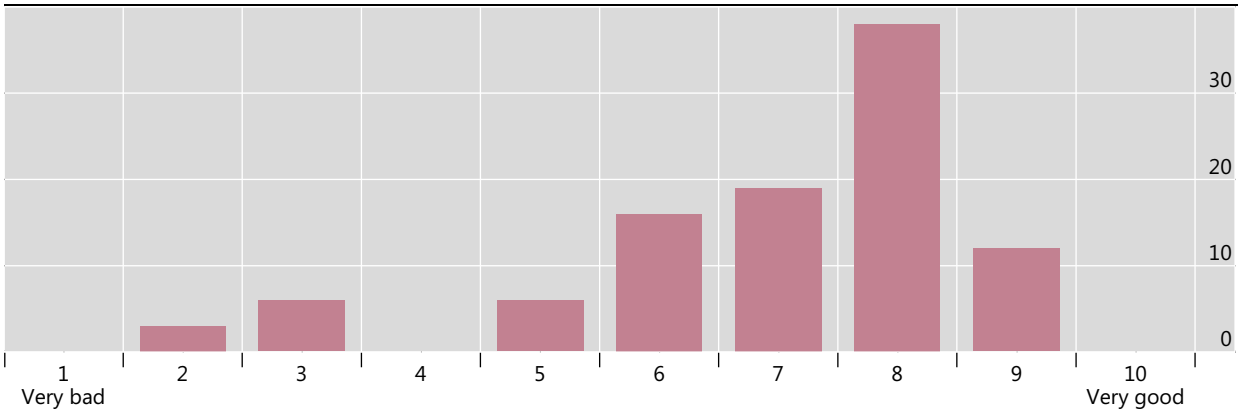


Source: IFC survey on BI systems in central bank statistics (2019).

The survey also showed that there is quite a high **degree of satisfaction** with the use of BI tools within central banks, as illustrated in Graph 5.

What is your level of satisfaction in using BI tools?

Percentage of central banks using BI Graph 5



Source: IFC survey on BI systems in central bank statistics (2019).

5. Technology stack involved in statistical computing

BI tools

Graph 6 shows the most used BI tools in central bank statistics, with three main groups within the large variety of BI products available in the marketplace:

- A **first group** comprises two tools that appear particularly popular, each used by approximately 50% of the central banks, namely Microsoft BI (eg the business

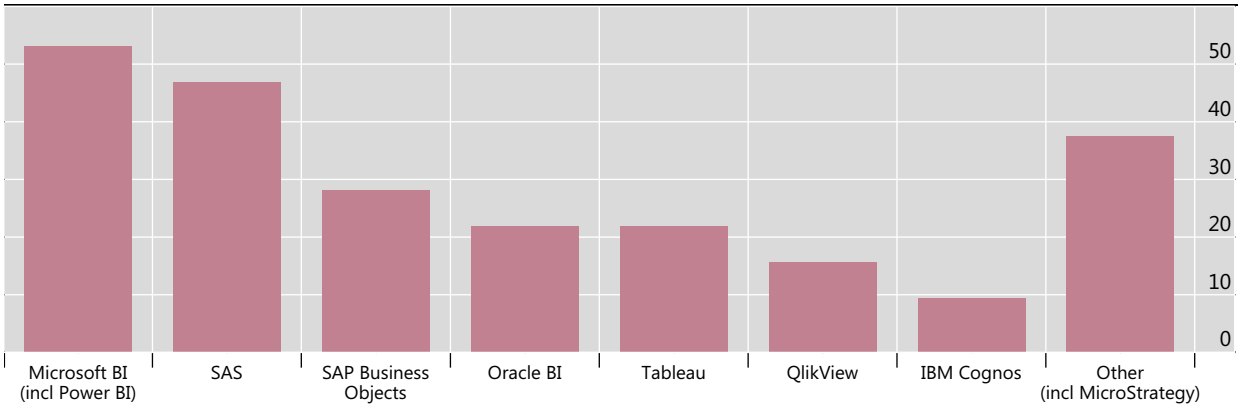
analytics service provided by Microsoft) and the analytics software provided by SAS

- A **second group** of tools, each used by approximately 20% to 30% of central banks, comprises SAP Business Objects (eg information management), Oracle BI (eg data analytics), Tableau (eg data visualisation) and MicroStrategy (eg data mining)
- A **third group** relates to other tools used by a relative minority of central banks, such as QlikView (eg dashboard compilation) and IBM Cognos (eg performance management tasks)

In addition to the major packaged solutions that are commercially available off the shelf, the survey found that several less widespread tools are also used by a significant number of central banks – involving in some cases internal solutions requiring important customisation work.

What is the name of the BI tool(s) used in your organisation?

Percentage of central banks using BI (multiple answers possible) Graph 6



Source: IFC survey on BI systems in central bank statistics (2019).

Statistical package / programming language

Many BI tools are available on a self-service basis so that users can access the information through predefined interfaces that do not require extensive coding as for "traditional", ad hoc programming languages – they are thus often referred to as "fourth generation programming languages" (4GL). These tools can also be customised to better address users' needs; in addition, the complexity of statistical processes within central banks typically requires the use of programming languages for conducting specific analysis.

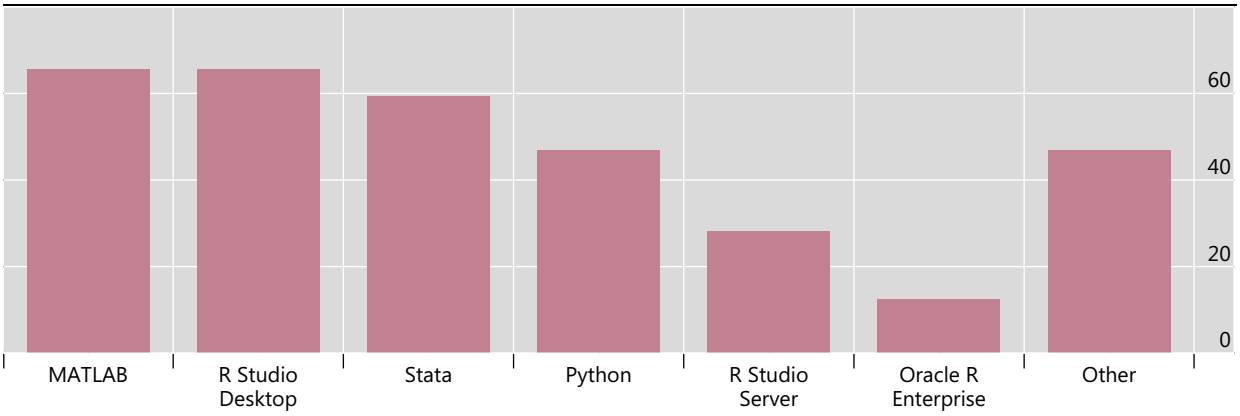
The survey highlighted the following three main product categories:[12]

[12] These results should be interpreted with caution since software package preference may vary across units within the same central bank. For instance, anecdotal evidence suggests that MATLAB is typically preferred by economists and econometricians, while statisticians may tend to rely on other software. In addition, the software/hardware choices reported by respondents may reflect various motivations, such as the ability to address specific business requirements, budget constraints, internal skills etc.

- A first group comprising the **three major products** that are used by around two thirds of central banks (Graph 7): MATLAB (eg time series analysis), Stata (eg panel data econometrics) and the open source R language – noting that several respondents reported using various products based on this language (eg R Studio Desktop, R Studio Server, Oracle R Enterprise).
- **Python** was mentioned by almost half of respondents. Anecdotal evidence suggests that the use of Python is increasing in the central bank community, reflecting its readability characteristics and widespread use within the data science community.
- **Other products** mentioned as important for central banks, though to a lesser extent, were eViews (eg time series econometrics) and the Structured Query Language (SQL), designed, in particular, for querying and managing relational databases (with data organised in tables or "relations").[13]

What is the statistical package / programming language used for statistical computing and graphics?

Percentage of central banks using BI (multiple answers possible) Graph 7



Source: IFC survey on BI systems in central bank statistics (2019).

6. Data visualisation and publications

One important issue when making sense of data relates to visualisation. Data visualisation is an important element when communicating data, since publishing statistics is not just about supplying numbers. How the numbers are presented is important too (Ten Bosch and de Jonge (2008)). In order to make the statistics useful and for knowledge to be generated from bare figures, statisticians must disseminate information that is relevant; present it in a way that users can relate to their own interests; and use language and tools coherent with those used in other contexts (Grossenbacher (2009)). In particular, users get most of their information about the

[13] In addition a growing number of central banks are interested in using non-traditional databases structures – eg "not only SQL" (NoSQL) databases such as Mongo DB; see Fournier (2016) for the related initiative developed at the Bank of France to store heterogeneous data.

world through visual perception, and a well designed graphic can help someone understand the underlying data much better than a simple table of numbers (Carson (2009)).

This is certainly the case for internal communication, and a key task for central bank statisticians is to facilitate the understanding of data for their researchers and to provide background analysis to policymakers. Yet it may be stated that the above sentence is even more important for external communication. Central banks are in charge of disseminating statistics and economic analysis to the public, and they have to carefully manage and monitor the impact of this communication (Nymand-Andersen (2019)). In addition, central banks must provide factual evidence when communicating their policy decisions, as this can be a decisive ingredient supporting their accountability and independence.[14]

The survey addressed **three important issues** faced by central banks in the area of data visualisation. One relates to the choice of the publication medium when communicating to the public, considering the increasing importance of web-based dissemination strategies. The second being how users can interact with the information provided to them. Finally, the third relates to the choice of tools for displaying this information.

Medium for publication/dissemination

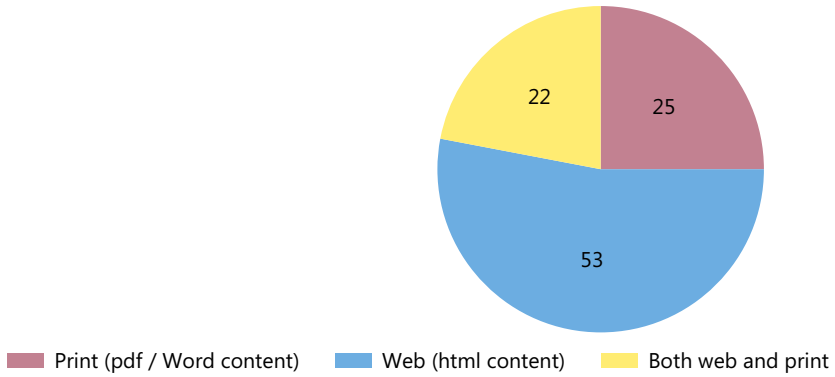
The internet provides an entirely new communications medium that differs in fundamental ways – perceptual and cognitive – from traditional paper-based publications (Levi (2006)). The survey shows that a large majority of central banks have developed **an internet-based dissemination strategy**. Three quarters of the respondents indicated that the web, or a mix of both web and print, is the preferred chosen medium for dissemination and publication processes in their central banks (Graph 8).[15]

[14] On the role of external communication for supporting central bank policy, see eg Blinder et al (2008) and Weidmann (2018).

[15] It should be noted that some central banks have also developed specific mobile applications ("apps") to allow access to their statistics and that offer predefined data visualisations on various topics (see eg the ECB's ECBstatsApp at www.ecb.europa.eu/stats/html/statsapp.en.html).

What is the primary medium for the publication/dissemination process in your organisation?

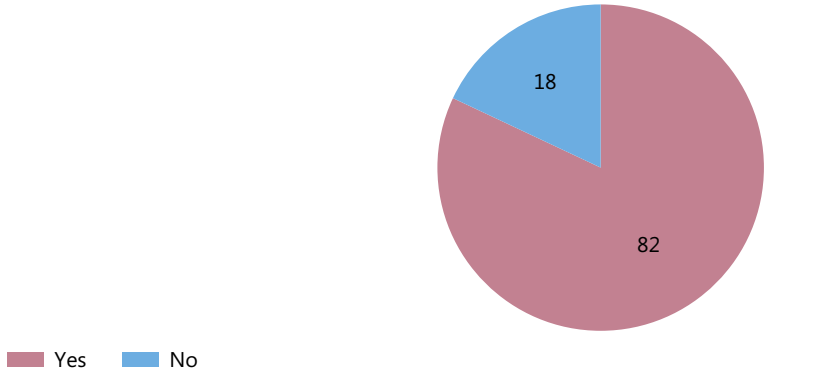
Percentage of central banks using BI — Graph 8



Source: IFC survey on BI systems in central bank statistics (2019).

If print, do you intend to move to the web within the next five years?

Percentage of central banks relying on print as their primary medium — Graph 9



Source: IFC survey on BI systems in central bank statistics (2019).

In contrast, only one quarter of central banks reported that they exclusively rely on printed documents for dissemination/publication purposes. Among those central banks relying on print as their primary medium, there is a clear trend to move to the web as the primary medium, with the vast majority (82%) intending to make the transition over the next five years (Graph 9).

Static versus dynamic graphical content

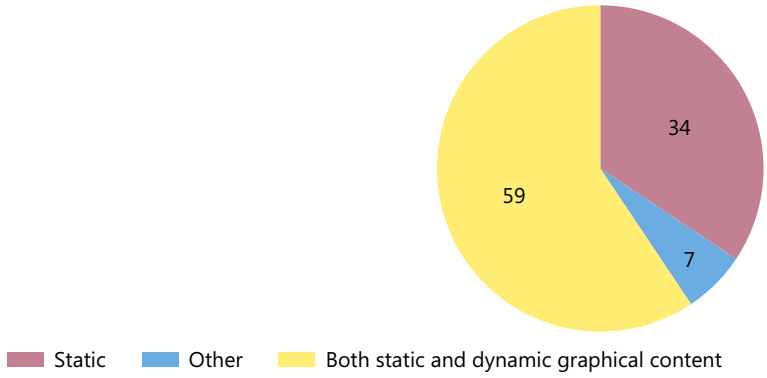
With the greater importance of internet-based data dissemination, it is becoming easier to offer users the possibility to interact with central banks' publications in a dynamic way. In contrast to print, which is by definition static, graphs displayed on the web can be updated in real time as new information becomes available, and can allow users interact with them and make modifications depending on their particular

needs – for instance, to select a specific time frame, frequency or list of countries.[16] Recent years have been marked by an increasing demand for more complex visualisation techniques, with users asking for more appealing graphs and/or advanced analytical tools (eg in the area of network analysis; IFC (2019)). Many central banks have started to develop dashboard content, typically through tools such as Tableau and Microsoft BI, thereby allowing the grouping of data from different sources and providing a visual interface for users to view and interact with multiple topics of interest.

Nearly two thirds of central banks reported that they use **a mix of both static and dynamic graphical content** when presenting information on their websites (Graph 10). Yet one third continue to disseminate only static content. Moreover, no reporting central bank has completely abandoned this practice and exclusively offer dynamic graphs alone.

Do you disseminate static and/or dynamic graphical content on your corporate website?

Percentage of central banks using BI — Graph 10



Source: IFC survey on BI systems in central bank statistics (2019).

Software for generating graphical content

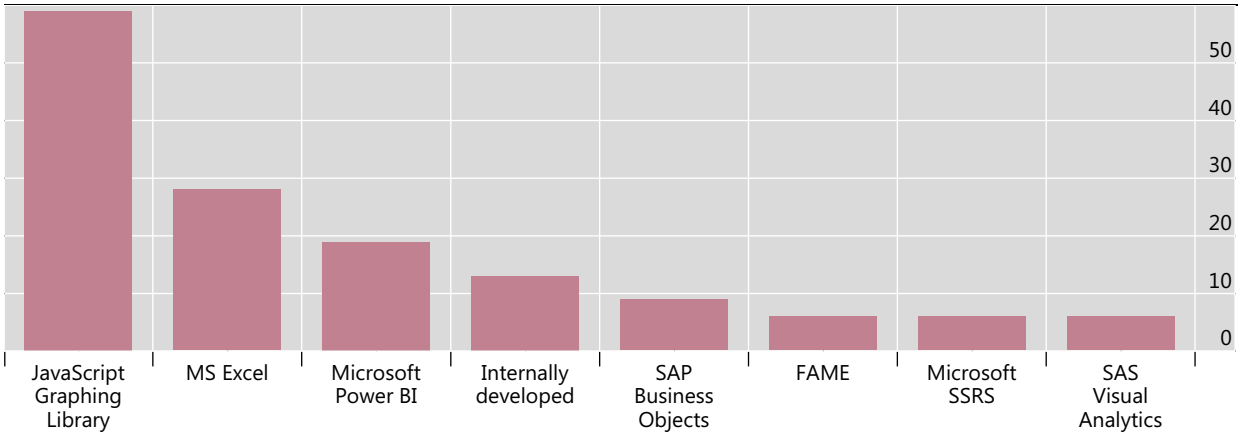
The final section focused on the **most popular methods** used to generate graphical content for publications and the corporate website. A variety of analytics and/or reporting packages is used within each institution to serve this need. The survey shows that most central banks use JavaScript[17] graphing libraries to produce graphs for dissemination purposes (Graph 11), as indicated by over half of the respondents.

[16] For example, the dashboard-based online interactive disclosure tool developed by the Reserve Bank of New Zealand to promote the dissemination of banking regulatory disclosures (Irrcher (2019)).

[17] JavaScript (JS) is a high-level, interpreted scripting language, typically used for the development of web pages.

Which software do you use to generate graphical content for publications and the corporate website?

Percentage of central banks using BI (multiple answers possible) Graph 11

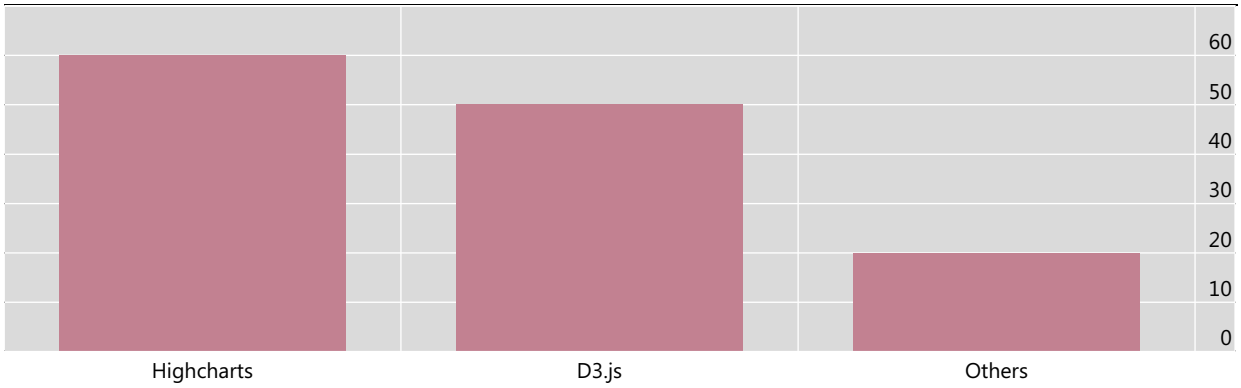


Source: IFC survey on BI systems in central bank statistics (2019).

Within the JavaScript libraries, **Highcharts and D3.js** are the most common, being used in at least 50% of the respondents (Graph 12).[18] Almost one third of the respondents indicated using more than one JS Library within their institution.

Which JavaScript graphing library do you use?

Percentage of central banks using a JavaScript graphing library (multiple answers possible) Graph 12



Source: IFC survey on BI systems in central bank statistics (2019).

Other tools such as the spreadsheet (MS Excel) and business analytics service (MS Power BI) provided by Microsoft are also used frequently in central banks, by approximately one third and one fifth of the respondents, respectively. In addition, a significant number of institutions (more than a tenth) have developed customised

[18] Highcharts is a JS software library for publishing charts directly on the website; D3.js (or Data-Driven Documents) is a library for producing dynamic, interactive data visualisations in web browsers.

internal solutions for this purpose, in most cases by transforming "off the shelf" packages (Graph 11). Lastly, a few central banks reported using statistical computing software such as MATLAB and R (Graph 7) also for publication.

References

Blinder, A, M Ehrmann, M Fratzscher, J De Haan and D-J Jansen (2008): "Central bank communication and monetary policy: a survey of theory and evidence", *NBER Working Paper*, no 13932, April.

Carson, C (2009): "Numbers as pictures: examples of data visualisation from the Business Employment Dynamics program", *Statistical Survey Paper*, US Bureau of Labour Statistics, October.

Drozdova, A (2017): "Modern informational technologies for data analysis: from business analytics to data visualization", *IFC Bulletin*, no 43, March.

Fournier, J (2016): "Regulatory reporting à la française: the Banque de France Data Lake", presentation at the European Institute of Financial Regulation Workshop, 20 September.

Grossenbacher, A (2009): "Official statistics beyond web 2.0. Developments, rewards and risks to come", presentation at the conference "Statistics – investment in the future 2", Prague, 14–15 September.

Irrcher, T (2019): "Enriching disclosures: the Bank Financial Strength Dashboard", *IFC Bulletin*, no 49, January.

Irving Fisher Committee on Central Bank Statistics (IFC) (2017a): "Big data", *IFC Bulletin*, no 44, September.

——— (2017b): "Data needs and statistics compilation for macroprudential analysis", *IFC Bulletin*, no 46, December.

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

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

——— (2015): *The financial crisis and information gaps – Sixth Implementation Progress Report of the G20 Data Gaps Initiative*.

Lacroix, R (2019): "The Bank of France datalake", *IFC Bulletin*, no 50, May.

Levi, M (2006): "Statistics in cyberspace: how has the internet changed the U.S. Bureau of Labour Statistics", ECE/CES/SEM.54/17, paper presented at the Joint UNECE/Eurostat/OECD Seminar on the Management of Statistical Information Systems (MSIS), Sofia, 12 May.

Müller, S (2019): "Creating comprehensive data worlds using standardisation", *IFC Bulletin*, no 49, January.

Nishimura, K (2012): "Market intelligence, market information and statistics in central banking", speech at the 6th IFC Conference, Basel, 29 August.

Nymand-Andersen, P (2019): "Modernising statistics: communicating the wealth of statistics in a digital age", paper presented at the 62nd congress of the International Statistical Institute, Kuala Lumpur, August.

Prokunina, E (2019): "Creating integrated data world within the Bank of Russia: strategic vision and first results", paper presented at the 62nd congress of the International Statistical Institute, Kuala Lumpur, August.

Staab, P (2017): "The Bundesbank's house of micro data: Standardization as a success factor enabling data-sharing for analytical and research purposes", *IFC Bulletin*, no 43, March.

Ten Bosch, O, and E de Jonge (2008): "Visualising official statistics", *Statistical Journal of the International Association of Official Statistics* (IAOS), vol 25, pp 103–116.

Tissot, B (2016): "Development of financial sectoral accounts: new opportunities and challenges for supporting financial stability analysis", *IFC Working Papers*, no 15, November.

——— (2017): "Using micro data to support evidence-based policy", paper presented at the 61st congress of the International Statistical Institute, Marrakesh, July.

——— (2018): "Providing comparable information to assess global financial stability risks", *Eurostat Statistical Reports*, KS-FT-18-001, January.

United Nations Economic Commission for Europe (UNECE) (2019): *Generic Statistical Business Process Model (GSBPM)*, version 5.1, January.

Weidmann, J (2018): "Central bank communication as an instrument of monetary policy", lecture at the Centre for European Economic Research, Mannheim, 2 May.

Appendix: List of countries that responded to the survey

1. Argentina
2. Australia
3. Austria
4. Brazil
5. Canada
6. Chile
7. Croatia
8. Denmark
9. Estonia
10. Finland
11. Germany
12. Hungary
13. Italy
14. Japan
15. Korea
16. Latvia
17. Lebanon
18. Luxembourg
19. Malaysia
20. Malta
21. Mexico
22. Montenegro
23. Morocco
24. New Zealand
25. Nigeria
26. North Macedonia
27. Poland
28. Portugal
29. Russia
30. Singapore
31. Slovakia
32. South Africa
33. Spain
34. Sweden
35. Turkey