Claudia M. Buch

Vice-President of Deutsche Bundesbank

**Can Technology and Innovation Help?
New Data Generating Possibilities**[i]

Speech prepared for the
European Central Bank Conference on Statistics
Frankfurt, 12 July 2018

New data, technologies, and IT developments significantly alter production processes and distribution channels in the economy. Comparative advantage depends, to an increasing degree, on the capability to select the right data and to take "data driven" decisions. At the same time, regulators are challenged in the areas of consumer protection, competition policy, and data confidentiality.

Central banks are affected as well – although they do not "compete" on markets. Mandates of central banks comprise the provision of public goods: price and financial stability. This, in turn, requires analytical strength and making best use of data and methodologies. For policy analysis and research, statistics and reliable data based on standardization and quality assurance are crucial inputs. New technologies can contribute to a better quality of data and better analytical work.

In this talk, I argue that technology and innovation affect the production of central bank statistics and the production of knowledge. I outline how existing and new sources of data can be used more efficiently and how new analytical tools can improve production processes (Section 2). At the same time, challenges need to be managed (Section 3). The demand for skills changes away from manual to analytical and communication tasks, cost-benefit analysis needs to be improved, and knowledge sharing within and across central banks is crucial.
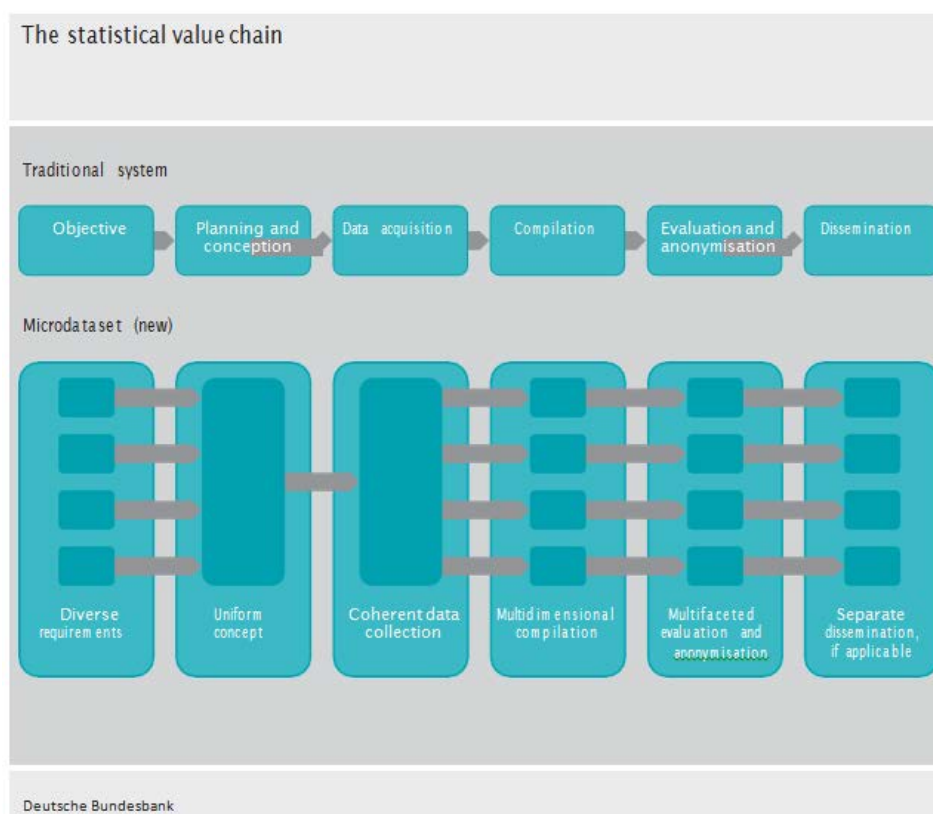
1 What's new?

Digitalization has the potential to overhaul the traditional system of compiling (central bank) statistics. In the traditional system, reporting forms are used to obtain targeted information meeting a specific purpose as, for instance, the compilation of balance of payments statistics. Interest is in aggregated data, not the underlying micro- or granular data.

This philosophy has changed: Digitalization lowers the costs of processing and storing data due to general technological advancement in hard- and software components. Digital storage and data processing power has increased substantially throughout the past decades, and costs have fallen. As a result, central banks are in a position to collect larger

amounts of data which can be organized and processed automatically. Given the ever increasing amounts of data, some degree of automation is inevitable. Granular, micro- data sets, which comprise security-by-security or loan-by-loan can be processed and used for multiple purposes. Such data can be tailored to specific needs of users. All this fundamentally alters the statistical value chain (Figure 1).

Figure 1: The statistical value chain

The statistical value chain

Traditional system

| Objective | Planning and conception | Data acquisition | Compilation | Evaluation and anonymisation | Dissemination |

Microdataset (new)

| Diverse requirements | Uniform concept | Coherent data collection | Multidimensional compilation | Multifaceted evaluation and anonymisation | Separate dissemination, if applicable |

Deutsche Bundesbank

The following trends are not new – but their speed has accelerated:

1. Granularity of data: Central banks increasingly move away from the collection of aggregate data, compiled according to pre-defined frameworks and designed for specific purposes. Instead, reporting agents transmit microdata, which are then compiled at the central bank, according to the specific needs of users.

2. New sources of data: New sources of data and in particular "Big Data" have become a buzzword. "Big Data" feature characteristics of Volume, Velocity, Variety, Variability, and Veracity. New datasets differ from traditional sources of statistics because they are unstructured or because they are provided by private sources.

3.   New analytical technologies: Machine learning (ML) and artificial intelligence (AI) date back to the late 1950s.[2] AI research deals with the study of "intelligent agents" – a system which reacts to changing circumstances and learns from experience.[3] ML methods are used to learn about patterns in data and to use this information in order to make predictions for other, similar cases.[4]

4.   New IT developments: Significant progress has been made in terms of dealing with large volumes of data. Moore's law, which was formulated in 1965, predicts that the complexity of integrated circuits doubles every one to two years.[5] Not only has the hardware infrastructure become more efficient, distributed computing and cloud computing also allows for the decentral processing of tasks. Already in the early 2000s, commentators claimed that "IT doesn't matter"[6] due to the rapid price deflation of computer chips, data storage, and data transmission.

Why have all these trends accelerated and reinforced each other in recent years? In 1934, Schumpeter argued that innovations basically result from the "changes of the combinations of the factors of production […] They consist primarily in changes in methods of production […]"[7] Innovation thus happen as innovators recombine existing productive components throughout a "creative" process. In this context, "General Purpose Technologies" are crucial because they potentially impact several sectors or the economy and allow for new methods of production by re-combining factors of production. For example, the combination of electricity with other means of production made innovations in the entire economy possible. Similarly, the internet boom of the late 1990s has been described as an example of "combinatorial innovation", resulting in a situation with relatively little scarcities in input factors and very low marginal costs.[8]

New data, IT-Infrastructure and analytical tools may bring about a new wave of "combinatorial innovation". Most input factors are of immaterial nature as are statistical methods and software. This could potentially explain the enormous speed with which new applications occur.

Given these trends, how do central banks generate knowledge? According to Boos (2018), data are the starting point. Connecting different data points generates "infor-

mation", bringing this information into context generates "knowledge", using it generates "intelligence", and reflecting on the use of data generates "wisdom".[9] Each of these stages requires the input of different production factors, reminiscent of a production function:

$$Y = f(A, K, L, D)$$

where Y can represent statistics, knowledge, or intelligence.

The IT-infrastructure is captured by (K), i.e. any hardware related component that is necessary for the processing of data. The costs for IT-Infrastructure have fallen tremendously throughout the years which, viewed in isolation, would reduce the costs of producing knowledge. Data (D) are collected by the central banks from reporting agents, according to regulations, or bought from commercial providers. Technological progress often augments the inputs of labor (L) as it gives the statistician tools at hand that make her work more efficient. Either more data can be dealt with or the quality of data could be improved, while holding working-hours constant. Technology (A) represents software-tools, algorithms and statistical methods needed for the handling and processing of data.

Knowledge produced by central banks is essentially the result of a combination of the input factors data, labor, capital, and technology. Importantly, technology leverages any unit of labor that is invested into the production of knowledge. Machine learning algorithms are an example of how technology (A) can augment the efficiency of labor (L) because it lowers the quantity of routine tasks that need to be performed manually. This, in turn, allows the statistician to focus on non-routine tasks, e.g. interpretation or communication with users. ML-algorithms can support checks of data quality. Algorithms can provide predictions of the results that manual data quality checks would have produced by checking each data point for plausibility.

Record linkages are another example for how algorithms can support statistical work: Matching of large, granular data sets without unique identifiers manually is cumbersome, time-consuming, and prone to produce errors. Machine learning approaches can be used to match data sets based on probabilistic scores.[10, 11]

Using new technologies cannot only allow using existing factors of production more efficiently, thus potentially generating returns to scale, it may also enhance the division of labor. For example, within a team of data scientists, each team member could focus on a specific subset of methods, say, in the field of machine learning, for example one in supervised and one in unsupervised learning. Cooperation with colleagues, each one bringing in "expert knowledge" at a particular stage of the project, creates a process of mutual learning with the potential to increase productivity.

Reaping synergies along the production process is another potential benefit of technological progress. Increases in quantities of data, particularly if collected and stored according to well-defined information models, allow for more standardized procedures. Certain tasks can be centralized along the production chain. The "statistical value chain" changes: the diverse requirements and objectives that statistics faces are not treated in isolation anymore. Rather, uniform concepts of data collection, processing and dissemination make it possible that each of these heterogeneous needs is dealt with by a specialized unit dedicated to one step of the production chain.

2   What are the implications for the production of statistics?

Making better use of (existing) data

Traditionally, central banks collect a large number of datasets which are then stored in different business areas. In the Bundesbank, for example, bank-level data are collected by the Statistics Department and the Banking Supervision Department, information on payments transactions are collected by the Payments System Department, and additional micro-data are available in the Markets Department.

These data contain potentially useful information for analytical policy work in different business areas. However, combining the data has traditionally not been feasible due to legal restrictions because of confidentiality requirements or data protection rules. Very practically, company identifiers or software tools often differed. The fixed costs of using these data have, therefore, been high, which severely constrained their use in day-to-day

policy work. Merging different micro-level datasets requires, for example, identifiers, cumbersome matching algorithms and/or decisions to be taken for matching reporting entities. Studies using micro data are often not readily comparable because analysts made different choices when compiling datasets. As a result, existing micro data have often not been used efficiently in the past for policy work.

In order to address these issues and making best use of existing data, Deutsche Bundesbank initiated, in 2013, its project IMIDIAS – the Integrated MIcroData Information and Analysis System. IMIDIAS focuses on the cross-sectional microdata storage, integration and usage of relevant microdata sets. The initiative consists of different components:

- an in-house steering committee representing all owners and users of data across the business areas of the Bundesbank,

- the House of Microdata (HoM) as an integrated data warehouse, the content of which are datasets which are relevant for internal analysis and research,

- the Research Data and Service Centre (RDSC) as a facilitator for internal and external researchers to get access to selected microdata of high quality and as a contact point for data providers and data users.[12] The RDSC offers access for non-commercial research to (highly sensitive) micro data of the Bundesbank. All researchers (internal and external) have the same transparent access to microdata of the Bundesbank. The RDSC has been accredited by the German Data Forum.[13]

Currently, 20 employees are working at the RDSC, which provides 12 work places for external guest researchers. Currently, over 300 active projects of more than 160 institutions are affiliated with the RDSC. In 2017 alone, the RDSC received about 130 project applications, of which 73 were realized. Reasons why projects are not realized range from inadequacy of the available data, the project not being classified as independent research, to idiosyncratic factors. In order to preserve confidentiality of sensitive microdata, 250 output controls for research projects were conducted. By now, the first papers of "RDSC users" have been published.

Developing new sources of data

Analytical work in central banks traditionally focused on macroeconomic aggregates like GDP, inflation, credit, or interest rates. The advantage of these data is that they can easily be shared and combined within and across institutions. The fixed costs of using the data are thus relatively small. Yet, the analytical insights based on aggregate statistics can be limited.

For example, understanding the build-up of risks in financial systems and the transmission of monetary policy shocks requires sufficiently granular data. Micro-level heterogeneities within and across sectors and countries, play a crucial role in determining macro-level aggregates.[14] Therefore, the analysis of macroeconomic measures alone proved to be insufficient as they cannot account for the relevant heterogeneities and frictions.

The analysis of micro-level, granular data has made significant progress over the past two decades. Availability of new data sources has improved, not least through research data centers, and new empirical methodologies have been developed. Many of these improvements have taken place in the areas of labor economics, international trade and investment, or in the identification of effects of monetary policy.

Within the European System of Central Banks (ESCB), there has been a paradigm shift away from the production of aggregates towards more granular statistics:

- Since 2005, granular data on securities have been collected and proceeded in the Centralised Securities Database (CSDB). This dataset contains information on 6.8 million actively traded securities (1.9 million from Germany) issued by approximately 670.000 issuers (30.000 from Germany).

- Since 2014, the Eurosystem has been collecting quarterly data on securities holdings with the Securities Holdings Statistics Database (SHSDB). 6 million observations on sector-by-sector holdings are reported from 25 national central banks, roughly 2 million observations by the Bundesbank.

- From September 2018, the Analytical Credit Datasets (AnaCredit) will provide information on loans provided to legal entities and credit risk, providing crucial inputs for monetary policy analysis and the surveillance of financial stability risks. The novelty of AnaCredit is that it provides information on banks and borrowers in a harmonized way at the European level, unlike existing national credit registries which often differ in scale and scope.

- For all reporting systems, reference data are collected in the Register of Institutions and Affiliates Data (RIAD), including almost all financial but also non-financial firms. High quality reference data is a necessary condition for the production of statistics as they allow to uniquely identifying the reporting population and, for example, to derive group structures of reporting agents. Hence, information stored in RIAD will allow matching information across different data sets.

Collecting and processing these data is costly. Costs arise for the reporting firms which need to maintain IT infrastructures based on in-house information systems which provide the relevant data for external users. But costs arise also within central banks as new infrastructures have to be set up. Planning new micro-data projects and balancing costs and benefits thus requires a longer-term horizon. The AnaCredit project, for example, has been initiated in the year 2011 and data will become available in 2018 – i.e. after seven years. This shows that careful planning, cost-benefit analysis, and consultation with the industry must be crucial elements of any new data project.

In parallel, researchers and statisticians at central banks have also tested other external, sometimes less structured and more "unconventional" data sources. Matching internal with external data is often useful. At the Bundesbank, pilot projects make use of rather unstructured, external data stemming from internet sources. For example, Google search data are used in several research projects to analyze the behavior of depositors during the financial crisis[15], of developments on the German residential mortgage market[16], or potential improvements in forecasting models. [17]

New analytical tools

The increasing number of microdata sets also poses challenges for data production and analysis. New tools are needed for handling these data and for the extraction of the relevant knowledge. Therefore, several projects in Statistics at Bundesbank use machine learning techniques:

- Quality assurance and checking for errors in the data is, typically, a manual task. Such routine tasks are labor intensive and may not detect all errors. In order to verify implausible securities data for the Securities Holdings Statistics, a random forest algorithm is thus used. This algorithm is based on decisions taken in the past by compilers to flag reported securities as erroneous. A machine learning algorithm approximates the human decision making process.

- The Register of Institutions and Affiliates Data (RIAD) system develops a prototype for record linkages based on machine learning tools.

- The Bundesbank's Research Data and Service Center (RDSC) makes use of record linkage methods that are based on supervised machine learning to automatically link several data sets on companies. The result is a mapping table with company identifiers.

- Seasonal adjustments are improved by using random forests of conditional inference trees. This approach assesses the seasonal nature of a given time series with a high degrees of accuracy and can be used to improve tests for seasonal patterns in the data.

Given the increasing importance of European statistical projects and the potential to learn from experience, international cooperation is crucial. Under the auspices of the ECB's Statistical Committee (STC), a Task Force on machine learning has been established with the aim to exchange experiences and ideas for machine learning applications in central banks statistics.

3 What are the implications for the production of knowledge?

The production of central bank statistics is not an end in itself but a central element in the creation of knowledge – knowledge about the structure and functioning of the financial system, knowledge about its interaction with the real economy, and, not least, knowledge about the effects of policy measures affecting the financial system. The last step – the ex-post evaluation of policy effectiveness – is a crucial element of the transparency and the accountability of central banks.

Data alone do not provide the necessary information. Data need to be combined with appropriate models – be they formal or informal – and with appropriate empirical techniques to generate knowledge. Only then can we make statements about causes and effects of policy measures. In short, we need to go beyond correlations and should aim at establishing causalities.

Improving identification through heterogeneity

Identification of policy effects using aggregate data is challenging and often not possible. The identification of causal effects thus requires sufficiently detailed data. Heterogeneity across institutions, individuals, or countries is often the key to identification. Seminal theoretical work on the functioning of credit markets, relevant frictions, and implications for monetary transmission dates back 20 years.[18] For empirical research in this field, loan-level microdata comprising detailed information on lenders, financing conditions, and the provision of collateral is of utmost importance. Throughout the past years, studies based on micro- (firm-level) data have provided important insights:

- Analyzing the allocation of bank credit to firms and the implications for productivity requires detailed micro-level data. Such data show that, in OECD countries, lending to weakly performing firms has tended to be stronger since the mid-2000s, with negative consequences for the performance of stronger firms. Performance is measured in terms of investment and employment.[19] In a related approach, using

European data, this mechanism is linked to the use of unconventional monetary policies.[20]

- Work with micro-level data on banking has a long tradition in Italy. Using granular data for a period from 2004–2013, one result of this research is that weaker banks have been lending to weaker firms during the Eurozone financial crisis.[21] Yet, this misallocation of credit at the firm level does not seem to be a key driver of the aggregate volume of credit. But "distress lending" slowed down the exit of weak firms. This adjustment along the "extensive margin" may be a channel through which credit allocation also affects overall performance.

- A potential misallocation of credit is, however, not confined to countries with weak growth performance. Small and mid-sized firms with negative returns on asset and net investment increased their leverage in the period from 2010 to 2014. These patterns are observable, both in stressed economies (such as Greece, Ireland, Portugal, Slovenia, and Spain) and in non-stressed economies (such as Germany and France).[22]

- Since 2014, the Nederlandsche Bank has been collecting loan level data on mortgages which provide important information for the analysis of risks building up on mortgage markets.[23] These detailed data contain information on the individual mortgage lenders portfolio at quarterly frequency. The variety of attributes that are recorded allow for the computation of various descriptive statistics as (joint) distributions of interest rates, outstanding amounts, the ratio of performing loans, or the age of borrowers. Similar data for the German mortgage market a not available, as has recently been highlighted by the IMF.[24]

4   Managing the challenges

New, granular sources of data, new analytical technologies, and new IT-developments promise significant improvements for the production and enhancement of (existing) statistics, the generation of new information, and ultimately for enhancing knowledge. Yet,

reaping these benefits and coping with the challenges, requires managing the changing demand for skills, improved cost-benefit analysis, and improved knowledge sharing and international cooperation.

Managing the demand for skills

New data and technologies alter the demand for skills in the production of statistics and knowledge. The use of "algorithms" does not lead to a reduced demand for labor due to the "replacement" of employees by machines. Quite to the contrary: Tasks now being performed by humans will move away from manual routine tasks towards the provision of services, including the explanation of statistics to users. New technologies will complement, not substitute workers, enabling them to be more productive and to enhance their existing skills.

To give an example: The arrival of the personal computer or new spreadsheet software profoundly altered the way finance and accounting departments work. But this did not make the affected employees disappear or obsolete at all.[25] Another example in this context is the introduction of ATMs and employment in banking. Contrary to initial expectations, the rapid spread of ATMs in the U.S. from the 1970s on did not lead to a significant decrease of bank teller jobs.[26]

The key to making best use of new technologies and managing the transition is training – at school, universities, but also continuous learning on the job. Many universities already alter their curricula by introducing study programs that better prepare future graduates for the change in required skills. One example is the European Master in Official Statistics (EMOS), a network of postgraduate master degree programs fostering the cooperation between universities and official statistics as data producers. The main goal is to train postgraduate experts in order to be able to cope with the rapid changes that statistical offices are facing today.

As regards the production of knowledge, skill sets focusing on the use of aggregate statistics and empirical methodologies using macro data will increasingly be complemented by skills for the use of micro data and identification mechanisms based on micro-econometric tools. Many promising developments in empirical economics bridge the gap between

macro- and micro-analytical tools. These include factor-augmented vector-autoregressive (FAVAR) models, which combine traditional structural VAR-models with factor analysis, summarizing large numbers of time series to a smaller number of "factors". This, in particular, allows to condition VAR models on a richer set of information which are likely to better represent actual decision makers' information sets. Making best use of such methodological developments, which are particularly useful in "data-rich" environments, may require explicit training programs for central bank analysts traditionally working with more aggregate datasets.

There are several initiatives within the Bundesbank as well as in cooperation with other institutions which allow employees to acquire the necessary skills for e.g. working with microdata. The Bundesbank is represented in the Task Force Big Data of the European Statistical System, in order to exchange views and experiences on how new data sources could be implemented into official statistics. At the national level, Deutsche Bundesbank and the Federal Statistical Office have conducted a joint workshop on the use of machine learning techniques for data production. Further, internally, the Bundesbank runs several training activities that allow statisticians to improve their technical (programming) and methodical skills.

Improving cost-benefit analysis

At first sight, granular data seem to increase the reporting costs for industry. However, in the medium to long term, granular reporting frameworks have the potential to reduce costs for the reporting population and central banks, where at the same time the analytical value of the data increases. Ultimately, these benefits may also accrue to industry as better analytical tools in central banks improve policymaking. In addition, feedback loops can provide reporting firms with input for their own analysis and benchmarking.

Improving the production of statistics and knowledge requires sound cost-benefit analysis, industry consultation, and long-term strategic planning. Costs of implementing new reporting systems should amortize over time. As stability of new reporting frameworks has materialized, aggregate statistics and indicators can be calculated from underlying microdata. This, however, requires consistency between different data sets. To achieve this

objective, the long-term approach of the European System of Central Banks (ESCB) is to collect data from banks in a standardized way to integrate existing ESCB statistical frameworks, as far as is possible, across domains and countries. The main objective of this approach is to increase the efficiency of reporting and to lower reporting costs, while providing users with high-quality data. This approach relies on clear concepts, definitions and requirements, which help automate data processing and enhance data quality. The banking industry and other reporting agents are involved in this process already today through consultations in the course of the ECB Merits and Costs Procedure and in public consultations on new or enhanced ECB Regulations on European statistics.[27]

One element of the strategic approach of the ESCB is the ESCB Integrated Reporting Framework (IReF).[28] This projects aims at integrating existing ESCB statistical data requirements for banks into a single framework. It consists of two parts: First, an integrated set of reports for banks, which can ultimately replace national reporting templates. Second, transformation rules compiling the statistics required by authorities, possibly to be shared with the stakeholders involved.

The other element is the Banks' Integrated Reporting Dictionary (BIRD).[29] One key objective of BIRD is to both, lower the costs of reporting statistics and enhance data quality. BIRD consists of a harmonized data model which describes data to be extracted from banks' internal IT systems to derive reports required by authorities, together with rules on how to transform the data. In this sense, BIRD is a "public good" that can be used by banks and other data providers. Participation is voluntary. At this moment, BIRD covers the reporting requirements of AnaCredit and securities holdings statistics (SHS), while coverage of the Financial Reporting for supervisory purposes (FinRep) is under development.

Enhancing knowledge sharing

Sharing knowledge and information within and across central banks can contribute significantly to making best use of new technological developments. Three initiatives can serve as examples:

- The International Network for Exchanging Experience on Statistical Handling of Granular Data (INEXDA) shares information on methods for harmonizing access procedures and metadata structures, thus fostering the efficiency of statistical work with granular data.[30] Its ultimate aim is to facilitate the use of granular data for analytical purposes, and it provides a platform also for users outside the participating institutions, of course observing confidentiality regulations. Beside knowledge sharing, INEXDA promotes the G20 Data Gaps Initiative II (DGI-2), in particular recommendation 20, addressing the accessibility of granular data.

- The International Banking Research Network (IBRN) is a global network of central bank researchers which aims at making better use of micro data in central banks without having to share the (confidential) data. Teams of researchers across central banks collectively develop analytical models, apply these models to their national datasets, and the findings are compiled in a meta analysis. Research in the network contributes to an improved understanding of the role of global banking in the transmission of (policy) shocks across countries.[31]

- Internal processes may need to be adjusted in order to benefit from new technologies and data. Procedures for the governance of data need to be improved with regard to structuring and organizing knowledge. Currently, information that is available within organizations may not always be systematically collected and may be stored in a decentralized manner. As a consequence, the accessibility, i.e. "searchability" of these data is not always granted for all potential users. In order to structure such information and to make it available, the Bundesbank is working on a dynamic and adaptable repository which connects the relevant stakeholders like data producers or researchers in an interactive and learning knowledge map.

5 Summing up

Overall, technology and innovation can contribute to improvements in the production of statistics and, ultimately, knowledge in central banks. Reaping these benefits requires good planning, commitment of leadership, and good dialogue with all stakeholders.

At this point in time it seems useful to "start small" and learn from experience. Different input factors need to be combined to generate "innovative processes", and it is not clear a priori which combination generates the best results. Experimenting with different models can thus be useful.

New approaches and ideas are needed and welcome. The "statistical value chain" is changing profoundly given to the increased demand for microdata. This in turn is likely to have implications – for required skills, data-governance, and the organization of work. We need to "think outside the box" to best management these challenges.

In this sense, the establishment of units which think and organize work across the traditional "silos", which establish a culture of knowledge sharing can be useful to experiment with and implement useful, innovative ideas.

[i] I would like to thank Stefan Bender, Jürgen Häcker, Claudia Wieck, and Simon Oehler for most helpful contributions to an earlier draft. Any errors and inconsistencies are my own.

[2] Samuel, Arthur L. (1959). Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development.

[3] Poole, D., Mackworth, A., Goebel, R. (1998). Computational Intelligence: A Logical Approach. New York: Oxford University Press.

[4] Prediction here is meant in a wide sense and is particularly not restricted to the case of forecasting in longitudinal settings.

[5] Moore, Gordon, E. (1965). Cramming More Components onto Integrated Circuits. Electronics, pp. 114–117.

[6] Carr, Nicholas G. (2003). IT doesn't matter. Harvard Business Review, May Issue.

[7] Schumpeter, J. (1934). The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle. p. 66.

[8] Varian, Hal R. (2003). The Economics of Information Technology. Raffaele Mattioli Lectures. Cambridge University Press.

[9] Based on an interview with Chris Boos in *Frankfurter Allgemeine Sonntagszeitung*: http://www.faz.net/aktuell/wirtschaft/diginomics/kuenstliche-intelligenz-ki-pionier-boos-im-interview-15619692.html.

[10] Cagala, T. (2017). Improving Data Quality and Closing Data Gaps with Machine Learning. IFC Bulletin, 2017, 46.

[11] Schild, C.-J., Schultz, S. and F. Wieser (2017). Linking Deutsche Bundesbank Company Data using Machine-Learning-Based Classification. Technical Report 2017-01. Deutsche Bundesbank Research Data and Service Centre.

[12] https://www.bundesbank.de/Navigation/EN/Bundesbank/Research/RDSC/rdsc.html

[13] www.ratswd.de.

[14] See Amiti, M. and Weinstein, D. (2017). How Much do Idiosyncratic Bank Shocks Affect Investment? Evidence from Matched Bank-Firm Data. Journal of Political Economy; or Gabaix, X. (2011). The Granular Origins of Aggregate Fluctuations. Econometrica, Econometric Society, Vol. 79(3), pages 733-772, 05.

[15] Fecht, F., Thum, S. and Weber, P. (2018). Fear, Deposit Insurance Schemes, and Deposit Reallocation in the German Banking System. Available at SSRN: https://ssrn.com/abstract=3180107.

[16] Oehler, S. (Forthcoming). Developments in the residential mortgage market in Germany – What can Google data tell us?

[17] Götz, Thomas B., Knetsch, Thomas A. (2017). Google data in bridge equation models for German GDP. Bundesbank Discussion Paper No 18/2017.

[18] Examples are: Bernanke, B., Gertler, M., Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. Handbook of Macroeconomics, Volume 1; Kiyotaki, N. and Moore, J. (1997). Credit Cycles. Journal of Political Economy, Vol. 105, No. 2.

[19] McGowan et al. (2017). The Walking Dead? Zombie Firms and Productivity Performance in OECD Countries. OECD Working Paper No. 1372.

[20] Acharya, V., Eisert, T., Eufinger, C., and Hirsch, C. (2017). Whatever it takes: The Real Effects of Unconventional Monetary Policy. SAFE Working Paper, No. 152.

[21] Schivardi, F., Sette, E., and Tabellini, G. (2017). Credit Misallocation During the European Financial Crisis. Center for Economic Policy Research (CEPR). Discussion Paper 11901.

[22] Storz, M., Koetter, M., Setzer, R., Westphal, A. (2017). Do we want these two to tango? On Zombie firms and stressed banks in Europe. ECB Working Paper Series No 2104.

[23] Mastrogiacomo, M., van der Molen, R. (2015). Dutch mortgages in the DNB loan level data. DNB Occasional Studies, Vol. 13 – 4.

[24] IMF Country Report No. 18/208 Germany. 2018 Article IV Consultation: Press Release, Staff Report, and Statement by the Executive Director for Germany.

[25] For a discussion on the general effects of digitalization on the German labor market see: Weber, E. (2017). Digitalizing the economy: The future of employment and qualification in Germany. IAB-Forum.

[26] Autor, D. (2015), Why are there still so many jobs? The History and Future of Workplace Automation, Journal of Economic perspectives, Vol. 29.

[27] https://www.ecb.europa.eu/stats/ecb_statistics/governance_and_quality_framework/html/merits_costs_procedure.en.html
https://www.ecb.europa.eu/stats/ecb_statistics/governance_and_quality_framework/html/transparency_for_ecb_regulations_on_european_statistics.en.html

[28] https://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/reporting/html/ecb.escb_integrated_reporting_framework201804.en.pdf

[29] http://www.public-test.banks-integrated-reporting-dictionary.eu/

[30] https://www.bundesbank.de/Navigation/DE/Bundesbank/Forschung/FDSZ/INEXDA/inexda.html

[31] https://www.newyorkfed.org/ibrn