*Data & Policy* (2026), 8: e28
doi:10.1017/dap.2026.10073

RESEARCH ARTICLE

# Linking data for better evidence-based policy: A landscape review in central banks

Olivier Sirello[1] and Bruno Tissot[1,2]

[1]Monetary and Economic Department, Bank for International Settlements, Switzerland
[2]Irving Fisher Committee on Central Bank Statistics, Switzerland
**Corresponding author:** Olivier Sirello; Email: olivier.sirello@bis.org

**Received:** 30 September 2025; **Revised:** 06 February 2026; **Accepted:** 03 May 2026

**Keywords:** central banks; data sharing and access; evidence-based policy; official statistics; statistical quality

**Abbreviations:** BIS: Bank for International Settlements; IFC: Irving Fisher Committee on Central Bank Statistics of the BIS.

## Abstract

The growing availability of information sources has offered central banks new opportunities to enhance their statistical, analytical, and policy functions. By linking—or integrating—various data sets, they have been able to produce more granular, timely, and diverse statistics in a cost-efficient way. These advancements have also enabled a better use of information available in society, such as administrative records, to improve statistical agility in responding to user needs. Yet integrating alternative data—often generated as a by-product of other processes—also raises challenges, including concerns over accuracy, representativeness, and reliability. This paper aims to review systematically the opportunities and limitations of data integration in central banks, taking stock of their experience thus far. Results underscore the need for strengthening the global statistical infrastructure through adequate data governance, management, and public resources.

### Policy Significance Statement

In an era of increasing digitalisation, the staggering proliferation of data presents unprecedented opportunities to improve decision-making and analysis. Yet much of these data often remains untapped and fragmented across silos. In this context, linking or integrating data sources holds significant policy relevance, enabling policymakers to maximise the potential of available information while reducing reporting burdens and resource demands. Drawing on the longstanding central banks' experience as both users and producers of data, this paper highlights the numerous benefits associated with data integration and underscores the importance of adopting a holistic approach to data governance and management to make the most of these opportunities.

## 1. Introduction

The growing digitalisation of economies has driven a staggering proliferation of data. This surge has created various opportunities, not least to better inform policymaking by enabling the production of more granular, timely, and diverse data and public statistics. Meanwhile, such an abundance of data also presents significant challenges. In particular, issues of quality arise, as novel data—often generated as a by-product of idiosyncratic, independent processes—may lack the consistency, accuracy, and independence typical of official statistics, which, by contrast, must meet rigorous quality standards and principles

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.

(MacFeely, 2020; Schubert, 2025; UNSD, 2025). There is also a risk of lower-quality data overshadowing and perhaps even crowding out reliable sources, as the latter are often more costly and time-consuming to produce.

Addressing these various trade-offs has raised interest in the concept of data integration. This process, also known as linking or fusion, basically refers to combining multiple data sources. The main goal is to fill information gaps effectively and efficiently, by leveraging the various existing (official and non-official) data while ensuring their highest quality. This calls for having adequate methodological frameworks, information standards, and infrastructure, as well as, perhaps more fundamentally, a systematic and holistic view of the available data being managed and (re-) used.

In recent years, extensive literature has emerged on this topic, particularly in the context of official statistics. Most of this research concentrates on the methodological approaches and techniques to link data in ways that ensure adequate quality and draws on use cases from national statistical offices (NSOs), which are typically the main suppliers—but not the only ones—of official statistics. Interestingly, however, research has been relatively scarcer in the context of central banks, which are also responsible for producing reference statistical information (Dilip et al., 2026). These not only have developed significant experience in linking various sources and types over the past decades, particularly with the spread of big data; more significantly, data integration appears to be highly relevant to them as both data users and producers to eventually support their statistical, analytical, and policy functions.

This paper addresses this gap by conducting a systematic review of the benefits and challenges of data integration in central banks. First, it conceptualises data integration within the broader literature related to official statistics (Section 2). It then presents a systematic review of the opportunities (Section 3) as well as the various practical, methodological, and organisational challenges of linking data (Section 4). Finally, Section 5 outlines a number of recommendations based on the central banking experience to make the most of the available information, underscoring the need for an efficient use of the data sources at hand.

## 2. Conceptualising data integration in central banks: related literature and analytical framework

Data integration refers to the process of combining data and spans three key dimensions (Figure 1). First, it encompasses merging diverse sources, typically distinguished as "primary" (i.e., data collected for statistical purposes in line with international statistical standards, such as the UN National Quality Assessment Framework—NQAF) and "secondary" (i.e., data that have not been *primarily* collected for statistical purposes and can be a by-product of another activity, such as administrative, regulatory and commercial data; UNECE, 2024a). Second, it may involve linking data sourced from multiple areas, such

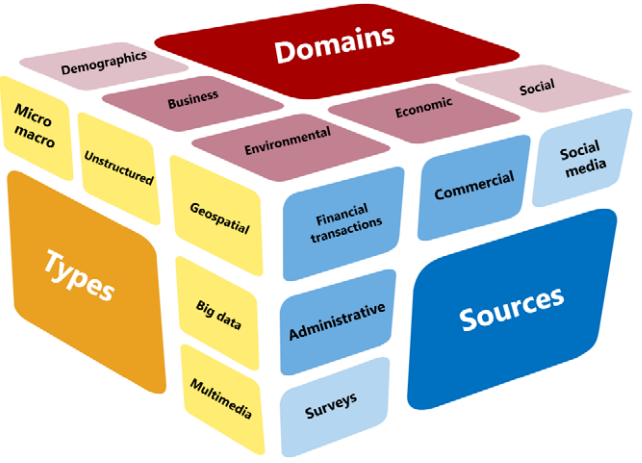



***Figure 1.*** The three key dimensions of data: domains, types, and sources. Source*: Authors' elaboration.*

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.

as social, economic, and environmental statistics. A prominent case is the production of national accounts, which relies on a mix of various statistical domains. Third, data integration may combine different information types, such as "traditional" statistical time series with novel data types (IFC, 2024a). As an example, the integration of geospatial information with textual and sensor data can help produce experimental statistics on the environment (OECD, 2018). Another typical case is the reconciliation of micro-level data with aggregate statistics, such as security-by-security databases in the balance of payments (BoP) (Dilip and Tissot, 2024).

A substantial body of literature has emerged on the subject of data integration over the past years, particularly related to official statistics. It can be categorised in three main strands focusing on (i) linking multiple data types and sources, (ii) integration techniques and standards, and (iii) applications in official statistics, including among NSOs and central banks.

The first strand of literature focuses on the integration of non-traditional data types. For one, the widespread adoption of big data over the past years has garnered significant attention for its potential to drive data-driven innovations across various sectors (Mondal et al., 2023). Research has, in particular, explored the applications of big data in finance, economics, and statistics, including in terms of real-time analytics and more granular indicators (Varian, 2014; Israël and Tissot, 2021; Tosi et al., 2024). A number of studies have highlighted challenges associated with using big data for statistical purposes, particularly regarding data quality, scalability, privacy, integration, and ethical considerations (Struijs et al., 2014; Kitchin, 2015; Radermacher, 2018). Secondly, attention has been devoted to leveraging administrative data to complement traditional statistical sources, particularly for censuses (Wallgren and Wallgren, 2014; Aragona and Zindato, 2016; Harron et al., 2017). Last but not least, the integration of geospatial information (GIS) into official statistics has also gained traction, highlighting the importance of ensuring compatibility and consistency across diverse data sets, often through robust tools and processes (Hassani et al., 2025).

The second strand delves into the techniques and standards underpinning data integration. Building on seminal work on probabilistic record linkage (Fellegi and Sunter, 1969), statistical and data matching (D'Orazio et al., 2006; Christen, 2012), machine learning and artificial intelligence (AI) techniques have recently gained prominence for automating entity resolution, handling complex relationships, and using ontologies to achieve semantic integration of heterogeneous data (Leulescu and Agafitei, 2013; Lewaa et al., 2021; Nunna and Sahu, 2025). In parallel, there has been a growing focus on data architectures for integration processes, both implying technical (Doan et al., 2012; Bogdanov et al., 2020; Haryono et al., 2020) and management considerations (Dehghani, 2022; Östberg et al., 2022). Another important area of attention is privacy-enhancing technologies, especially to enable the linking of granular data while safeguarding confidentiality rules (Ranbaduge et al., 2024; Ricciato, 2024). Lastly, a significant portion of this literature discusses how standardisation and, broadly, interoperability can support data integration (Gregory, 2024; United Nations Economic Commission for Europe (UNECE) 2025a). Specifically, related works show that statistical standards—such as Statistical Data and Metadata eXchange (SDMX)—can support semantic alignment to ensure that the data to be integrated are clearly and unambiguously interpreted (IFC, 2023a, 2025a).

The third strand concentrates on the applications of data integration in organisations involved in the production of official statistics. Much of this research focuses on NSOs, typically discussing the opportunities and limitations of linking various sources and types to produce statistical indicators (Ascari et al., 2020; Allin, 2021; Holmberg, 2025). An area of ongoing attention has been the integration of non-traditional data, such as administrative records, to enhance the quality of traditional statistical surveys and address the issues posed by non-responses. Central banks have also emerged as a key area of focus because of their important role in compiling statistics for policy and the public good at large (Dilip et al., 2026). Research has, particularly, examined the use of financial big data in central banking (Tissot, 2017; Doerr et al., 2021; Cornelli et al., 2022) as well as the integration of supervisory and financial information to support central banks' various functions (Tissot, 2019; Turner, 2021), especially during periods of traditional data supply disruption or "statistical darkness" (de Beer and Tissot, 2020).

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.

Despite this growing body of research, comprehensive and systematic reviews on the benefits and challenges of data integration in the context of central banks remain scarce. Yet these represent a relevant case study for several reasons, not least because of their extensive experience in combining various information sources and types for fulfilling their statistical, analytical, and policy functions effectively. To address this gap, this paper conducts a systematic review harvesting a vast body of primary and secondary sources, including public reports, working papers, and other presentations. Primary sources typically relate to official publications by central banks as well as by the Irving Fisher Committee on Central Bank Statistics (IFC) of the Bank for International Settlements (BIS), which has accumulated over time a substantial volume of documentation on data and statistical issues related to central banks. Secondary sources encompass supplementary materials, including research papers, reports, and analyses that provide additional context on the topics covered in the primary sources. Turning to methodological considerations, inductive thematic analysis is used to extract common patterns in the benefits and challenges identified across the reviewed materials. This technique is especially well-suited for qualitative studies as it provides a flexible yet systematic approach to identifying and interpreting themes (Guest et al., 2012). However, as with most qualitative methods, it may have certain limitations, such as the reliance on expert judgment, which inherently involves a controlled though unavoidable degree of subjective interpretation (Nowell et al., 2017). To address this, the paper ensures robustness by transparently referencing primary and secondary sources used, as well as by tapping into a broad range of literature, which strengthens the reliability and generalisability of the findings.

## 3. A review of the main opportunities and challenges associated with data integration in central banks

This section presents the results of a review of the opportunities offered by data integration in central banks (Section 3.1)—including enhanced insights, improved data use, and accuracy—alongside the related challenges (Section 3.2), such as fragmented standards, resource constraints, data quality issues, organisational silos, and ethical concerns.

### *3.1. Benefits and opportunities*

Common feedback from central banks shows that opportunities include the ability to derive additional analytical insights and statistical agility (Section 3.1.1), make better use of existing data (Section 3.1.2), and achieve enhanced data accuracy (Section 3.1.3).

#### *3.1.1. Additional analytical insights and statistical agility*

Combining various sources can facilitate addressing the wide range of user demands for multidimensional information in central banking. Climate change statistics are a prime example of the benefits offered by data integration, both in terms of aggregation levels—that is, micro and macro—and the types—that is, textual documents (e.g., sustainability reports, press releases), images (e.g., satellite imagery, street view), and sensor data (Aurouet et al., 2023, Doll et al., 2026). Data integration can also help meet policy demands in the areas of financial stability and supervision. For example, detailed assessments of banks' credit exposures often require linking data on loan activity with business registers and financial statements to identify counterparties (Elgg et al., 2025). Furthermore, combining various data sources can also be useful for central banks' economic monitoring and forecasting tasks, such as measuring policy uncertainty through text analysis (Ghirelli et al., 2021).

Linking sources also enables additional and more diverse analytical perspectives. In external statistics, combining business registers and BoP data can help map global financial risks (Diz Dias et al., 2025) as well as complement traditional residency-based statistics with those that are nationality-based (McGuire et al., 2024). Linking BoP statistics with mirror data has also proven useful to better track international finance (Pradhan and Silva, 2019). Beyond external statistics, integrating data from payments and financial statements can unlock a wide array of new types of analysis, such as estimating consumption

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.

through credit card transactions and tourism statistics, monitoring the financial cycle, estimating loan-to-value ratios, and, more broadly, tracking economic conditions through mobility indices.

Lastly, integrating micro and macro sources can enhance statistical agility, with benefits for both data users and compilers (Rosolia et al., 2021). For users, it allows answering questions with greater agility, for instance, by "zooming in" on specific events without losing sight of macro or system-wide perspectives (Israël and Tissot, 2021). For producers, combining micro and macro data can allow the timely compilation of new aggregates by swiftly rearranging granular inputs (e.g., "Lego bricks") instead of setting up new data collections that are typically slower to deploy.

#### *3.1.2. Better use of existing data*

Effective data integration allows for maximising the use of existing data sources with three main benefits, namely, (i) limiting reporting burden, (ii) filling information gaps—especially in times of statistical disruptions—and (iii) improving timeliness.

As regards reporting burden, data integration can help identify and potentially streamline redundant data collections. This is because similar information may be collected multiple times in practice due to the absence of data inventories, inadequate sharing, or governance limitations. A typical case relates to statistical and supervisory data exercises, which often collect identical information from the same reporting agents. To address this issue, the ECB's Integrated Reporting Framework (IReF) project aims to integrate the Eurosystem's statistical requirements for banks and, ultimately, statistical and prudential reporting.

Another advantage of data integration is to close information gaps in a cost-effective way. For users, their data requests can be more easily met without setting up new data collections. Turning to producers, integrating various data from administrative, credit, and business registers can enhance statistical coverage, potentially moving away from probability samples as the main basis for official statistics (Fosen et al., 2025). A case in point is the Bank of Korea's use of administrative data to complement the 2020 economic census to enhance the coverage of e-commerce and freelancers. Relatedly, Eurostat has developed a micro-data-linking project to leverage data from business registers—notably the EuroGroups Register—to derive additional breakdowns of structural business statistics. Further, in the area of external statistics, linking BoP aggregates with security-by-security holdings data can provide additional information on the hidden wealth of households and firms.

The benefits of data integration to fill information gaps can be particularly critical to cope with sudden stops in data collection and overcome periods of "statistical darkness." The COVID-19 pandemic was a prime example, also serving as a wake-up call for official statistics to complement traditional sources with alternative data as a way to maintain statistical production during periods of distress (de Beer and Tissot, 2020). This episode underscored, in particular, the importance of building "rainy day data funds" or "statistical buffers" in case of shortages in the information supply chain, while also allowing for the flexible measurement of phenomena that traditional methods cannot adequately capture in times of crisis (Veronese et al., 2020).

Third, combining sources can also improve timeliness, in particular to address urgent information needs. For instance, leveraging online indicators offers numerous opportunities thanks to their immediate accessibility and availability (Cavallo, 2015). A well-known example relates to the measurement of consumer price indices that can benefit from web-scraped data to nowcast inflation, produce higher frequency indicators, and/or complement official statistics in those jurisdictions characterised by data collection difficulties. Examples of such advancements also include the compilation of consumer expenditure indicators leveraging scanner data or the analysis of market sentiment by integrating textual data with more traditional indicators (Araujo et al., 2025).

#### *3.1.3. Enhanced data accuracy*

Combining information from various sources can be instrumental in supporting quality checks for data collected by a statistical compiler like a central bank. In particular, it can assist with consistency

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.

assessments, for example, by benchmarking the data against other, well-identified sources to ensure that they provide a coherent description of the economic indicator being measured. Relatedly, plausibility checks can also benefit from having multiple sources, especially for detecting outliers or incorrect patterns. Finally, combining information can also facilitate mirror analysis by allowing the comparison of the same phenomena with independently collected observations (Jellema et al., 2020).

Additionally, leveraging secondary data as auxiliary sources also has the potential to support overall statistical quality through adequate assurance frameworks and processes. They can assist in a broad range of editing tasks, such as locating missing data or filling gaps due to delayed reporting. This can be especially useful to cope with growing non-response rates for surveys (Scotti et al., 2024). Further, they can support post-survey adjustments, such as imputation techniques to estimate missing values (Sakshaug and Steorts, 2023) or estimating the valuation of unlisted shares in financial accounts through financial statements data (Paúl, 2024). Finally, mixing data from different sources may play a decisive role in protecting sensitive information, for instance, by enhancing statistical disclosure techniques (UNECE, 2023).

Furthermore, the concept of data integration is not solely an in-house issue. It can indeed be extended to the broader perimeter of the data ecosystem, allowing, for instance, central banks to benchmark their data against that produced by NSOs and international organisations. An example is the joint work by the Bank of France and the French National Institute of Statistics and Economic Studies to ensure that their coverage of economic activities is consistent and, in turn, more accurate (Barut-Etherington and Golfier-Chataignault, 2024). Similarly, the Bank of Spain and the Spanish National Statistics Institute collaborate closely to check and improve the quality of data on foreign-controlled corporates by combining multiple sources, such as business registers, the central bank balance sheet office, foreign affiliated trade statistics, national accounts, and BoP (García del Riego and Paúl, 2024).

### *3.2. Challenges and limitations*

Combining multiple data sources and types does not come without challenges and costs, according to central banks' experience. These typically include fragmented information standards and identifiers (3.2.1), IT, security, and resource constraints (3.2.2), data quality (3.2.3), statistical continuity (3.2.4), organisational siloed structures (3.2.5), and ethical and legal concerns (3.2.6).

#### *3.2.1. Fragmented information standards and identifiers*

One key issue limiting effective data integration is the absence of consistent information standards covering the various types of data and statistical phases. For example, SDMX is often used for macroeconomic statistics (IFC, 2025a), while other standards are considered for regulatory data (i.e., eXtensible Business Reporting Language, or XBRL), cross-border payments (ISO 20022), metadata and surveys (Data Documentation Initiative, or DDI), and geospatial information (ISO 19111 and 19115). While each of these standards is tailored to specific needs, the presence of several standards may also increase fragmentation, costs, and maintenance difficulties for data compilers. Ultimately, this complexity can hinder efficient data use Gregory et al. (2024).

Another challenge is the limited coverage of common identifiers. These are critical to join different sets of information, for instance, through record linkage. Over the past few decades, there have been a number of international initiatives to make progress in this regard, such as the Global Initiative on Unique Identifiers for Businesses (GIUIB), the Legal Entity Identifier (LEI), and the International Securities Identification Number (ISIN) for traded securities. Yet, in practice, global identifiers still have many limitations. First, they typically rely on voluntary adoption. For example, the LEI is not mandatory across all jurisdictions and includes only firms accessing financial markets, so that it may poorly represent countries with limited financial services (Pilgrim and Ang, 2024). Second, global identifiers often compete with existing regional or national ones, which underpin most business registers and administrative data sets but may not be fully consistent. Finally, most private data brokers, such as Bloomberg

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.

Ticker and Refinitiv Instrument Code (RIC), have their own identifiers, making data linking more complicated.

*3.2.2. IT, security, and resources issues*

Linking data can raise various IT challenges. Key ones include investing in adequate solutions to handle diverse and complex data formats as well as to manage, store, and process large data sets (Berman, 2018; Doerr et al., 2021). This raises the issue of how to migrate from legacy systems—typically designed to handle macro and structured data—to more agile, scalable, and flexible solutions that can offer higher-performance computing capabilities (IFC, 2020). Yet, migrations may also lead to a heterogeneous IT environment, with the (costly) need to maintain dual-system operations at least on a temporary basis and increased security risks (IFC, 2023a).

One associated challenge relates to the option of adopting cloud services, but this poses difficult trade-offs. On the one hand, the increased interest in the cloud by central banks reflects the need to find cost-effective solutions for storing and handling diverse sources like web-scraped text, scanner data, and mobile phone records, which often require high scalability, computational power and, increasingly, sophisticated tools that are not available on premises (IFC, 2025b). On the other hand, cloud adoption also raises important risks, particularly in terms of potential data leaks and violations of sovereignty, as well as concerns about excessive (and often quite expensive) third-party dependency (UNECE, 2024b).

A second major issue is ensuring adequate IT security, particularly to protect restricted and/or micro-level information. This typically involves developing and implementing solutions to prevent data breaches by hardening cybersecurity measures, which can, however, be costly to maintain. The fast-paced development of quantum computing may also threaten existing cryptographic techniques and, more broadly, current practices to safeguard data confidentiality and authentication (Auer et al., 2025).

A third challenge is that the implementation of IT solutions may face a shortage of in-house skills, calling for recruiting or training qualified staff to keep up with the pace of technological developments. It also demands developing new, "hybrid" professional roles that have traditionally been rare in central banks and statistical offices, such as data engineers and data scientists (Antonucci et al., 2023). In low- and middle-income countries, the lack of staff resources, combined with intense global competition for IT skills, may critically hinder the modernisation of their IT infrastructure (Hammer et al., 2017).

*3.2.3. Keeping integrated data fit for purpose and consistent*

One key methodological concern with secondary sources is whether they are fit for the intended statistical purpose, which is a golden rule when producing high-quality official statistics (UN, 2014). Since alternative sources are typically generated as a by-product of other processes, there is no guarantee that their collection is aligned with existing methodological frameworks and best practices (UNECE, 2021a). Moreover, a significant part of these indicators is made available by private providers with limited transparency and metadata information. The absence of clear documentation can create ambiguity regarding how to integrate the data, potentially leading to inconsistent outcomes, even when identical analyses are performed on the same original data sources.

The lack of robust methodologies underpinning most alternative sources might exacerbate a number of statistical biases. A prominent one is related to big data samples, whose representativeness can be limited due to the biased coverage of most non-traditional sources (Tissot, 2019; Cipollone, 2022). Another one is hallucination, with the possibility of compounding errors by integrating inaccurate data sources. This clearly raises the risk of polluting good data with bad ones by incorporating inaccurate information into well-established statistical processes (the "garbage in—garbage out" or "data contamination" syndrome).

One particular issue also relates to the incorporation of micro-level information into existing statistical frameworks that often rely on macroeconomic data. This integration can result in major discrepancies, not least because of inconsistencies in terms of sector classification. Further, micro data are generally organised around legal units, such as consolidated global groups spanning across several jurisdictions, rather than institutional units identified in specific sectors of the domestic economy (Tissot, 2016).

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.

Dealing with these caveats may call for developing and implementing adequate statistical methods and practices. These include acknowledging data limitations in a transparent way (Gelman and Henning, 2017), providing adequate methodological guidance, and communicating clearly how experimental indicators are manufactured, especially when these do not fit established statistical standards, to partially address users' immediate needs.

#### *3.2.4. Statistical continuity*

Combining multiple data sources could increase the likelihood of discontinuities, in turn undermining the overall quality of the statistics produced and generating breaks in final outputs, which can be hard to explain. One example of such limitations is that big data sets are often highly volatile and can create inexplicable jumps in time series (Braaksma and Zeelenberg, 2015). Another is the typically limited historical depth of a number of secondary sources, which might prevent cross-time consistency and, in turn, the compilation of long series (Woloszko, 2024). A final issue relates to the risks posed by the discontinuation of the business processes generating alternative indicators, for instance, when a social media or mobile phone company stops its activities.

One possible response to this challenge is to continuously evaluate the stability of the sources being integrated by engaging closely with all the stakeholders involved. Having formal data-sharing agreements with explicit clauses on data provision continuity could be another mitigation factor. In any case, it may be essential to work on contingency plans, limit overreliance on certain providers, and promote the diversification of sources (Ascari et al., 2020).

#### *3.2.5. Siloed structures*

Compartmentalised organisational structures may hinder effective linking of data, as the information tends to be guarded by its owners in "data silos." While this can reflect important considerations (e.g., confidentiality protection or institutional arrangements), it also prevents any holistic approach to the information available within the institution.

In practice, the presence of silos can generate a number of challenges for data integration projects as well as overall data reusability. First, fragmented data collections may lead to inconsistent formats and sources across reported data, complicating the integration process, especially when there is scarce semantic and system interoperability. Second, silos can result in inefficiencies, such as with redundant data stored in multiple locations, leading to poor scalability and high storage costs. And third, silos may undermine collaboration between data users and producers, causing potentially costly errors, flawed analytical outcomes, and users' redundant data requests.

#### *3.2.6. Ethical and legal issues: balancing utility with social acceptability and independence*

Linking sources can raise privacy, confidentiality, and ethical concerns. For one, integration could lead to centralising multiple information sources, resulting in negative public perceptions should individual data not be adequately safeguarded, inappropriately used and/or handled without consent (Sexton et al., 2018). A potential way to resolve this issue is to remove the identifiability of individual records, but this might also significantly decrease the usefulness of the data, in turn undermining further integration—for instance, through record linkage methods.

These considerations call for establishing robust and clear rules underpinning data integration exercises, not least to protect sensitive information. At the international level, the Principles governing international statistical activities by the Committee for the Coordination of Statistical Activities (CCSA) state that "individual data […] are to be kept strictly confidential and used exclusively for statistical purposes or for purposes mandated by legislation." Another option is developing legal arrangements to facilitate access to administrative data while ensuring the protection of privacy rights.

In addition, it is important to take adequate operational measures to protect data confidentiality at the operational level. A number of central banks have put in place information security management guidelines and notices, for example, to prevent breaches through stringent protocols for access and

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.

control (IFC, 2023b). Further, they have established data centres or labs, particularly for researchers, to enable the integration of various granular data sets in a secure and protected environment (Brault et al., 2024). Privacy-enhancing technologies such as aggregation, synthetic data generation, or anonymisation have also been developed to enhance data protection while also allowing access to a wider range of users (UNECE, 2021b; Araujo et al., 2025).

Finally, the use of multiple sources—notably administrative and private ones—for statistical purposes has to be carefully weighed against the risk of excessive reliance on external providers, which could impact—directly or indirectly—the professional independence of compilers (Ljones, 2011). Strengthening the adherence to the Fundamental Principles of Official Statistics and safeguarding the independent role of central banks' entities responsible for producing official statistics is crucial in this regard.

## 4. Maximising data use through multisource statistics: key lessons and ways forward

The above review of central banks' experience suggests that maximising data use and value through integration while mitigating the associated challenges calls for making further progress in four areas: data governance (Section 4.1) and management (Section 4.2), adequate data access and sharing (Section 4.3), data quality and curation (Section 4.4), international cooperation (Section 4.5), as well as interdisciplinary skills, communication, and literacy (Section 4.6).

### *4.1. Data governance*

Data governance is a multifaceted concept, which can be referred to as "a system of decision rights and accountabilities for the management of the availability, usability, integrity, and security of the data and information to enable coherent implementation and coordination of data stewardship activities as well as increase the capacity to better control the data value chain" (UNSD, 2025). More generally, governance aims to secure the overall quality of statistical information. Sound data governance can play an important role in enabling multisource integration, for at least five reasons.

First, it helps define comprehensive assessments, identifying the various roles, responsibilities, and accountabilities involved when combining data sources both at the organisational and broader ecosystem levels (MacFeely et al., 2025a). These include clear rules on who controls the assets, who can access them, and how they can use and reuse them.

Second, it also underpins the setup of adequate structures to facilitate coordination of data activities and eventually break down organisational silos (Moreno, 2021). One example is the growing relevance of data stewards in addition to the role of data owners to facilitate the use and reuse of data across the organisation. In addition, several central banks have taken actions to foster a common bank-wide coordination function for consolidating and streamlining data activities across the various stakeholders involved.

Third, a data governance framework can help coordinate data management processes, in particular to prevent redundancies in collections, licences, and associated costs, as well as inadequate storage solutions and limited access.

Fourth, its implementation requires developing clear strategies for data (re)use. This calls for recognising the importance of data as a strategic asset to generate high-quality information. More broadly, it also underscores the fully fledged role of data resources in modern societies as a public good, not least because of their fundamental implications in terms of representation, equity, and trust in public policies. As such, high-quality data should foster transparency in decision-making and promote inclusive access to public information (MacFeely et al., 2025b).

Fifth, a data governance framework helps to properly and actively manage data throughout their entire lifecycle (the so-called "data curation" process), ensuring their preservation, quality, reliability, integrity, and access (Križman and Tissot, 2022). This should ultimately strengthen the role of reliable statistics as a key public good infrastructure in an increasingly global competitive information environment (UN-CEB, 2023).

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.

### *4.2. Data management*

Integrating various sources to compile official statistics also calls for sound data management to execute and supervise the organisation, maintenance, and improvement of the necessary IT infrastructure. One first action point is to develop and document data catalogues as a unique inventory of the information assets available in the organisation. Such catalogues are key instruments helping users identify, search, and make use of existing data sets. Another focus point is to implement well-recognised standards to ensure that data can be properly found, exchanged, accurately interpreted, and reused, for instance by fostering semantic interoperability. A case in point, SDMX has been widely adopted by statistical organisations, including central banks, to support the various stages of the data life cycle, including harmonisation of metadata through common registries (IFC, 2025a).

Additionally, developments in AI are also opening new avenues for data management, especially for data discovery applications. Search engines are being improved to capture users' inputs through semantics rather than predefined keywords. New AI-based tools can also enhance the overall quality and accuracy of classifications, which is essential for ensuring the reliability of data catalogues. Taken together, these developments can foster better data integration processes.

### *4.3. Data access and sharing*

Fostering appropriate access to and sharing of information is essential to facilitate data reuse and make the most of the benefits of data integration.

As regards data access, central banks and other statistical organisations have already developed various solutions de Carvalho et al (2024). These notably include easily accessible data portals to foster the findability and discoverability of the disseminated statistics (IFC, 2024c); common data platforms that pool data together across multiple institutions; and, finally, single access data points to foster the reuse of information by making it available to the public through centralised platforms (Araujo et al., 2025).

Central banks have been actively promoting organisation-wide data sharing (e.g., between statistical and supervisory functions), across the national statistical system (e.g., between the central bank and the NSO) as well as internationally (e.g., the GSIBSs data collection undertaken under the BIS International Data Hub; Bese Goksu and Tissot, 2018). Specific projects include the development of open data interfaces, data libraries, and data spaces which support data sharing in secure and privacy-preserving environments, as well as the creation of data centres serving as entry points for all data assets in the organisation, including micro data (Bender et al., 2019). Statistical offices and central banks are also leveraging innovative data techniques, such as privacy-enhancing technologies to promote better access to and sharing of information while also safeguarding confidentiality (Kim et al., 2025).

Moreover, central banks have been engaging with other statistical organisations to promote the exchange of best practices and recommendations on data access and sharing at the international level. An example is recommendations 13 and 14 of the third phase of the G20 Data Gaps Initiative, which aim to establish internationally agreed principles and taxonomies on data sharing and access, particularly to guide the collaboration of statistical organisations with private companies. Nevertheless, there are still important and understandable operational, legal, and ethical factors constraining the scope of data sharing.

### *4.4. Quality assurance and curation*

To be successful, data integration exercises need to incorporate the various data sources (especially secondary ones) into well-established statistical and data quality frameworks through rigorous assurance and curation processes. One key issue is to address the arguably lower quality of emerging alternative data to be used for statistical purposes (Viggo Sæbø and Hoel, 2023). Additionally, the ongoing adoption of AI is underscoring the importance of a wide range of data quality aspects to be fostered, including availability, openness, and machine readability, particularly of metadata (MacFeely, 2025).

Addressing these quality issues calls for focussing efforts in three areas. The first one is documenting how non-official sources can be used for statistical purposes, for instance, by making recommendations or

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.

publishing dedicated handbooks based on current experiences—such as the one currently pursued by the UNECE High-Level Group for the Modernisation of Official Statistics. Central banks can play an important role in this endeavour, not least because their various functions naturally call for integrating data across multiple domains. A second area is advancing metadata quality frameworks, particularly to better respond to the emerging needs for AI-ready data (e.g., the "FAIR-R" principles; Verhulst et al., 2025). A third objective is to further promote the development and use of data maturity frameworks that can help better identify the opportunities and challenges posed by data integration. Such frameworks provide organisations with a high-level benchmark to review their various data processes against different levels of sophistication (i.e., the maturity degrees). One recent example is the maturity framework developed by the DGI to assess institution-wide capabilities in accessing and sharing data in official statistics.

### *4.5. Collaboration and cooperation*

Integrating the large variety of available data sources calls for strengthened collaboration with all the stakeholders involved.

Fortunately, central banks can leverage their well-established partnerships within the statistical system, notably with NSOs, other government bodies, and international organisations. Such collaborations can broaden the palette of available sources, particularly administrative data, in turn helping to measure complex phenomena. The example of large cases units (LCUs)—that is, specialised units collecting data on multinational enterprises—showcases how cooperation between NSOs, central banks and other government institutions can lead to a better "profiling" of international firms and understanding of the global value chains. Moreover, collaboration between central banks and other statistical actors can be instrumental in optimising data collection. An example is the partnership between the NSO, the Bank of England, and the Financial Conduct Authority in the United Kingdom, which has helped to maximise the use of existing data reported from financial institutions (Benford, 2024a).

Collaboration may also expand to other important stakeholders in the data ecosystem, especially the private sector, academia, and citizens. The starting point is to recognise that companies are generating immense amounts of data, which can be used for statistical purposes (UNECE, 2025b). For example, information from mobile network firms can be leveraged to produce external statistics (IFC, 2024b). Regarding academia, collaborative alliances with statistical authorities can offer various advantages to promote a better use of novel data and leverage experimental techniques, while also fostering statistical literacy (Hansen et al., 2024). Turning to the general public, its involvement ("citizen science") can contribute to statistical work, for instance by using social media information in the sustainable development area (Fritz et al., 2019). Similarly, citizens' participation in public data collections through crowdsourcing may enhance official statistics, for instance, on payments (by leveraging digital transactions) and migration (i.e., geo records from smartphones).

While beneficial, collaboration with private stakeholders may raise some challenges. One question is how to tap into corporate information while also guaranteeing an adequate level of accessibility, equity, and inclusiveness, particularly for countries with limited statistical capacity (Amutorine et al., 2024). Another challenge involves crowdsourcing efforts that require providing incentives to encourage citizen participation in data collections while safeguarding their privacy. Finally, statistics generated by the private sector may fall short of the rigorous quality standards of official statistics, emphasising the need for users to enhance statistical literacy to better identify limitations.

### *4.6. Interdisciplinary skills, communication, and literacy*

Maximising the use of various data sources obviously calls for fostering cross-functional data expertise for both users and producers, especially to achieve synergies across data domains and, at an operational scale, across units (Nelson et al., 2024). For example, the collection of loan-by-loan information can serve the needs of both monetary policy and macroprudential supervision. But, to be effective, it may also

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.

require compilers to combine knowledge of monetary and financial statistics with prudential as well as accounting expertise. Furthermore, users of this information also have to understand the various methodologies involved and develop adequate skills to perform data analysis. This puts a premium on developing adequate communication strategies to better explain the statistics produced, foster greater interpretability, and, in turn, reduce the risk of their misuse (de Carvalho Campos et al., 2024).

Hence, a key lesson moving forward is that the integration of the constantly evolving data landscape requires a generalist approach to effectively handle multiple sources and techniques. The emergence of hybrid profiles, such as data scientists—at the intersection of IT specialists, statisticians, and subject matter experts (Araujo et al., 2023)—is perhaps one solution to this issue (Antonucci et al., 2023). More generally, central banks as data-dependent organisations are likely to be increasingly interested in cultivating flexible and interdisciplinary skills among staff, for instance by investing in a broad set of learning resources (Damouras et al., 2021).

## 5. Conclusion

In conclusion, the review of central banks' experience with data integration shows that this process offers significant benefits, including enhanced analytical insights, better use of existing data, and improved data accuracy. It enables central banks and statistical organisations to meet complex user demands, fill information gaps, and ensure statistical agility in an evolving and uncertain policy landscape. However, the integration process is not without challenges, including still-fragmented standards, IT infrastructure limitations, ethical concerns, and organisational silos.

Addressing these challenges requires robust data governance, sound data management, improved access and sharing mechanisms, quality assurance frameworks, and, perhaps more importantly, strengthened collaboration with stakeholders. Furthermore, fostering interdisciplinary skills and statistical literacy is instrumental to maximising the potential of integrated data sources. By embracing these strategies, central banks can unlock the full value of multisource statistics, ensuring they remain relevant and effective in supporting policy decision-making.

**Data availability statement.** Data availability is not applicable to this article as no new data were created or analysed in this study.

**Acknowledgments.** The authors thank Douglas Araujo, Mauro Bucci, InKyung Choi, Archana Dilip, Eurydice Fotopoulou, Zlatina Hofmeister, Edward Lambe, François Robin, Luís Teles Dias, Kevin Tracol, Gabor Vincze, colleagues at Banco de Portugal and the European Central Bank, as well as four anonymous referees for their valuable comments and suggestions. The views expressed are those of the authors and do not necessarily reflect the views of the BIS, the IFC, and the various organisations mentioned in this guidance note. All errors are the authors' sole responsibility.

**Author contribution.** Conceptualization, Formal analysis, Investigation, Writing - original draft, Writing - review & editing: O.S; B.T.

**Funding statement.** None.

**Competing interests.** A statement about any financial, professional, contractual or personal relationships or situations that could be perceived to impact the presentation of the work or `None' if none exist.

## References

**Allin P** (2021) Opportunities and challenges for official statistics in a digital society. *Contemporary Social Science 16*(2), 156–169. https://doi.org/10.1080/21582041.2019.1687931.

**Amutorine M**, **Lawrence N and Montgomery J** (2024) Increasing data sharing and use for social good: Lessons from Africa's data-sharing practices during the Covid-19 response. *Data and Policy 6*, e53. https://doi.org/10.1017/dap.2024.43.

**Antonucci L**, **Balzanella A**, **Bruno E**, **Crocetta C**, **Di Zio S**, **Fontanella L**, **Sanarico M**, **Scarpa B**, **Verde R and Vittadini G** (2023) Data science skills for the next generation of statisticians. *Statistical Journal of the IAOS 39*(4), 773–782. https://doi.org/10.3233/SJI-230060.

**Aragona B and Zindato D** (2016) Counting people in the data revolution era: Challenges and opportunities for population censuses. *International Review of Sociology 26*(3), 367–385. https://doi.org/10.1080/03906701.2016.1244926.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.

**Araujo D**, **Bruno G**, **Marcucci J**, **Schmidt R and Tissot B** (2023) Data science in central banking: Applications and tools. *IFC Bulletin 59*, 1–28.

**Araujo D**, **Cap A**, **Mattei I**, **Schmidt R**, **Sirello O and Tissot B** (2025) Data science in central banks: Unlocking the potential of data. *IFC Bulletin 64*, 1–22.

**Ascari G**, **Blix K**, **Brancato G**, **Burg T**, **McCourt A**, **Delden AV**, **Krapavickaitė D**, **Ploug N**, **Scholtus S**, **Stoltze P**, **Waal TD**, **Zhang L-C** (2020) Quality of multisource statistics – The KOMUSO project. *The Survey Statistician 81*, 36–51.

**Auer R**, **Dodson D**, **Dupont A**, **Haghighi M**, **Margaine N**, **Marsden D**, **McCarthy S and Valko A** (2025) Quantum-readiness for the financial system: A road map. *BIS Papers 158*. https://doi.org/10.21314/JFMI.2025.006.

**Aurouet D**, **De Sanctis A**, **Giudice D Del**, **Franke J**, **Herzberg J**, **Osiewicz M**, **Peronaci R and Willeke C** (2023) Indicators of granular exposures to climate-related physical risks for central banks' analytical purposes. In Arslanalp S, Kostial K and Quirós-Romero G (eds.), *Data for a Greener World: A Guide for Practitioners and Policymakers*. Chapter 4: IMF, pp. 59–77.

**Barut-Etherington M-L and Golfier-Chataignault C** (2024) Three good reasons for central banks to (also) produce statistics. *Bank of France Bulletin 250*, 4.

**Bender S**, **Hausstein B and Hirsch C** (2019): "An introduction to INEXDA's metadata schema", *IFC Bulletin 49*, January.

**Benford, J** (2024a): "Enhancing our national asset: maximising the value of data collections across the financial sector", speech at the Regtech Summit, London, 3 October.

**Berman J** (2018) "Providing structure to unstructured data", In *Principles and Practice of Big Data*, 2nd ed., pp. 15–51.

**Bese Goksu E and Tissot B** (2018) Monitoring systemic institutions for the analysis of micro-macro linkages and network effects. *Journal of Mathematical and Statistical Science 4*(4), JMSS17082201.

**Bogdanov A**, **Degtyarev A**, **Shchegoleva N**, **Khvatov V and Korkohv V** (2020) Evolving principles of big data virtualization. *Computational Science and its Applications – ICCSA 12254*, 67–81.

**Braaksma B and Zeelenberg K** (2015) Re-make/re-model: Should big data change the modelling paradigm in official statistics? *Statistical Journal of the IAOS 31*(2), 193–202. https://doi.org/10.3233/sji-150892.

**Cavallo A** (2015) Scraped data and sticky prices. *NBER Working Paper Series 21490*, 30.

**Christen P** (2012) *The Data Matching Process" in Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-31164-2.

**Cipollone P** (2022) "The experience of 10 years of data in central banking – from gathering real-time data and big data to challenges like storage or skills", speech at the international conference "Future of central banking", Vilnius, 29 September.

**Cornelli G**, **Doerr S**, **Gambacorta L and Tissot B** (2022) Big data in Asian central banks. *Asian Economic Policy Review 17*(2), 255–269. https://doi.org/10.1111/aepr.12376.

**D'Orazio M**, **Di Zio M and Scanu M** (2006) *Statistical Matching: Theory and Practice*. Wiley. https://doi.org/10.1002/0470023554.

**Damouras S**, **Gibbs A and MacFeely S** (2021) Training official statisticians for adaptive statistical practice. *Statistical Journal of the IAOS 37*(3), 887–898. https://doi.org/10.3233/SJI-210851.

**de Beer B and Tissot B** (2020) Implications of Covid-19 for official statistics: A central banking perspective. *IFC Working Papers 20*.

**de Carvalho Campos L**, **Nunes LM**, **Tissot B and Trincão MS** (2024) Communication on central bank statistics: Unlocking the next level. *IFC Bulletin 60*,1–27.

**Dehghani Z** (2022) *Data Mesh: Delivering Data-Driven Value at Scale*. O'Reilly Media.

**Dilip A and Tissot B** (2024) "Development and maintenance of a security-by-security database", *IFC Guidance Note* 5, May.

**Dilip A**, **Mdingi Z**, **Sirello O and Tissot B** (2026) The evolution of the central bank statistical function. *IFC Bulletin 66*.

**Diz Dias J**, **Ramos MCG**, **de Andrés EM**, **Pastoris F**, **Schmitz M**, **Sirello O and Tissot B** (2025) External statistics in a fragmented and uncertain world: Addressing novel analytical challenges. *Statistical Journal of the IAOS 41*(2), 368–384. https://doi.org/10.1177/18747655251335759.

**Doan A**, **Halevy A and Ives Z** (2012) The future of data integration. In *Principles of Data Integration*. Elsevier. https://doi.org/10.1016/B978-0-12-416044-6.00019-3.

**Doerr S**, **Gambacorta L and Serena JM** (2021) Big data and machine learning in central banking. *BIS Working Papers 930*.

**Doll HC**, **Kormanyos E**, **Walter S and Werb GA** (2026) The climate data iceberg – A depth of information to integrate, *IFC Bulletin 66*.

**Elgg D**, **Park T**, **Sirello O**, **Tissot B and Pastor JMV** (2025) Leveraging corporate financial statements for policy insights. *IFC Bulletin 65*.

**Fellegi IP and Sunter AB** (1969) "A theory for record linkage". *Journal of the American Statistical Association 64*, (328). 1183–1210. https://doi.org/10.1080/01621459.1969.10501049.

**Fosen J**, **Holmberg A**, **Jansson I**, **Krapavickaitė D and de Waal T** (2025) Preface to the special issue on integrating data from multiple sources for production of official statistics. *Journal of Official Statistics 41*(2), 539–546. https://doi.org/10.1177/0282423X251332063.

**Fritz S**, **See L**, **Carlson T**, **Haklay M**, **Oliver JL**, **Fraisl D**, **Mondardini R**, **Brocklehurst M**, **Shanley LA**, **Schade S and Wehn U** (2019) Citizen sciences and the United Nations sustainable development goals. *Nature Sustainability 2*(10), 922–930.

**García del Riego M and Paúl I** (2024) "Spanish approach on FCC statistics: MNE statistics at sight?", presentation at the OECD Conference on National Economic and Financial Accounts, November.

**Gelman A and Henning C** (2017) Beyond subjective and objective in statistics. *Journal of the Royal Statistical Society Series A: Statistics in Society 180*(4), 967–1033. https://doi.org/10.1111/rssa.12276.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.

**Ghirelli C**, **Hurtado S**, **Pérez J and Urtasun A** (2021) New data sources for central banks. In Consoli S, Recupero DR and Saisana M (eds.), *Data Science for Economics and Finance*. Cham: Springer, pp. 169–194. https://doi.org/10.1007/978-3-030-66891-4_8.

**Gregory A** (2024) "Cross-domain interoperability framework (CDIF)", WorldFAIR: Global cooperation on FAIR data policy and practice.

**Guest G**, **MacQueen KM and Namey E** (2012) *Applied Thematic Analysis*. Thousand Oaks: SAGE Publications.

**Hammer C**, **Kostroch D and Quiros-Romero G** (2017) Big data: Potential, challenges and statistical implications. *IMF Staff Discussion Notes 17*(6), 1. https://doi.org/10.5089/9781484310908.006.

**Hansen JC**, **Cespedes LMS**, **Oyola LT**, **Dimakos XK**, **Walsh B**, **Bueno RS**, **Kabo-Bah AT**, **Seidu O and Nielsen VO** (2024) "Collaboration between national statistical offices and academia: Benefits, conditions, areas of collaboration and practical level experience in countries". *Statistical Journal of the IAOS 40*(1) , 161–172. https://doi.org/10.3233/SJI-230117.

**Harron K**, **Dibben C**, **Boyd J**, **Hjern A**, **Azimaee M**, **Barreto ML and Goldstein H** (2017) Challenges in administrative data linkage for research. *Big Data & Society 4*(2), 205395171774567. https://doi.org/10.1177/2053951717745678.

**Haryono E**, **Fahmi ATW**, **Gunawan I**, **Hidayanto A and Rahardjia U** (2020) "Comparison of the E-LT vs ETL method in data warehouse implementation: a qualitative study", 2020 International Conference on Informatics, Multimedia, Cyber and Information System, pp. 115–120.

**Hassani H**, **Mashhad LM**, **Stewart S and MacFeely S** (2025) Integrating GIS and official statistics using gisintegration. *Applied Mathematics 5*(4). https://doi.org/10.3390/appliedmath5040166.

**Holmberg A** (2025) Future pathways embracing multisource statistics and novel data sources at national statistical offices. *Journal of Official Statistics 41*(3), 795–803. https://doi.org/10.1177/0282423X251349082.

**Irving Fisher Committee on Central Bank Statistics (IFC)** (2020) "Computing platforms for big data analytics and artificial intelligence", *IFC Report* 11, April.

**Irving Fisher Committee on Central Bank Statistics (IFC)** (2023a) "Central banks' use of time series products", *IFC Guidance Note* 4, December.

**Irving Fisher Committee on Central Bank Statistics (IFC)** (2023b) "Data-sharing practices", *IFC Guidance Note* 3, March.

**Irving Fisher Committee on Central Bank Statistics (IFC)** (2024a) "Granular data: new horizons and challenges", *IFC Bulletin* 61, July.

**Irving Fisher Committee on Central Bank Statistics (IFC)** (2024b) "External statistics in a fragmented and uncertain world", *IFC Bulletin* 62, August.

**Irving Fisher Committee on Central Bank Statistics (IFC)** (2024c) "Communication on central bank statistics", *IFC Bulletin* 60, April.

**Irving Fisher Committee on Central Bank Statistics (IFC)** (2025a) "SDMX adoption and use of open source tools", *IFC Report* 17, February.

**Irving Fisher Committee on Central Bank Statistics (IFC)** (2025b) "Governance and implementation of artificial intelligence in central banks", *IFC Report* 18, April.

**Israël J-M and Tissot B** (2021) Incorporating microdata into macro policy decision-making. *Journal of Digital Banking 6*(3). https://doi.org/10.69554/MVIQ5259.

**Jellema T**, **Pastoris F and Aguilar CP** (2020) "Using synthetic indicators to assess the quality of macroeconomic statistics via mirror data", ECB Statistics Paper Series, no 34, March.

**Kim S**, **Jansen R**, **Jug M**, **Buckley D**, **de Fondeville R**, **Fadel AC**, **Saxena A**, **Stahl J and Hsiao W** (2025) A Delphi study on the role of privacy enhancing technologies (PETs) in data sharing ecosystems. *Statistical Journal of the IAOS 41*(3), 836–845. https://doi.org/10.1177/18747655251355706.

**Kitchin R** (2015) The opportunities, challenges and risks of big data for official statistics. *Statistical Journal of the IAOS 31*(3), 471–481. https://doi.org/10.3233/SJI-150906.

**Križman I and Tissot B** (2022) Data governance frameworks for official statistics and the integration of alternative sources. *Statistical Journal of the IAOS 38*(3), 947–955. https://doi.org/10.3233/SJI-220025.

**Leulescu A and Agafitei M** (2013) *Statistical Matching: A Model Based Approach for Data Integration*. Eurostat Working Papers.

**Lewaa I**, **Hafez MS and Ismail MA** (2021) Data integration using statistical matching techniques: A review. *Statistical Journal of the IAOS 37*(4), 1391–1410. https://doi.org/10.3233/SJI-210835.

**Ljones O** (2011): "Independence and ethical issues for modern use of administrative data in official statistics", *Statistical Journal of the IAOS 27* (1–2), 25–29. https://doi.org/10.3233/SJI-2011-0715.

**MacFeely S** (2020) In search of the data revolution: Has the official statistics paradigm shifted? *Statistical Journal of the IAOS 36*(4), 1075–1094. https://doi.org/10.3233/SJI-200662.

**MacFeely S** (2025) "The role of generative AI in generating official statistics", speech at the UNECE workshop on generative AI and official statistics, Geneva, 12 May.

**MacFeely S**, **Me A**, **Baeven R**, **Costanzo J**, **Passarelli D**, **Rossini C**, **Schueuer F**, **Veerappan M and Verhulst S** (2025a) Further reflections on the journey towards an international framework for data governance. *Statistical Journal of the IAOS 41*(2), 337–346. https://doi.org/10.1177/18747655251325830.

**MacFeely S**, **Me A**, **Schueuer F**, **Costanzo J**, **Passarelli D**, **Veerappan M and Verhulst S** (2025b) Towards a set of universal data principles. *Statistical Journal of the IAOS 41*(1), 150–155. https://doi.org/10.1177/18747655241307544.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.

**McGuire P**, **von Peter G and Zhu S** (2024) International finance through the lens of BIS statistics: Residence vs nationality. *BIS Quarterly Review*, 73–88.

**Mondal SS**, **Mondal S and Adhikari SK** (2023) Applications of big data in various fields: A survey. In Bhattacharyya S, Das G, De S and Mrsic L (eds), *Recent Trends in Intelligence Enabled Research: Advances in Intelligent Systems and Computing*. Singapore: Springer. https://doi.org/10.1007/978-981-99-1472-2_19.

**Moreno M** (2021) "Data governance: an orchestra of people, processes and technology", IFC Bulletin, no 54, July.

**Nelson A**, **Hogle P**, **Zanti S**, **Proescholdbell S and Tenenbaum J** (2024) A governance and legal framework for getting to 'yes' with enterprise-level data integration. *Data and Policy 6*, e31. https://doi.org/10.1017/dap.2024.23.

**Nowell LS**, **Norris JM**, **White DE and Moules NJ** (2017) Thematic analysis. *International Journal of Qualitative Methods 16*(1). https://doi.org/10.1177/1609406917733847.

**Nunna VVSB and Sahu R** (2025) Entity resolution in distributed systems: From fuzzy matching to knowledge graph integration. *Journal of Computer Science and Technology Studies 7*(8), 489–495.

**Organisation for Economic Co-operation and Development (OECD)** (2018) "Which strategies for NSOs in the digital era? Towards 'smart data' strategies", Committee on Statistics and Statistical Policy, August.

**Östberg P-O**, **Vyhmeister E**, **Castañé G**, **Meyers B and Van Noten J** (2022) Domain models and data modeling as drivers for data management: The ASSISTANT data fabric approach. *IFAC Papers 55*(10), 19–24. https://doi.org/10.1016/j.ifacol.2022.09.362.

**Paúl I** (2024) "The use of financial statements on NFC to estimate the valuation of unlisted shares and other equity of resident units in financial accounts", presentation at the OECD Conference on National Economic and Financial Accounts, November.

**Pérez AC and Huerga J** (2016) "The Centralised Securities Database (CSDB) – standardised micro data for financial stability purposes", IFC Bulletin 41, May.

**Pilgrim G and Ang S** (2024) "The OECD-UNSD multinational enterprise information platform", *OECD Statistics Working Papers* 1, February.

**Pradhan S-K and Silva J-F** (2019) "Using mirror data to track international banking", *IFC Working Papers* 19, December.

**Radermacher WJ** (2018) Official statistics in the era of big data opportunities and threats. *International Journal of Data Science and Analytics 6*(3), 225–231. https://doi.org/10.1007/s41060-018-0124-z.

**Ranbaduge T**, **Vatsalan D and Ding M** (2024) Privacy-preserving deep learning based record linkage. *IEEE Transactions on Knowledge and Data Engineering 36*(11), 6839–6850. https://doi.org/10.1109/TKDE.2023.3342757.

**Ricciato F** (2024) Steps toward a shared infrastructure for multi-party secure private computing in official statistics. *Journal of Official Statistics 40*(1), 3–15. https://doi.org/10.1177/0282423X241235259.

**Rosolia A**, **Stapel-Weber S and Tissot B** (2021) New developments in central bank statistics around the world. *Statistical Journal of the IAOS 37*(4), 1055–1060. https://doi.org/10.3233/SJI-210887.

**Sakshaug J and Steorts R** (2023) Recent advances in data integration. *Journal of Survey Statistics and Methodology 11*(3), 513–517. https://doi.org/10.1093/jssam/smad009.

**Schubert A** (2025) The eternal quest for quality in official. In Schipper U, Kirchner R and Walter J (eds), *The Eternal Quest for Quality in Official*. Cham: Springer, pp. 171–180. https://doi.org/10.1007/978-3-031-77193-4_12.

**Scotti C**, **Rondinelli C**, **Bottone M**, **Mattevi E and Neri A** (2024) Central bank business surveys: Version 2.0. *VoxEU 9*.

**Sexton A**, **Shepherd E**, **Duke-Williams O and Eveleigh A** (2018) The role and nature of consent in government administrative data. *Big Data & Society 5*(2). https://doi.org/10.1177/2053951718819560.

**Struijs P**, **Braaksma B and Daas PJH** (2014) Official statistics and big data. *Big Data & Society 1*(1). https://doi.org/10.1177/2053951714538417.

**Tissot B** (2016) Globalisation and financial stability risks: Is the residency-based approach of the national accounts old-fashioned? *BIS Working Papers 587*.

**Tissot B** (2017) "Big data and central banking", *IFC Bulletin* 44, September.

**Tissot B** (2019) Making the most of big data for financial stability purposes. In Strydom M and Strydom S (eds.), *Big Data Governance and Perspectives in Knowledge Management, IGI Global*, pp. 1–24.

**Tosi D**, **Kokay R and Roccetti M** (2024) 15 years of big data: A systematic literature review. *Journal of Big Data 11*(1). https://doi.org/10.1186/s40537-024-00914-9.

**Turner J** (2021) "OeNB's reporting data model as regtech/suptech solution", *IFC Bulletin* 53, April.

**United Nations** (UN) (2014) "Fundamental Principles of Official Statistics", resolution adopted by the General Assembly, A/RES/68/261.

**United Nations Economic Commission for Europe (UNECE)** (2021a) *Measuring the value of official statistics: testing and developing a measurement framework*, January.

**United Nations Economic Commission for Europe (UNECE)** (2021b) *Guide to sharing economic data in official statistics*, February.

**United Nations Economic Commission for Europe (UNECE)** (2023) *Synthetic data for official statistics – a starter guide*, January.

**United Nations Economic Commission for Europe (UNECE)** (2024a) In-depth review of linking data across domains and sources, Conference of European Statisticians, ECE/CES/2024/5, April.

**United Nations Economic Commission for Europe (UNECE)** (2024b) *Cloud for official statistics*, March.

**United Nations Economic Commission for Europe (UNECE)** (2025a) *Generic Statistical Business Process Model*, May.

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.

**United Nations Economic Commission for Europe (UNECE)** (2025b) *The future of national statistical offices: a call to action*, May.

**United Nations Statistics Division (UNSD)** (2025) *Handbook on management and organization of national statistical systems*, May.

**United Nations System Chief Executives Board for Coordination (UN-CEB)** (2023): International data governance – pathways to progress, CEB/2023/1/Add.2, May.

**Varian HR** (2014) Big data: New tricks for econometrics. *Journal of Economic Perspectives 28*, *28*(2), 3.

**Verhulst S**, **Zahuranec A and Chafetz H** (2025) Moving toward the FAIR-R principles: Advancing AI-ready data. *SSRN 5164337*,10.

**Veronese G**, **Rosolia A**, **Venditti F and Biancotti C** (2020) Saving economic data from Covid-19. *VoxEU 12*.

**Viggo Sæbø H and Hoel M** (2023) Official statistics: Quo vadis? *Statistical Journal of the IAOS 39*(1), 39–45. https://doi.org/10.3233/SJI-220100.

**Wallgren A and Wallgren B** (2014) *Register-Based Statistics: Statistical Methods for Administrative Data*. Wiley. https://doi.org/10.1002/9781118855959.

**Woloszko N** (2024) Nowcasting with panels and alternative data: The OECD weekly tracker. *International Journal of Forecasting 40*(4), 1302–1335. https://doi.org/10.1016/j.ijforecast.2023.11.005.

**Cite this article:** Sirello O and Tissot B (2026). Linking data for better evidence-based policy: A landscape review in central banks. *Data & Policy*, 8: e28. doi:10.1017/dap.2026.10073

Downloaded from https://www.cambridge.org/core. 17 Nov 2026 at 14:18:38, subject to the Cambridge Core terms of use.