

Guidance Note 3 – Data-sharing practices

This guidance note¹ is based on the reported experience from the IFC network on central banks' data-sharing practices. It reviews a number of specific issues, related in particular to:

- *Ongoing initiatives to promote data-sharing;*
- *The use of common statistical identifiers and agreed standards;*
- *National practical experiences to share granular data; and*
- *The balance between confidentiality and users' needs.*

1. Data-sharing – general context

Data-sharing, which is usually understood as the practice of making data resources (collected for a specific intention) available to persons and/or entities for various purposes,² has become a key issue for official statisticians. It can certainly help to improve the quality of economic statistics, but it also raises a number of operational (eg identification, IT infrastructure) as well as ethical (eg confidentiality protection, privacy) challenges.³

In fact, a key prerequisite for ensuring the quality of (national) official statistics is that *"data sharing and coordination among data-producing agencies are adequate"*, as emphasised in the IMF's Data Quality Assessment Framework (DQAF; § 0.1.2). Turning to the global level, **the push for developing statistical data-sharing has clearly gained momentum since the Great Financial Crisis** (GFC) of 2007–09. This reflects the recognition that the exchange of detailed information has become more important for better informed decision-making in the increasingly globalised economic and financial environment where entities are more interconnected within and across borders. In particular, the three phases of the Data Gaps Initiative (DGI) developed by the G20 after

¹ This analysis produced by the IFC Secretariat ([Bruno.Tissot@bis.org](mailto: Bruno.Tissot@bis.org); [IFC.Secretariat@bis.org](mailto: IFC.Secretariat@bis.org)) does not necessarily reflect the views of the BIS or IFC members and the various institutions mentioned in this note. Redentor Paolo M Alegre Jr, Marcia Cavalinhos, Cécile Golfier, Robert Kirchner, Maria do Carmo Moreno, Li Ming Ong, Silke Stapel-Weber, Luís Teles Dias and Caroline Willeke are very much thanked for their comments and suggestions.

² A more detailed [definition of data-sharing is proposed by Amazon Web Services \(AWS\)](#) as follows: *"Data sharing is the process of making the same data resources available to multiple applications, users, or organizations. It includes technologies, practices, legal frameworks, and cultural elements that facilitate secure data access for multiple entities without compromising data integrity. Data sharing improves efficiency within an organization and fosters collaboration with vendors and partners. Awareness of the risks and opportunities of shared data is integral to the process."*

³ For a comprehensive overview of data-sharing in official statistics, see United Nations Economic Commission for Europe (UNECE) (2020): [Guide to Sharing Economic Data in Official Statistics](#), ECE/CES/STAT/2020/5.

the GFC have comprised several specific plans to promote the sharing and accessibility of economic and financial statistics.⁴

In this post-GFC context, **seven key recommendations** have been identified to foster data-sharing, namely:⁵

- Promoting the use of common statistical identifiers;
- Promoting the exchange of experience on statistical work with granular data and improving transparency;
- Balancing confidentiality and users' needs;
- Linking different datasets;
- Provision of data at the international level;
- Consideration of ways of improved data-sharing of granular data; and
- Collection of data only once.

Central banks, as both users and compilers of official statistics, have been at the forefront of the above initiatives to develop data-sharing both at national and international levels, not least to support public policy.⁶ They have called for more and better cooperation between stakeholders, *eg taking stock of useful projects, promoting best practices and identifying novel methods*. They have in particular underlined the essential role that data-sharing can play in ensuring data availability in the broadest and most efficient way, while keeping the reporting burden under control.

On its side, the **IFC has advanced work on specific areas**, such as the sharing of banking data between supervisors and central bank statisticians (including the identification of a set of "good practices"; [IFC Report no 1](#)); the access to, and sharing of, micro data sources ([IFC Bulletin no 53](#)); and related data governance issues ([IFC Bulletin no 54](#)). In addition, the Committee recently collected feedback from its members to take stock of the practices being developed to promote the use of common statistical identifiers and agreed standards (Section 2); (ii) develop the exchange of experiences when dealing with granular data (Section 3); (iii) and balance confidentiality and users' needs (Section 4).

⁴ The [first phase \(2009–15\)](#) of the DGI called for investigating "the possibility of improved collection and sharing of information on linkages between individual financial institutions". The [second phase \(2016–21\)](#) aimed to "promote and encourage the exchange of data and metadata among and within G-20 economies, and with international agencies, to improve the quality (eg, consistency) of data, and availability for policy use"; the G-20 economies were also "encouraged to increase the sharing and accessibility of granular data, if needed by revisiting existing confidentiality constraints". The [third phase \(2022–27\)](#) aims to address priority policy needs in the areas of "access to private sources of data and administrative data, and data sharing to improve the timeliness and granularity of official statistics".

⁵ Inter-Agency Group on Economic and Financial Statistics (IAG) (2017): [Update on the Data Gaps Initiative and the outcome of the Workshop on Data Sharing](#), March.

⁶ Reflecting the recognition of the importance of sharing data and information to support public policy and the implications for central banks and providers of statistical information; see Buch, C (2021): "[Data sharing for better policy making](#)", Introductory statement at the G20 DGI-2 Workshop on Data Sharing, 24 March.

2. The use of common statistical identifiers and agreed standards

Background

As compilers of official statistics, central banks have a long experience in collecting and aggregating information from various data sources, which are typically based primarily on reports filled by financial corporations.⁷ In practice, such compilation exercises require the use of a “data identifier”, ie “an identifying data attribute that, alone or in combination with other identifiers, ensures that each record or entry can be (uniquely) identified in an entity table. Identifiers help to maintain data integrity and to identify relationships between datasets”.⁸

Various lists of identifiers have been developed over time to support statistical compilation exercises in different domains, reflecting the need to correctly identify individual reporters and derive meaningful aggregations. These lists, typically under the control of statistical agencies (and thus central banks for the specific areas under their responsibility), can differ markedly in practice among national compilers, and even among statistical domains within the same institution. Needless to say, these differences are even more marked across jurisdictions.

The existence of different types of identifiers can raise several challenges. One is complexity, as individuals or firms have to keep track of their various identification numbers used by data collectors (eg social security number, tax account number, etc.). A second problem is the difficulty to merge data sets based on different identifiers. A third issue is the obstacle posed to data-sharing, since valuable information cannot be easily exchanged between statistical agencies that do not/cannot share a common identification list.

The development of common identifiers that may be used by different agencies can help to address these challenges.⁹ The feedback collected by the IFC shows that, while several central banks continue to rely on their own statistical identifiers especially in the context of banking supervision exercises, there is clear trend towards the use of shared identifiers.

Related initiatives

At the **national level**, central banks and national statistics institutes (NSIs) in various jurisdictions have already agreed to use common identifiers to make data reporting exercises simpler and more efficient. One successful example has been the development of the Simplified Corporate Information in Portugal (with the local acronym IES) since the mid-2000s: as a result, Portuguese firms can submit their information of accounting, fiscal and statistical natures simultaneously to four distinct public entities (including Banco de Portugal and the NSI) through one single electronic submission.¹⁰ This initiative has in particular helped to reduce the number of reporting exercises in

⁷ There are certainly different situations depending on the role played by central banks in the organisation of official statistics in a given jurisdiction, and a number of central banks do also collect micro-level information from non-financial firms or households for specific purposes.

⁸ Cf ECB’s [AnaCredit Reporting Manual – Part I – General methodology](#).

⁹ As emphasised by Knot, K (2020): “[The value of the LEI for identification in digital processes](#)”, speech at the GLEIF Meet the Market Event, Amsterdam, 4 February.

¹⁰ See Banco de Portugal (2008): “[Simplified reporting: inclusion of IES in the statistics on non-financial corporations from the Central Balance-Sheet Database](#)”, *Statistical Bulletin*, Supplement 1|2008, May.

the country and also contributed to enhance the quality and usefulness of the information collected – for instance, to better track the role of multinational enterprises (MNEs).¹¹ A related initiative is the Austrian central bank's reporting data model that leverage on the joint AuRep platform set up by the main Austrian commercial banks with the view of making more efficient the overall reporting of their data to public authorities.¹² Another example of initiative to standardise banks' data reporting is the Eurosystem Integrated Reporting Framework (IReF) being developed in Europe, which aims to integrate existing statistical data requirements for banks into a single, standardised reporting framework applicable across the euro area.¹³

At the **global level**, one key project endorsed by global authorities after the GFC has been the introduction of the Legal Entity Identifier (LEI), a 20-digit reference code to uniquely identify legally distinct entities that engage in financial transactions, which has already become mandatory in several jurisdictions. In addition, work is ongoing to develop principles and standards for collecting and consolidating information on the direct and ultimate parents of legal entities, through the identification of parent relationships between LEIs.¹⁴ Furthermore, additional progress is under way in the standardisation of the reporting financial operations – including the definition of a unique transaction identifier (UTI) and unique product identifier (UPI).

The experience of IFC members shows that the **LEI is already used extensively for statistical work in several jurisdictions**, especially in Europe. Among euro area central banks, for instance, it is reported to support important compilation exercises such as the Money Market Statistical Reporting (MMSR), the Register of Institutions and Affiliates Database (RIAD), the Securities Holdings Statistics Database (SHSDB), the Central Securities Database (CSDB), the European Market Infrastructure Regulation (EMIR) and the AnaCredit exercises coordinated by the ECB.¹⁵ The LEI can also be used as an identifier, in addition to other sources, to report granular information of relevant financial institutions to the BIS.¹⁶ It is also considered as an important element for enhancing cross-border payments.¹⁷

Yet, **in practice the LEI is often used as an additional identifier in addition to existing national ones**. Those can, depending on specific jurisdictions, be used extensively by central banks

¹¹ Cf Pinto, A B, J A Neves and T P Pereira (2019): "[Keeping track of MNEs through business group databases: the experience of Bank of Portugal](#)", *IFC Bulletin* no 49, January.

¹² See Kienecker, K, G Sedlacek and J Turner (2018): "[Managing the processing chain from banks' source data to statistical and regulatory reports in Austria](#)", Oesterreichische Nationalbank.

¹³ See ECB (2022): "[The Eurosystem Integrated Reporting Framework: an overview](#)", September.

¹⁴ See <https://www.leiroc.org/>, and in particular Legal Entity Identifier Regulatory Oversight Committee (LEIROC) (2016): "[Collecting data on direct and ultimate parents of legal entities in the Global LEI System – Phase 1](#)", March; and LEIROC (2022): "[Progress Report 2019-2021](#)", January. The Global LEI Foundation (GLEIF) also provides on its website a centralised database of LEIs and corresponding reference data. In Europe, the EMIR European regulation imposes, for certain derivative products, the use of a unique identifier (the LEI) for parties to financial transactions.

¹⁵ For instance, the ECB displays on its [website](#) the related list of all national identifiers used in the context of AnaCredit reporting.

¹⁶ One example relates to the actual sharing of granular information on global financial institutions (see Bese Goksu, E and B Tissot (2018): "[Monitoring systemic institutions for the analysis of micro-macro linkages and network effects](#)", *Journal of Mathematics and Statistical Science*, vol 4, April).

¹⁷ See Committee on Payments and Market Infrastructures (CPMI) (2020): "[Enhancing cross-border payments: building blocks of a global roadmap – technical background report](#)", July.

for instance to cover other firms as well as customers of financial services.¹⁸ One particular issue is that the LEI coverage for non-financial entities, especially small to medium-sized enterprise (SMEs), remains modest, hindering the effective integration of financial and non-financial information. There are indeed a number of dedicated initiatives underway to overcome this challenge.¹⁹ Moreover, and despite the development of the LEI, several central banks are still using other identifiers for reporting financial institutions.²⁰ Lastly, there are other important global identifiers for different domains that can facilitate the sharing of micro-level information, such as the ISIN (International Securities Identification Number) for securities.²¹

A related but important dimension for the international sharing of information relates to **standardisation**, so that common data formats and types as well as descriptive information on the data themselves (ie, “metadata”) can be exchanged. One example is the Statistical Data and Metadata eXchange standard ([SDMX](#)), already used by many central banks for the handling of macro variables. In particular, the latest version SDMX 3.0 comprises new features supporting microdata use in addition to aggregated time series.²² The guidelines developed by international bodies, such as the IMF special data dissemination standard ([SDDS](#)) and the Enhanced General Data Dissemination System ([eGDDS](#)), are also important references governing the international exchange of statistical data and metadata. The experience of IFC members shows that central banks have been actively committed to follow such international conventions, not least to facilitate their participation in global statistical fora and benefit from adequate international technical assistance for the compilation, management, dissemination, and improvement of statistics.

3. Sharing granular data: taking stock of national experiences

While there is a widespread recognition of the benefits of using and exchanging granular data to better support statistical work as well as policy-making, one important question is **how to organise data-sharing both within and among various authorities** such as statistical agencies, government offices and other public bodies as well as with academic researchers and the general public. The reported experience of IFC members sheds interesting light on these aspects, covering the four main areas of focus for central banks: (i) their internal users; (ii) public authorities, including

¹⁸ An illustration of this complementarity between the LEI and a national identifier can be found on the [INSEE SIRENE register](#) (the French national identification system for natural and legal persons and their establishments).

¹⁹ See European Systemic Risk Board (ESRB) (2020): “[Recommendation of the European Systemic Risk Board on identifying legal entities](#)”, ESRB/2020/12, 24 September; and ESRB (2021): “[The benefits of the Legal Entity Identifier for monitoring systemic risk](#)”, *Occasional Paper*, no 18, September.

²⁰ As mentioned above, one example is the European AnaCredit initiative that relies on the sharing of information on national identifiers that can be validated in each country.

²¹ For instance, ISIN underpins the European Centralised Securities Database (CSDB), a security-by-security database accessible by the entire European System of Central Banks (ESCB); cf Pérez, AC and J Huerga (2016): “[The Centralised Securities Database \(CSDB\) - Standardised micro data for financial stability purposes](#)”, *IFC Bulletin* no 41, May.

²² Cf IFC (2016): “[Central banks’ use of the SDMX standard](#)”, *IFC Report*, no 4, March. The new 3.0 version launched in 2021 will be key to support data-sharing and official statistics more generally, as analysed in Tissot, B (2021): “[SDMX as a standard for data sharing / standardization](#)”, G20 DGI-2 Workshop “Promotion of Data Sharing”, March. Moreover, work is under way, for instance in [Europe](#), to enhance the mapping between SDMX and other standards such as XBRL (eXtensible Business Reporting Language) supporting the “data point model” used in prudential and resolution reporting. Lastly, the new DGI launched in 2022 will work on a possible microdata standard in close cooperation with the ongoing development of the SDMX standard for microdata (among other initiatives).

their counterparts in National Statistical Systems (NSSs); (iii) academic researchers; and (iv) the general public more broadly.²³

Internal users

One often understated issue is the importance of promoting data-sharing **among internal users in central banks**. Because most of the data collected and processed are for specific purposes (eg payments monitoring, statistical production, supervision), they are often stored and administered in separate, independent operational systems that are accessible only to limited groups of internal staff. This can result in “data silos”, with isolated infrastructures in terms of data bases and related IT tools, and in turn potential inefficiencies when making use of the information collected. Other drawbacks may comprise the sending of competing data requests to external reporters, limited internal awareness of the available data within organisations, heterogeneous data quality assurance practices, etc. These challenges are particularly evident as regards the internal sharing of supervisory data, even when the central bank is in charge of financial supervision (cf [IFC Report no 1](#)).

To address these issues, **several central banks have established committees/programmes to foster the internal sharing of, and access to, granular data**. Such initiatives typically aim to set up standardised procedures for granting information access for internal analysts. They can further include an intensive stock-taking of all the granular data sources available and a review of the costs/benefits related to their usage. One example is the Bundesbank IMIDIAS (Integrated Micro Data-based Information and Analysis System) initiative, which provides an integrated, cross-divisional information system for analysis and research purposes.²⁴ This system is micro data-based and aims at different user groups within the Bundesbank. It consists of three components: the House of Microdata (HoM), the Research Data and Service Centre (RDSC; cf below) and an integrated data management process through common identifiers and data models. Internal users of IMIDIAS may be granted direct access to the HoM, while external researchers can only access micro data through the RDSC. The initiative has helped to take stock of all the data sources available, conduct cost/benefit usage analyses, review the various data quality processes, and establish standardised access procedures. Another example is the Integrated Data Management (IDM) programme of Banco de Portugal, which fosters a greater and better use of the data available in the central bank and promotes transparency and internal sharing based on a dedicated governance framework.²⁵

²³ As regards international data-sharing, however, there has been arguably more limited progress in exchanging granular information, except in specific regions such as the European Union or in the context of BIS global statistical activities (cf Tissot, B (2017): “[Data sharing: what, why and how? The BIS experience](#)”, October).

²⁴ See Schönberg, T (2019): “[Data access to micro data of the Deutsche Bundesbank](#)”, *Technical Report*, no 2019-02.

²⁵ Covering various aspects, including the integrated data collections models, a data catalogue, a single reference data management system, a corporate data warehouse, and a data architecture enabling efficient data exploration; cf Moreno, M (2021): “[Data governance: an orchestra of people, processes and technology](#)”, *IFC Bulletin* no 54, July.

Public authorities

Central banks also report close cooperation to enhance data-sharing with other public authorities. The main issue is to clearly define who is responsible for collecting which data. In addition, formal agreements have been developed to facilitate access by all relevant parties. The aim is to allow data to be exchanged at a granular level, in line with well-documented procedures, and be further disseminated in a more aggregated way depending on the respective purposes of the institutions involved.

Needless to say, and for instance in the context of the European Union (EU), such data-sharing arrangements are often well advanced between central banks and their **NSIs counterparts**. In particular, a specific Memorandum of Understanding ([MoU](#)) sets out the respective areas of responsibility in economic and financial statistics at the Community level of the ECB (Directorate General Statistics) and the Commission (Eurostat) and provides a framework for the exchange and reproduction of data. Moreover, efforts are under way to promote the cross-border exchange of information, for instance to cover MNEs. Nevertheless, an important task is to remove remaining hurdles for an effective exchange of confidential data for statistical purposes between central banks and NSIs of different countries in this context.

The [IMF SDDS Plus webpage](#) provides detailed information on the related existing arrangements in a number of countries, in terms for instance of data-sharing and coordination among data producing agencies as well as on the confidentiality of individual reporters' data. In particular, countries that adhere to the SDDS Plus standard make a commitment to provide [detailed country information](#) about their data and data dissemination practices. One example is [France](#), where the [Official Statistical System](#) includes the NSI (INSEE) and 16 Ministerial Statistical Offices who carry out statistical operations in their field of competence. The Bank of France is independent from this official system and the exchange of statistical information is organised with INSEE under a specific MoU. A different example is [Portugal](#), where under a 2008 law the NSS is comprising the NSI, the central bank, two regional statistical services, and the producers of official statistics by delegation of powers from the NSI. The sharing of data between these authorities is also regulated by law, with signed MoUs for operational purposes.

Another important group for central banks comprises public **authorities in charge of financial regulation and prudential supervision** (when such tasks are not performed by central banks). Several initiatives have been developed to allow the information collected for supervisory purposes (on, *eg*, *banks, pension funds, insurance companies*), which is to be shared and used by central bank statisticians depending on national cooperation agreements. One illustration is the French Unified Financial Reporting System ("RUBA"), under which the financial information collected from banks and other reporting institutions can be used for the different needs within the Bank of France and the French supervisory authority (mainly for statistical and prudential purposes).

Researchers

Central banks have launched important efforts to develop research data centres to allow external academics to access their data for analytical purposes. These initiatives reflect the growing recognition that using individual data is necessary to accurately reflect the heterogeneity of observation units and perform relevant causal analyses. In practice, granular data cannot be made generally available, but can be used for clearly defined academic research purposes, in compliance with all legal and internal regulations on data processing and data disclosure.

Noteworthy examples include the [Bank of Italy's RDSC](#) that represents the centralised point of access for the microdata that the central bank makes available to researchers and institutions, exclusively for research purposes. Depending on the characteristics of each database, the data can be made accessible either as Public Use Files (PUF) or through the Bank of Italy's remote execution system. Another open data room example is the Bank of France's Confidential Data Access Portal ([CDAP](#)) which gives access to a large set of databases stored in the Secure Data Access Center ([CASD](#)), along with other official public statistics. This enables academic researchers (including a number of foreign research groups) to access a large set of databases and link microdata from different sources. Each application for accessing the data is evaluated by the Data Access Secretariat and methodological specialists. Turning to Germany, the [Bundesbank's RDSC](#) provides standardised access to selected microdata collected by the central bank to be used in independent scientific research projects. Moreover there is a strong coordination between the various RDSCs dealing with social, empirical and economic sciences in the country, including with those of the NSI and of the federal statistical institutes, organised in the context of the [German Data Forum](#). Similarly, the Banco de Portugal Microdata Research Laboratory ([BPLIM](#)) allows both internal and external researchers to gain access to well-documented and anonymised micro datasets customised to their particular needs.

Lastly, a key initiative to promote knowledge-sharing on micro data issues relates to the group of 13 central banks involved (as members or observers), together with a number of NSIs and international organisations, in [INEXDA, the International Network for Exchanging Experience on Statistical Handling of Granular Data launched in 2017](#). The IFC has energetically supported this international project, whose aim is to facilitate the use of granular data for analytical, research and comparative purposes by users outside the participating institutions within the limits set by the applicable confidentiality regimes. INEXDA work has comprised in particular the development of a metadata schema to describe granular data sets from different countries, the review of best practices for granting access to open software solutions and data, and the identification of common features across jurisdictions with a view to a potential harmonisation of data access procedures. This has allowed participating central banks to exchange on best practices, enhance their knowledge of data-sharing service offerings by external providers, and discuss concrete use cases for data-sharing, such as the reduction of bilateral asymmetries in international statistical exercises.²⁶

General public

Data-sharing with external stakeholders in the general public is also an important issue for central banks. One example of initiative to promote the use of granular data in compliance with confidentiality requirements is the [Financial Big Data Cluster](#) (FBDC) currently developed in Germany. This specific project aims at allowing a wide range of interested parties, including private financial service providers, authorities, firms as well as research institutions to use a common cloud-based data platform. Its general goal is to develop an industry-wide ecosystem that (i) makes financial data accessible to different user groups in gradations and (ii) provides analysis tools, data exchange tools and computing capacity. Another initiative is the [Open finance framework](#) planned by the European Commission to enable data sharing and third party access in the financial sector.

²⁶ For a review of INEXDA's role, see Members of the INEXDA network (2018): "[INEXDA – the Granular Data Network](#)", *IFC Working Papers* no 18, October; for a discussion of "mirror data-sharing", see Pradhan, S-K and J F Silva (2019): "[Using mirror data to track international banking](#)", *IFC Working Papers* no 19, December.

This open finance proposal relates to the EU digital finance strategy and covers a wide range of financial sectors and products, in line with data protection and consumer protection rules.

A further example is the [Open Data Philippines](#) (ODPH) Portal, which is an online repository of open data from different government agencies to address citizens' rights to information and data on matters of public concern in the Philippines. This initiative promotes transparent, participative, inclusive, and accountable governance for the collection of various datasets from different government agencies that are deemed "open" or public and can then be published on the single platform. In practice, the project has allowed multiple users, including the central bank, to find the specific data they need from a continuously growing collection of public datasets. It also facilitated the implementation of the "*collection of data only once*" culture in the country.

4. Balancing confidentiality and users' needs

Confidentiality is one of the key Fundamental Principles of Official Statistics. It means that "*individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes*". Data are deemed confidential when individual reporting statistical units (legal or natural persons) can be identified, either directly (through the disclosure of their formal identifiers) or indirectly (by inferring the respondent's identity from the data disclosed).²⁷ The general confidentiality protection principle means that the level of details of the information that can be passed to users is limited.²⁸ In addition, dedicated data protection frameworks have been set up across countries to regulate the use and sharing of granular information so as to respect specific confidentiality rules and privacy standards.²⁹

The reported experience of IFC members highlights a number of important points from this perspective:

- The **confidentiality of data and its protection is an important issue** for central banks that rely on public trust in pursuing their policy mandates. In this context, specific data and classification systems may be used effectively, to divide data into certain levels according to their confidentiality, and in turn determine how the data must be handled.

²⁷ There are many ways in practice to protect data confidentiality, which can be defined as "*a property of data, usually resulting from legislative measures, which prevents it from unauthorized disclosure*" and is typically implemented by defining which cells of a statistical table can be published or not; cf OECD (2008): [OECD Glossary of Statistical Terms](#), OECD Publishing, Paris. One typically distinguishes between the risks of "direct disclosure" (ie disclosure of information with identifiers) and "statistical disclosure" (ie the risk of re-identification of anonymised data, a risk that is growing with the increased IT computing power available and the multiplication of cross-sectional and longitudinal databases that one can try to link); cf National Research Council (2005): "[Expanding Access to Research Data: Reconciling Risks and Opportunities](#)", Washington, DC: The National Academies Press.

²⁸ See UNECE (2019): [Principles and guidelines on confidentiality aspects of data integration undertaken for statistical or related research purposes](#), Geneva.

²⁹ As regards the European Union, the data protection rules and related mechanisms were established in a specific regulation, the 2016 [General Data Protection Regulation \(GDPR\)](#), and [specific rules](#) apply to the processing of personal data collected for administrative, commercial, statistical or other purposes (including by central banks).

- It is essential that **reporting units can count on a sound confidentiality protection framework** when transmitting the data.³⁰ Most central banks have thus developed adequate, enterprise-wide data governance frameworks and policies that ensure data quality and integrity and support trust and confidence among participants.³¹ One option is to set up certificates ensuring a desired level of protection/security when collecting the data.³² Moreover, regular reviews should be conducted to check whether adequate disclosure control rules and practices are in place and are kept up to date.³³
- Such frameworks typically include various **sanctions or penalties for data mismanagement to support confidentiality protection**, as is common practice in a number of NSIs for their staff. In France, for instance, agents of the official statistical service are bound by the obligation of professional confidentiality, reinforced for confidential data by the “statistical confidentiality” rule which strictly prohibits the communication of individual data.³⁴ Central banks have been also clarifying the penalties being applied in case of unauthorised disclosure of confidential information (eg containing bank secrecy), for instance in Armenia.³⁵ More generally, the EU has set up rules governing penalties, including administrative fines, to be imposed for any infringement of the GDPR regulation.
- When making **data publicly available**, central banks ensure that the information is sufficiently aggregated so that it cannot be traced back to individual reporters. To do so, the potential number of reporters can be considered rather than their actual number – eg, in the context of the so-called “minimum three” (or “minimum five”) reporters rule traditionally applied for determining that a certain aggregate is deemed confidential.³⁶

³⁰ For a discussion on the importance of data’s confidentiality, not least to support the accuracy of official statistics and data sharing, see National Research Council (2005) above, esp. chapter 4 “Risks of access: potential confidentiality breaches and their consequences”.

³¹ On the link between data-sharing and data governance framework, see: Tiwari, S, S Sharma, S Shetty and F Packer (2022): “[The design of a data governance system](#)”, *BIS Papers* no 124, July. One example of a related initiative in the Philippines was the 2017 [Guidelines on information security management](#) issued by Bangko Sentral ng Pilipinas for supervised financial institutions, with specific considerations for information sharing and collaboration and the treatment of confidential/sensitive data. Similarly, the Central Bank of Ireland provides information about the ways in which it collects and uses personal data in its [data protection notice](#).

³² Such certificates will typically state that data identified to a specific individual will not be used or revealed unless for research and statistical purposes and on a need-to-know basis. They will also ensure that any recipient of this information is bound to these obligations and has to take adequate measures. In practice, staff accessing the data on a need-to-know basis are formally notified of their related confidentiality requirements and must agree to abide by them.

³³ One example is the regular [review](#) of the statistical confidentiality protection framework adopted by the ESCB.

³⁴ See INSEE (2023): “[Guide to statistical confidentiality](#)”, January. In terms of actual sanctions, for instance, agents are liable to criminal penalties of up to one year’s imprisonment and a fine of 15,000 euros in the event of a breach of the professional confidentiality obligation. Turning to the external parties being granted personalised authorisation to access the NSI secure remote access centre, they must sign a confidentiality agreement and are exposed to “severe penalties” in the event of non-compliance.

³⁵ For instance in Armenia; cf Central Bank of Armenia (2008): “[Statistical code of practice](#)”.

³⁶ One recent example of such an approach relates to Canada; cf UNECE (2017): “[New approach to data suppression in macroeconomic accounts](#)”, prepared by the Statistics Canada for the Group of Experts on National Accounts, Sixteenth session, 5 May. In this context, the statistical agency had updated its interpretation of the Statistics Act to develop a new approach to identify data cells to be suppressed for reasons of confidentiality, based on an assessment of the **potential** population rather than the reporting population.

-
- Applying “**passive**” **confidentiality rules** (that is, only labelling the information as confidential when the reporter asks for it) could be an effective option.
 - A common practice for **internal users of data** is to have them sign confidentiality agreements, and to perform regular reviews of the related access rights. Another is to have specific functions dedicated to manage/verify the granting of access rights and to work on the related operational aspects (eg IT, security, legal). For instance, to ensure that internal units do access the relevant information when the statistical database system is managed centrally in the central bank; maintain a proper data governance framework – eg drafting of a data governance policy, guiding manual for data users, supervision of the various data-sharing agreements; and promote (international) best practices.
 - As regards **external users**, the sharing of central banks’ data with national competent authorities is typically governed by formal MoUs to ensure confidentiality protection. Other contractual arrangements can include Non-Disclosure Agreements, to guard against the risk of publishing confidential data by third parties, as well as specific Service Level Agreements (SLAs) governing the particular purposes for which the data are shared.
 - **Innovative data access procedures** can be developed to facilitate the work on confidential data by other external users, such as academic researchers especially in the context of INEXDA (cf above). In particular, a number of ways are available to share the results of computations instead of actual data and to enable the protection of the privacy of sensitive information. Such privacy enhancing techniques include secure multiparty computation, homomorphic encryption, trusted execution environments, and differential privacy.³⁷

³⁷ For a more detailed description and examples of pilots conducted under the UN framework, see UNECE (2020): [Guide to Sharing Economic Data in Official Statistics](#), ECE/CES/STAT/2020/5, pp 23-25.