



EUROPEAN CENTRAL BANK
EUROSYSTEM

Statistics Paper Series

Antonio Perrella, Julia Catz **Integrating microdata
for policy needs:
the ESCB experience**

No 33 / February 2020

Disclaimer: This paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB or the Bank for International Settlements.

Contents

Abstract	2
1 Introduction	3
2 The importance of (integrated) microdata	6
2.1 Why microdata?	6
Box 1 Concentration risk	8
Box 2 Specific wrong-way risk	9
2.2 Challenges associated with microdata	9
3 ESCB microdata	11
3.1 The crucial role of RIAD	11
Box 3 Group structures in RIAD	12
Box 4 RIAD as a pivotal tool for monitoring compliance with collateral management rules	14
3.2 Microdata on credit	15
4 Linking the microdata	18
Box 5 Benefit of linking the microdata: unique identification of entities	18
4.1 Challenges	20
4.2 The ESCB approach	21
4.3 Limitations and way forward	22
5 Conclusion	25

Abstract

Since the financial crisis, central bank policymakers have expressed a need for more integrated microdata for monetary policy purposes and for macroprudential and microprudential supervision, with a stronger focus on lending. In response to this policy need, the European System of Central Banks (ESCB) has increased the scope and quality of instrument-level data (e.g. loan-by-loan) it collects. At the same time, the ESCB has further developed the Register of Institutions and Affiliates Data (RIAD), which is pivotal in ensuring the successful linking of the databases, because it ensures the unique identification of counterparties. RIAD allows data to be aggregated using various types of company information, such as industrial activity or geographical location, but it also offers the possibility of aggregating data according to multiple group structures based on different concepts of what a “group” is.

This paper discusses why there is a policy need for microdata and highlights some of the practical uses of the interlinked data. It also sheds more light on how information contained in different granular databases can be combined and aggregated in a flexible manner according to different business needs. It describes in detail the process of linking through a common stable identifier, points out current limitations and suggests a possible way forward.

Keywords: microeconomic data, granular data, unique identification, group structures, macroprudential, microprudential

JEL classification: C81, E44, G32

1 Introduction

Since the financial crisis, central bank policymakers have expressed a need for more granular data for monetary policy purposes and for macroprudential and microprudential supervision. As former ECB Executive Board Member Benoît Coeuré phrased it: “Authorities – including central banks – need high-quality financial data at a granular and aggregate level to perform several of their functions, including: conducting monetary policy, assessing systemic risks, supervising banks, performing market surveillance and enforcing and conducting resolution activities.”¹

The vast technological improvements we have seen in the last two decades have played a fundamental role in increasing the demand for granular data.² On the one hand, analysts now have the tools they need to process and analyse large sets of data; on the other hand, statisticians also have the information technology (IT) infrastructure they need to collect, store and manage large datasets. In a rapidly changing world, central banks are also evolving, and policymakers are keen to access as much detailed information as possible in order to identify behavioural patterns, drivers of systemic risk and potential sources of contagion.³ Insofar as the technology allows, policymakers will look at problems from multiple perspectives, exploiting the technological capacity to process more and more data.

Large, granular datasets, with a broad coverage of relevant topics and interconnected with each other, attract the attention of central bankers, regardless of the area in which they work. For instance, one goal of policymakers is to recognise and react to threats to the stability of the financial system through the early identification and assessment of potential sources of stress in financial markets. In this respect, it is useful to analyse the total exposure of creditors (lenders/holders) to debtors (borrowers/issuers) in order to monitor the soundness of the lending sector (banks and non-banks). On the other hand, it is also helpful to monitor potential risks stemming from an excessive accumulation of debt, analysing the total borrowings of debtors at the entity level and at different levels of consolidation. These “group structures” contain insights into the financial strengths of companies, especially in terms of potential contagion risks. The latter is a key metric that needs to be taken in to account in order to make the financial sector more resilient. The integration of multiple granular ESCB datasets (containing credit information) will thus help supervisors and other policymakers to monitor financial stability and conduct macroprudential analyses and quantitative risk assessments.⁴ The enriched information coming from the interconnected datasets will support stress tests conducted at both the entity level and the macro level in order to

¹ Coeuré, B., “[Setting standards for granular data](#)”, opening remarks at the Third OFR-ECB-Bank of England workshop on “Setting Global Standards for Granular Data: Sharing the Challenge”, Frankfurt am Main, March 2017.

² In this paper, the terms “granular data” and “microdata” are used interchangeably and refer to data on individual transactions, instruments and/or entities.

³ Carstens, A., “[Micro-data as a Key Input to Designing Macro-prudential Policy: The Mexican Experience](#)”, paper presented at the Eighth ECB Statistics Conference, Frankfurt am Main, 6 July 2016.

⁴ Heath, R.M. and Goksu, E.B., “[Financial Stability Analysis: What are the Data Needs?](#)”, IMF Working Paper, No 17/153, International Monetary Fund, 2017.

assess the vulnerability of the system as a whole.⁵ The additional information coming from the higher granularity will allow more flexibility during stress-test exercises for (i) models using granular data to tailor stress-test parameters and (ii) models using aggregated data to break down the impact of macro-level shocks.

A crucial outcome of the integration of datasets, which is especially significant for microprudential supervision, will be the assessment of creditworthiness of borrowers (e.g. “credit history”). In addition, supervisors will have the ability to assess the liquidity and solvency of supervised entities at both individual entity and consolidated group level. This will facilitate a multitude of uses in the supervisory process (both on-site and off-site), permit analyses not previously covered by any regular reporting and complement information from other reporting systems.⁶

Other ECB business areas which will benefit from the supply of granular data are the Directorate General Market Operations and the Directorate Risk Management, especially regarding the treatment of eligible collateral in ESCB credit operations. As mentioned by Ulrich Bindseil when he was Director General Market Operations: “Collateral availability is not an issue that can be understood at an aggregate level, as collateral scarcity is a phenomenon that will materialise at the individual bank level.”⁷ The monitoring of collateral at granular level is extremely important in order to understand the monetary policy transmission flow. This is because an absence of eligible collateral could block the monetary policy transmission channel; banks which run out of collateral would not be able to borrow money from the ECB, and consequently, would not be willing to provide new credit to the economy. The use of granular data would also be beneficial for monitoring compliance of individual banks with relevant collateral rules such as the “close link” prohibition and limits on the share of bank bonds issued by one issuer group in a collateral portfolio.

Finally, and importantly, central banking economic research will benefit from this new supply of granular data. Internal analysis and research are aimed at fine-tuning policy measures and their execution but may also find their way into publications, such as Economic Bulletin articles and ECB working papers, which will be enriched by the use of more granular and comprehensive information.

The above-mentioned analyses require the collection of various types of granular data on both entities and financial instruments (in particular loans and debt securities). In response to this policy need, the ECB and the euro area national central banks (NCBs) set up the Analytical Credit Datasets (AnaCredit) with detailed and timely information on bank loans in the euro area and some other EU countries. The information on loans in AnaCredit complements the Centralised Securities Database (CSDB) and the Securities Holdings Statistics Database (SHSDB), which are security-by-security datasets created to support the ECB’s decision-making processes.

⁵ Damia, V. and Israël, J.-M., “[Standardised granular credit and credit risk data](#)”, paper presented at the Seventh IFC Conference on “Indicators to support Monetary and Financial Stability Analysis: Data Sources and Statistical Methodologies”, Bank for International Settlements, 2014.

⁶ Damia, V. and Israël, J.-M., op. cit.

⁷ Bindseil, U., Corsi, M., Nicoloso, P., Rodríguez, F. and Segura, I. “[Micro data for monetary policy implementation – recent experience in the ECB](#)”, paper presented at the Eighth ECB Statistics Conference, Frankfurt am Main, 6 July 2016.

The power of granular data resides in being able to link all the pieces of the puzzle, i.e. calculating different aggregates in a flexible way. In order to have this flexibility, it is necessary to create a stable and unique link between the entities and the instruments issued or held by them. The ESCB's Register of Institutions and Affiliates Data (RIAD) is the central piece of the puzzle, it stores information on entity identifiers and is thus pivotal in ensuring the unique identification of counterparties. Moreover, it offers a representation of group structures at different levels of consolidation.

The rest of the paper is structured as follows: Section 2 explains in more detail why ESCB policymakers need more granular data and why it is important that these granular data can be linked. The boxes in Section 2 give a hint of the manifold uses of the information gained by combining such granular data. Section 3 looks at the crucial role of RIAD in linking a number of granular instrument-by-instrument ECB datasets. The boxes in Section 3 focus on group structure information in RIAD and how this information plays a crucial role in the ECB's collateral management. Section 4 then describes the actual work being done to link the granular databases, the challenges encountered and possible future developments. Section 5 concludes, providing an outlook on the way forward.

2 The importance of (integrated) microdata

2.1 Why microdata?

Given the increasingly complex challenges facing central banks today, spanning from monetary policy and the use of unconventional monetary policy tools to banking supervision and financial stability, decisions need to be data-driven. For this reason, statisticians are tasked with collecting, compiling and maintaining large datasets, as well as ensuring high data quality. Policymakers often seek answers that are not easily observable; indeed, the questions they pose can be answered from different angles and vary over time. In response to this policy need, statisticians have increasingly focused their activities on the collection of microdata.

Collecting and maintaining granular data has many advantages, all of which stem from their “granularity”. First, “granular” data can be collected in a timelier fashion, owing to the fact that no pre-aggregation is needed prior to their submission to the authorities, meaning that less transformation is needed by the reporting banks. This lowers the reporting burden on banks, which reduces reporting costs, as does the fact that granular reporting is stable once set up. Moreover, granular data give users more freedom to “play” with the information; for example, granular data can be used to disaggregate macrodata (analysing in detail what stands behind the aggregates) and can be aggregated in different ways, including across time (backwards), exponentially increasing the number of analyses that can be performed. It is evident that the possibility of interconnecting different types of granular data (e.g. reference data, balance sheet items and ownership relationships) goes beyond the mere aggregation of microdata to reach macro aggregates. This flexibility is, indeed, the main advantage of using granular data.

Traditionally, central bank policymakers have focused on aggregated, macroeconomic statistics in their analyses. One example is [ECB Convergence Reports](#), in which aggregated, macroeconomic statistics are used to assess the state of economic convergence in EU Member States seeking to adopt the euro. One such statistic is the government debt-to-GDP ratio of the countries under scrutiny. Both government debt and GDP are compiled according to internationally harmonised definitions and are part of an integrated set of interdependent macroeconomic statistics known as the national accounts.⁸ The debt-to-GDP ratios of EU Member States are checked and validated by Eurostat. However, this indicator only becomes available almost four months after the end of each year/quarter, and, in times of market turmoil and elevated economic uncertainty, policymakers need information that is timelier and more detailed than the debt-to-GDP ratio. For example, it is useful

⁸ See Regulation (EU) No 549/2013 of the European Parliament and of the Council of 21 May 2013 on the European system of national and regional accounts in the European Union ([OJ L 174, 26.6.2013, p. 1](#)) and Council Regulation (EC) No 479/2009 of 25 May 2009 on the application of the Protocol on the excessive deficit procedure annexed to the Treaty establishing the European Community (Codified version) ([OJ L 145, 10.6.2009, p. 1](#)).

to know which part of the total amount of outstanding government debt securities has to be refinanced in the next three months. Using the CSDB, it is possible to extract granular information with which to compute this amount (and other possible breakdowns) quickly after the end of each month. Moreover, the granularity of the data makes it possible to compute various aggregates that are of use to policymakers, such as short-term versus long-term issuances and nominal yields.⁹

The use of the CSDB for more information on government debt securities is just one example of how granular data can be used to provide timelier and more detailed information to policymakers to complement macro statistics. It became evident that not only microprudential analysts but also macroprudential analysts need granular microdata. To illustrate this, former ECB Executive Board member and former Vice-Chair of the Supervisory Board Sabine Lautenschläger gave the following example: “Assume that we see credit to businesses accelerating in a given country. We can think of several different developments underlying this ‘aggregate’ fact. It might be that solid companies, with low debt and good economic prospects, are taking out more or larger loans. Or it could be that fragile and highly indebted companies are borrowing more or are restructuring their debt just to survive. Credit might also be growing because more companies have access to bank financing. In turn, this might reflect better economic prospects and a greater appetite for investment – which is good – or just a deterioration of credit standards – which is not so good.”¹⁰ In short, granular data help identify the underlying developments that drive macroeconomic events and can thus help formulate an appropriate policy response.

As mentioned above, granular data also have the primary function of allowing the computation of non-conventional aggregates. Indeed, analysts have to perform increasingly complex analyses, trying to identify risks at an early stage and vulnerabilities coming from different directions. Macro aggregates do not always meet the research needs, and policymakers are often forced to rely on additional surveys that are both costly in terms of the ad hoc reporting burden and time-consuming (design and launch). The collection of granular data, on the other hand, offers much greater flexibility in responding to a new request by carrying out an appropriate aggregation of the data. Moreover, this approach will help save costs, because reporting agents do not have to fill out a new survey, leaving the aggregation burden to the central bank.

However, the maximum power is reached only when different types of microdata are brought together. For this reason, it is necessary to provide analysts with a flexible tool that breaks the silos. The first pillar is the creation of a single IT-platform where it is possible to access all the available data, even if those data are stored in different databases.

The existence of a single access point does not imply that the information is connected and hence exploitable. The second pillar is the creation of a unique link between the instrument-by-instrument, transaction-by-transaction and firm-level datasets. In this

⁹ For more detail, see Cornejo Pérez, A., Diz Dias, J. and Hartwig Lojsch, D., “[New and timely statistical indicators on government debt securities](#)”, *Statistics Paper Series*, No 8, ECB, June 2015.

¹⁰ Lautenschläger, S., “[Central bank statistics: moving beyond the aggregates](#)”, speech at the Eighth ECB Statistics Conference, Frankfurt am Main, 5 July 2016.

way, analysts have all the pieces of the puzzle in their hands. However, unlike a regular puzzle, there is no single solution; analysts can combine the information in different ways, tailoring the analysis according to their specific research needs.

Finally, once the data are linked, tools need to be developed to ensure that the data can indeed be combined in a meaningful way. As mentioned by Bank of Greece Governor Yannis Stournaras: “The point is that statisticians have their work cut out. We do not just ask them to collect the biggest amount of data and store them somewhere. We ask them to develop clever, versatile tools that can provide intelligible answers to policymakers’ queries.”¹¹ In this respect, much work has been devoted to the production of monitoring dashboards.

Box 1

Concentration risk

Concentration risk describes the level of risk in a bank’s portfolio arising from the concentration of exposures to a single counterparty, sector, currency or country. The risk arises from the observation that more concentrated portfolios are less diverse and therefore more risky because of the strong dependence on a particular asset or group of assets.

The simplest form of concentration risk is *single-name concentration risk* and arises in the absence of proper idiosyncratic risk diversification. A bank faces this risk when exposures to a single counterparty account for more than an infinitesimal share of the total portfolio,¹² meaning that there is still room for diversification and consequent risk reduction.

Another form of concentration risk is *sectoral concentration risk*, which occurs when investments are concentrated in asset classes with common features. This kind of risk arises from the fact that business conditions, and consequently default probabilities, are not fully aligned across the whole economy (e.g. the euro area); indeed, there may be areas (business areas or geographical areas) in which risks are higher. For instance, a bank with exposures that are not adequately diversified and therefore concentrated in one particular area is subject to sectoral concentration risk. The linked databases allow the performance of multiple sectoral concentration analyses with different counterparty breakdowns. For instance, a *size risk analysis* based on the size of the counterparties can be performed, or a *sectoral risk analysis* based on the activities¹³ in which they are engaged in order to identify potential risks arising from the cyclicity of a particular business activity. In addition, a *geographical risk analysis* (at different levels of granularity), exploiting location information,¹⁴ could be performed in order to measure potential *country risk*.

¹¹ Stournaras, Y., “[Micro data for monetary policy decisions: moving beyond and behind the aggregates](#)”, speech at the Eighth ECB Statistics Conference, Frankfurt am Main, 6 July 2016.

¹² Basel Committee on Banking Supervision, “[Studies on credit risk concentration: An overview of the issues and a synopsis of the results from the Research Task Force project](#)”, *BCBS Working Paper*, No 15, November 2016.

¹³ In accordance with the statistical classification of economic activities in the European Community (NACE) established by Regulation (EC) No 1893/2006 of the European Parliament and of the Council of 20 December 2006 establishing the statistical classification of economic activities NACE Revision 2 and amending Council Regulation (EEC) No 3037/90 as well as certain EC Regulations on specific statistical domains ([OJ L 393, 30.12.2006, p. 1](#)).

¹⁴ In accordance with the [Nomenclature of Territorial Units for Statistics \(NUTS\)](#).

Box 2

Specific wrong-way risk

Besides the above mentioned concentration risk, another important challenge for risk managers is the identification of *wrong-way risk*. While *general wrong-way risk (GWWR)* arises when the counterparty's likelihood of default is positively correlated with negative market events, *specific wrong-way risk (SWWR)* arises when future exposure to a specific counterparty is positively correlated with the counterparty's probability of default (PD) due to the nature of the transactions with the counterparty.¹⁵ A good example is bonds with credit-linked coupons which increase when the issuer's credit rating decreases. The exposure of the lender is then negatively correlated with the credit quality of the borrower (i.e. positively correlated with its PD).

SWWR may also arise through poorly structured transactions; for example, those collateralised by own or related party shares. Identifying SWWR in the case of collateralised transactions would help in the assessment of the quality of assets pledged as collateral by counterparties in market operations. The risk arises when the collateral pledged by the borrower is guaranteed by an entity within the same conglomerate as the borrower. SWWR exists because the posted collateral cannot be considered a strong credit risk mitigation tool. If a significant negative shock affected all entities within the conglomerate, the collateral power would be reduced. In extreme cases the security could be effectively uncollateralised. In other words, the exposure of the creditor to the borrower will increase as the PD of the latter increases owing to contagion¹⁶ within the conglomerate. The ability to identify such situations early on is an important diagnostic feature of any credit risk assessment system and requires the correct and unique identification of the issuers of securities. This is achieved through the full integration of datasets, with the unique identification of issuers of credit instruments and their allocation to the respective financial (or corporate) conglomerates. Up-to-date knowledge of banking group structures will serve to fill a number of information gaps and help policymakers to achieve higher quality standards.

2.2 Challenges associated with microdata

As may be expected, a centralised tool providing access to harmonised and integrated granular information does not come without costs and challenges. The most prominent are the handling of confidential data, the cost of developing and maintaining a complex IT infrastructure, the reporting burden and ensuring that the data collected are of high quality.

The major challenge in terms of the initial costs borne by the ESCB is setting up and maintaining statistical datasets that are sufficiently integrated. Such costs increase when the integration of separate silos is performed *ex post*. The parallel design of related statistical reporting requests ensures that unnecessary overlaps between the datasets are avoided.

¹⁵ See Article 291(1) of the Capital Requirements Regulation (CRR) – Regulation (EU) No 575/2013 of the European Parliament and of the Council of 26 June 2013 on prudential requirements for credit institutions and investment firms and amending Regulation (EU) No 648/2012 (OJ L 176, 27.6.2013, p. 1).

¹⁶ Default contagion is “the probability of an obligor's default conditional on another obligor defaulting being higher than the unconditional probability of default for the same obligor”. See Basel Committee on Banking Supervision, *op. cit.*, p. 20.

Data quality assurance (DQA) is a key goal for statisticians in order to make a positive contribution to the ESCB's decision-making process. However, DQA is not only performed ex post; it is a joint process that involves the financial industry, which provides the data, the NCBs, which report the data, and the ECB, which collects the data and performs a number of quality checks. It is therefore beneficial to improve the dialogue with financial industry (feedback loops), policymakers and academia in order to tailor reporting requirements and to make the best possible use of the data.¹⁷ With these considerations in mind, the ESCB has started a dialogue with the banking industry on ESCB statistics.¹⁸

The use of granular data almost always raises confidentiality issues. Information contained in company-level data is very often market-sensitive and is therefore classified as "confidential". The goal of central banks should be to develop best practices and adequate safeguards in order to exploit such information without breaching confidentiality. Possible solutions to allow researchers access to the data include the anonymisation of granular data or allowing external researchers to work temporarily at the ECB subject to the same confidentiality regime as ECB staff members.¹⁹

¹⁷ See Buch, C., "[Focus on micro data: Potential benefits for the industry?](#)", paper presented at the Eighth ECB Statistics Conference, Frankfurt am Main, 6 July 2016.

¹⁸ See "[Banking Industry Dialogue on ESCB statistics](#)" on the ECB's website.

¹⁹ See Bindseil et al., op. cit.

3 ESCB microdata

There are several granular statistical databases within the ESCB that contain information on different types of credit: the CSDB, the SHSDB and AnaCredit. As the repository of master data, RIAD is the linchpin of the integration process, allowing the unique identification of counterparties. This section describes the crucial role played by RIAD and the datasets containing microdata on credit.

3.1 The crucial role of RIAD

The ECB's Governing Council, advised by the ESCB Statistics Committee, decided in June 2016 to have a unique repository for reference data within the ESCB. It therefore agreed to extend RIAD – which previously only held data on entities in the financial sector – to make it a shared platform for the collection, (joint) management and provision of counterparty reference data for all ESCB granular datasets, including on non-financial corporations. As well as reference data on legal and other institutional units relevant for statistical and other business processes in the ESCB, RIAD also contains reference data used by supervisors in the Single Supervisory Mechanism (SSM).

RIAD is constructed as a shared dataset where NCBs and national competent authorities (NCAs) are responsible for data on entities resident in their own jurisdiction, but may also provide information on non-resident entities.

The reference data in RIAD serve four purposes:²⁰

- **Identification**

RIAD ensures the unique identification of counterparties via a large number of identifiers, including the Legal Entity Identifier (LEI)²¹ and specific national identifiers (e.g. tax codes, business register numbers). In addition, it contains information on the official names and addresses of entities. The unique identification of entities is crucial in the linking of different databases. At least one common identifier available in RIAD should be reported to the other systems to allow them be linked. This identifier should be unique and stable over time.²²

- **Stratification**

RIAD holds information on the economic activity classification (NACE), statistical sector classification (European System of Accounts, ESA 2010) and geographical location (NUTS) of entities. This information allows the population to be broken down

²⁰ In line with United Nations Economic Commission for Europe (UNECE) [Guidelines on Statistical Business Registers](#).

²¹ See “[Introducing the Legal Entity Identifier \(LEI\)](#)”, Global Legal Entity Identifier Foundation (GLEIF).

²² See Guideline (EU) 2018/876 of the European Central Bank of 1 June 2018 on the Register of Institutions and Affiliates Data (ECB/2018/16) ([OJ L 154, 18.6.2018, p. 3](#)).

into different layers and multiple analyses to be run for different purposes. The statistical sector classification is, for example, of relevance for the various lists of financial entities published by the ECB.²³ For credit risk analyses, these breakdowns can be used to assess exposures of or to certain sectors or industries.

- **Demographic developments**

The initial date of an entity's registration in the national business register ("birth date") is collected in RIAD, as well as the official de-registration date ("closure date"). Information is also collected on the business status of an entity (e.g. active, inactive) or whether the entity is under liquidation. In addition, NCBs report information on relevant corporate events, such as splits, mergers or changes of residency. A user can thus trace an issuer or holder of securities across time and corporate events, which may be relevant when assessing credit risks.

- **Relationships between entities**

There can be various types of relationships between entities: control, ownership, management, supervision, origination, branch, etc. Relationships are used to derive group structures in RIAD, as illustrated in Box 3. A particular type of relationship, "close links", is discussed in Box 4.

Box 3

Group structures in RIAD

The group structures in RIAD are derived from the bilateral relationships between entities in RIAD and are not separately reported. This approach provides flexibility to users who are interested in multiple analyses, looking at different interlinkages among institutions. Group structures can be defined on the basis of control, ownership or other links (e.g. links among supervised entities).

In RIAD, the group structures based on "control"²⁴ are determined as follows: first **the "head" of the group**²⁵ **is identified and** then the entities it controls. An entity is deemed to be a controlling parent if any of the following three conditions is met:

- **explicit direct control** over the subsidiary;
- **ownership of more than 50%** of the subsidiary's capital;
- **indirect control** over the subsidiary through two or more controlled subsidiaries.²⁶

The identified group reflects the financial accounting scope of consolidation as identified by IFRS 10. This group can then be enriched with information encompassing all equity investments of the group members, i.e. a group based on all non-controlling ownership relationships, as shown in Figure A.

²³ See "[Lists of financial institutions](#)" on the ECB's website.

²⁴ For a definition of "control", see Article 4(1)(37) of Regulation (EU) No 575/2013.

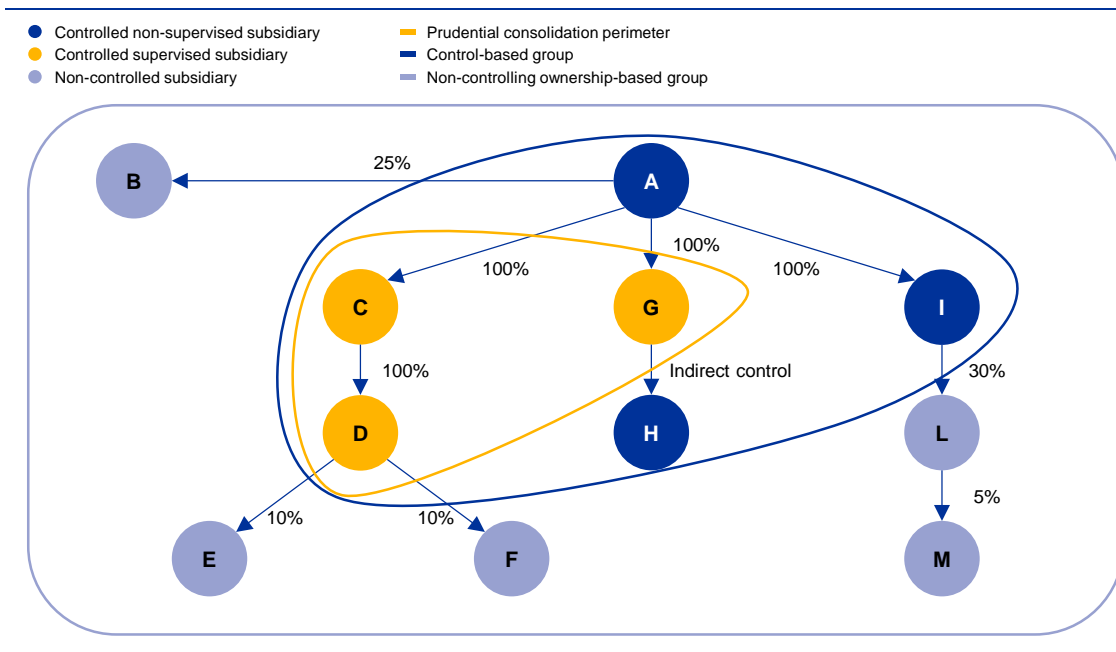
²⁵ The "head of the group" is an entity which is not itself subject to a control relationship but controls one or more other entities.

²⁶ Indirect control exists, for instance, when entity A and entity B each own 30% of entity C (owning in total 60% of the shares of C) and entity X controls both entity A and entity B. In this situation, it is possible to say that entity X indirectly controls entity C through its subsidiaries A and B.

Such an enriched group structure is of interest to, for example, ESCB policymakers working in market operations, risk management or macroprudential supervision.

Figure A

Group structures in RIAD



RIAD also contains group structures based on links between supervised entities within the “prudential consolidation perimeter” (i.e. the scope of prudential consolidation according to the CRR). The links (or absence thereof) between supervised entities determines the reporting requirements both of the ultimate supervised head of a group and of its supervised subsidiaries, liquidity sub-groups and stand-alone entities. Not all entities that are part of the same group within the financial accounting perimeter are supervised entities within the prudential consolidation perimeter. For example, an insurance undertaking (entity A in Figure A above), an industrial company or insurance-led mixed financial holding that controls a lower-level credit institution will not be considered the ultimate supervised head of the group. Owing to RIAD’s flexibility it will, however, be possible to compute groups according to both the financial accounting and the prudential consolidation perimeter and to compare overlaps and differences between the two group structures.

Information on group structures can be extremely useful in combination with data stored in AnaCredit. For instance, supervisors may be interested in calculating the non-performing loan (NPL) ratio²⁷ of significant institutions. Each bank in AnaCredit is identified by a single identifier (RIAD code) and all instruments issued by a bank are linked via this identifier. Moreover, loans that are classified as “non-performing” are flagged and it is easy to calculate an NPL ratio for a single bank. Supervisors may also be interested in understanding how this ratio changes if the same calculation is performed at group level and how it changes if the perimeter of consolidation is changed.

A type of relationship which RIAD does not have information on is “groups of connected clients”. This could be, for instance, a group of companies involved in the oil refining process and connected by

²⁷ Non-performing loans as a percentage of total loans.

business relationships because they are part of the same supply chain (e.g. upstream company, refining company and downstream company). A negative shock in the supply of crude oil could be transmitted to other entities in the supply chain. Situations like this, where economic (or business) dependency creates interlinkages among entities, are generally defined as a group of connected clients. Given its usefulness for risk analysis, it would be beneficial to have this type of relationship recorded in RIAD.

Box 4

RIAD as a pivotal tool for monitoring compliance with collateral management rules

When seeking credit from the ESCB, the financial industry faces many collateral requirements detailed in various “frameworks”.²⁸ The main requirements concern the type of asset, credit standards, place of issue, type of issuer, currency of issue and country of residence of issuer. Overall, the eligibility of assets is assessed by NCBs according to the criteria specified in the Eurosystem’s legal framework for monetary policy instruments.

Nevertheless, there are additional restrictions on the use of collateral aimed at limiting concentration risk, and the ESCB monitors compliance with these restrictions using the information contained in RIAD.

Close links

Close links are special relationships between issuers and holders of assets. The identification of these relationships plays a crucial role in the Eurosystem’s collateral management, in which only “eligible assets” are accepted as collateral.²⁹ Assets held by counterparties and issued by entities to which the counterparties have close links are not deemed eligible by the Eurosystem. In particular, close links are deemed to exist in any of the following situations:³⁰

(a) the counterparty owns directly, or indirectly through one or more other undertakings, 20% or more of the capital of that other entity;

(b) that other entity owns directly, or indirectly through one or more other undertakings, 20% or more of the capital of the counterparty;

(c) a third party owns, either directly or indirectly through one or more undertakings, 20% or more of the capital of the counterparty and 20% or more of the capital of the other entity.

This rule on close links is a risk control measure, since in the case of close links the credit quality of the counterparty is likely to exhibit a high correlation with the credit quality of the collateral submitted by the counterparty.³¹ To determine whether a close link exists, it is crucial to exploit the information on relationships between entities maintained in RIAD. A dedicated algorithm navigates the bilateral links between entities and calculates close links. These are flagged and brought to the attention of the ECB’s Directorate General Market Operations. This is one reason why RIAD is essential for the Eurosystem’s collateral management.

²⁸ See “[Collateral eligibility requirements – a comparative study across specific frameworks](#)”, ECB, July 2013.

²⁹ For more information on collateral, see [the ECB’s website](#).

³⁰ See Guideline (EU) 2015/510 of the European Central Bank of 19 December 2014 on the implementation of the Eurosystem monetary policy framework (ECB/2014/60) ([OJ L 91, 2.4.2015, p. 3](#)).

³¹ For more information on risk control, see [the ECB’s website](#).

Unsecured credit institution bonds limitation

In addition to the close link monitoring, the integration of microdata is beneficial for checking compliance with another important collateral rule: “A counterparty shall not submit or use as collateral unsecured debt instruments issued by a credit institution or by any other entity with which that credit institution has close links to an extent that exceeds 5% of the total value of the assets used as collateral by that counterparty after the applicable haircut. This 5% threshold shall not apply in either of the following cases:

- (a) if the value of such assets does not exceed EUR 50 million after any applicable haircut; or
- (b) if such assets are guaranteed by a public sector entity which has the right to levy taxes by way of a guarantee that complies with the features laid down in Article 114.”³²

In this case, the identification of credit institutions (through the ESA sector), the identification of close links and overall ownership relationships are extracted from RIAD. Reference data on the entities, once adequately linked to securities data (i.e. collateral value) stored in the CSDB, offers the full set of information for monitoring compliance with the above-mentioned rule.

3.2 Microdata on credit

This section briefly describes three of the granular ECB datasets containing instrument-by-instrument data, namely the CSDB, the SHSDB and AnaCredit. The list of granular datasets maintained at the ECB is, however, not limited to the ones mentioned.³³ The focus here is on the datasets containing microdata on credit (loans and debt securities), excluding credit granted on the money market.

The CSDB, the SHSDB and AnaCredit are regarded by the ESCB Statistics Committee as the first granular datasets to be fully integrated through the unique identification of counterparties offered by RIAD.

The CSDB³⁴

The financial crisis increased the relevance of granular security-by-security data as a means of ensuring better microeconomic analysis of financial markets and credit risks. The granular, security-by-security data in the CSDB provide a wide range of users within the ESCB³⁵ with complete, accurate, consistent and up-to-date information on individual securities, including:

- *securities issued by EU residents;*
- *securities likely to be held and transacted in by EU residents;*

³² See Article 141 of Guideline (EU) 2015/510 of the European Central Bank of 19 December 2014 on the implementation of the Eurosystem monetary policy framework (ECB/2014/60) (OJ L 91, 2.4.2015, p. 3).

³³ Not discussed in more detail here are the EMIR database and the MMSR dataset.

³⁴ See Cornejo Pérez, Diz Dias and Hartwig Lojsch, op. cit.

³⁵ A number of statistical indicators on government debt securities derived from the CSDB are available to the general public in [the ECB's Statistical Data Warehouse](#).

- *securities denominated in euro, regardless of the residency of the issuer or holder.*

The CSDB currently contains information on over seven million non-matured or “alive” debt securities, equities and mutual fund shares/units, plus approximately nine million matured or “non-alive” securities. It contains reference data on securities (e.g. outstanding amounts, issue and maturity dates, coupon and dividend information, statistical classifications), issuers and prices (market, estimated or defaulted) as well as more recently introduced information on ratings (of the security, issuer, guarantor or issuance programme).

The SHSDB³⁶

During the financial crisis, policymakers became aware that they did not have enough information on which holders of securities would be affected if the issuer of those securities were to go bankrupt. In response, the ESCB decided to collect information on securities holdings (both equity and debt securities).

The information on holdings can be divided into two categories:

- *securities held by institutional sectors³⁷*, which includes:
 - holdings of the institutional sectors in the euro area countries (e.g. households, non-financial corporations);
 - holdings of non-euro area investors deposited with euro area custodians;
 - holdings of non-euro area EU investors collected by non-euro area EU countries.
- *securities held by banking groups*: since September 2018 the holdings of all banking groups directly supervised by the ECB (significant institutions) have to be reported to the ECB.

AnaCredit³⁸

AnaCredit includes detailed information on credit and credit risk on an instrument-by-instrument basis. The following instruments (most of them types of loans) are covered by AnaCredit: deposits other than reverse repurchase agreements; overdrafts; credit card debt; revolving credit other than overdrafts and credit card debt; credit lines other than revolving credit; reverse repurchase agreements; trade receivables; financial leases; and other loans. AnaCredit covers credit granted by euro area credit institutions (including their branches outside the euro area) and by the euro

³⁶ For more information on securities holdings statistics, see [the ECB's website](#).

³⁷ The ECB publishes statistics on holdings of euro area sectors in its [Statistical Data Warehouse](#).

³⁸ For more information, see the [explanatory note](#) on the AnaCredit Regulation on the ECB's website.

area branches of non-euro area credit institutions to corporations and other legal entities.

The information collected comprises, inter alia, the type of credit extended, outstanding amounts, original and residual maturity, the interest rate and the currency of the instrument. In addition, information on any credit protection (i.e. guarantees and collateral) is also collected. Moreover, with a view to enabling the reliable identification of all debtors, the collection includes reference data for every counterparty related directly or indirectly to the instruments via RIAD.

Data are submitted to the ECB every month and every quarter (for monthly and quarterly reporting, respectively) pursuant to the reporting deadlines. The first AnaCredit reporting concerned data relating to September 2018.

4 Linking the microdata

Once AnaCredit, the CSDB and the SHSDB can be linked via RIAD, it will be possible to compute the total exposure of creditors (lenders/holders) to debtors (borrowers/issuers) and the total borrowings of debtors at entity level. In addition, these microdata will be able to be used to compute exposures at a more aggregated level (e.g. groups, sectors, geographical areas).

The benefits of this full integration are many. It will support, among other things, monetary analysis, monetary policy decision-making, the early detection of systemic risks and the conduct of macroprudential policies and microprudential supervision.

Box 5

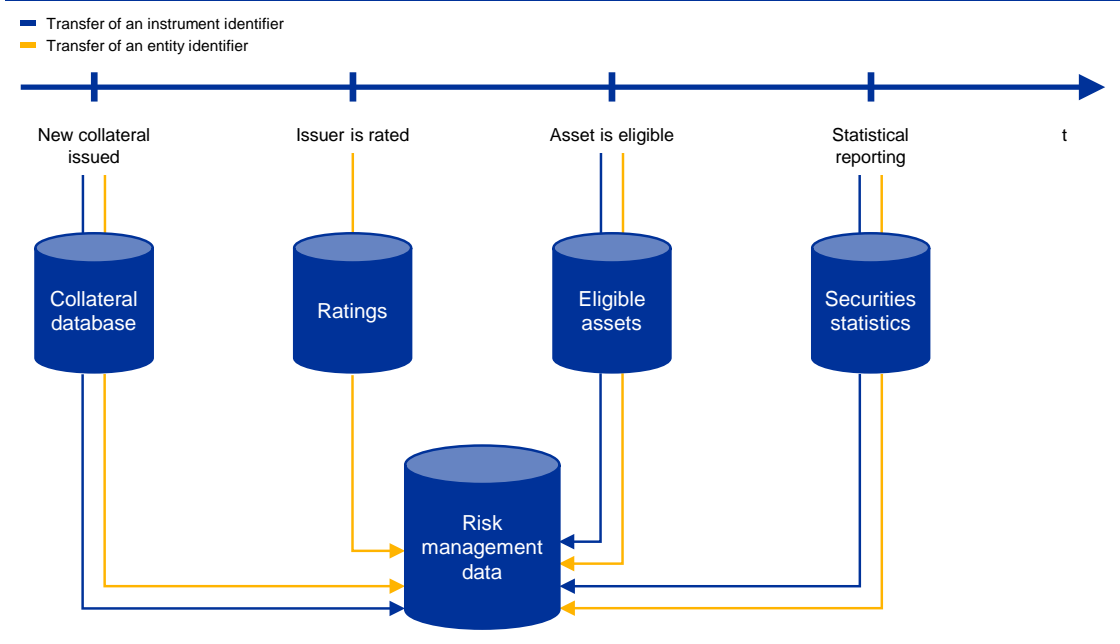
Benefit of linking the microdata: unique identification of entities

As highlighted in the other boxes, the interconnection of granular ESCB datasets will open the way for multiple uses of the data, especially exploring different perimeters of consolidation. The key element of this integration is, however, the unique identification of entities and instruments, which will avoid duplication of work and help streamline the day-to-day business of the central bank. The following example illustrates the importance of unique identification in the context of collateral management under the asset purchase programme (APP).

Example: a team of experts in risk management collects asset and issuer-specific information from various risk assessment and reporting datasets as soon as the information becomes available. For instance, when a new asset with an unknown issuer becomes available as collateral for Eurosystem credit operations, the relevant national experts register it in a dedicated dataset with an instrument identifier. This identifier is database-specific, meaning that it does not allow communication with other datasets (separate silos approach). At the same time, the ECB becomes aware of the issuer, since it is rated by at least one of the four eligible rating agencies. An entity identifier, provided by the rating agency is assigned to the issuer and attached to the asset. Consequently, the asset is recognised as “eligible” for purchase under the APP and is therefore registered in another specific dataset with a second instrument identifier. Finally, the risk management team receives monthly information on newly issued securities collected for statistical purposes (CSDB). The same instrument will have its own specific identifier as well as a new entity identifier related to the issuer. The example is illustrated in Figure A.

Figure A

Collection of microdata from multiple sources without the unique identification of counterparties



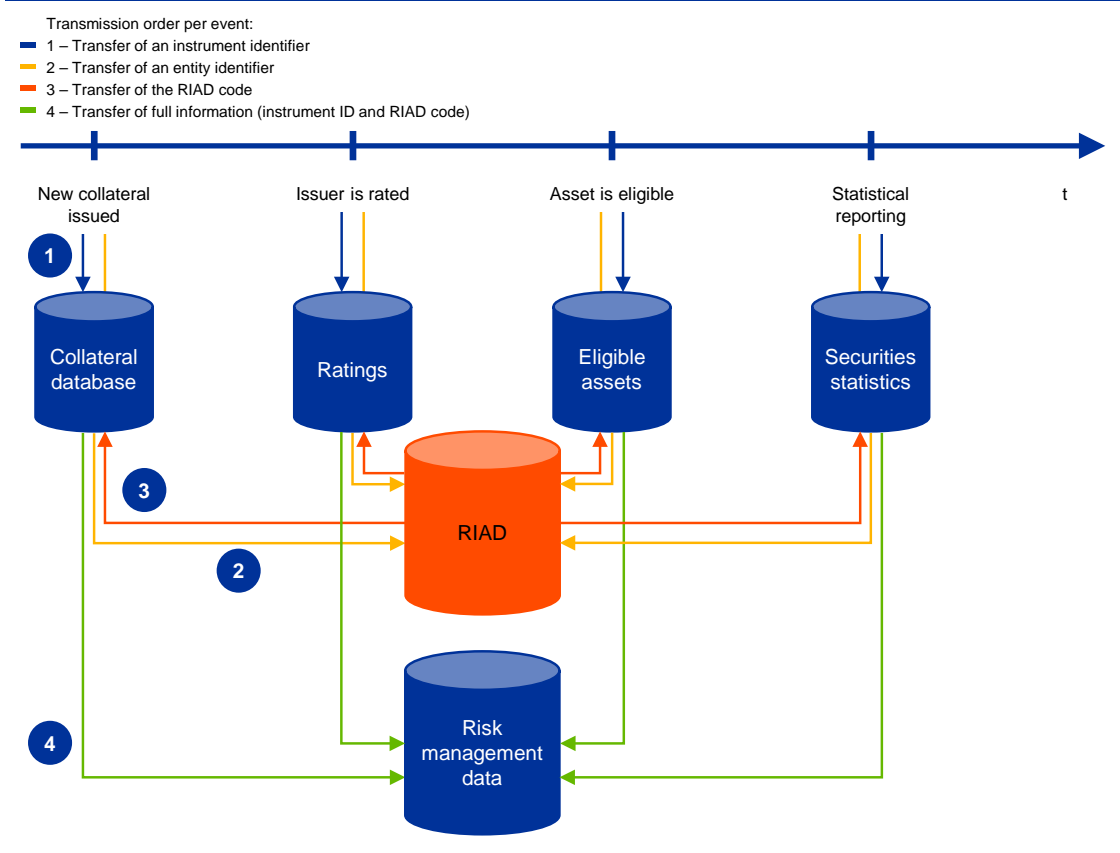
In the absence of a fully integrated system, there will be up to three different instrument identifiers and up to seven issuer identifiers (each rating agency may provide its own specific entity identifier).

The unique identification of the instrument is easier than the identification of the entity, owing to the common practice of reporting the International Securities Identification Number (ISIN) as the instrument identifier. The unique identification of entities is a much more complex task, because such identifiers are not standardised. In a situation like the one shown in Figure A, the recognition of the issuer is complicated, which creates confusion and amplifies operational risk. It is evident that there is a clear need for the unique management of institutions' reference data through a centralised repository such as RIAD.

Fortunately, RIAD is designed to manage multiple entity identifiers in an efficient way, and is able to connect different databases and enable them to interact with each other. As shown in Figure B, the risk management team will receive a complete package of information containing the instrument identifier and the RIAD code. The latter is the only entity identifier received by the risk management database, as the reconciliation is done externally in RIAD. The universal key allows all the (instrument) information stored in the relevant datasets to be accessed, recognising that there is only one instrument issued by one entity.

Figure B

Collection of microdata from multiple sources exploiting the unique identification of counterparties



As indicated above in Section 3, at least one common identifier available in RIAD should be reported to the other systems to allow them to be linked. This identifier should be unique and stable over time. The idea of a common, unique and stable identifier sounds very simple, but the actual implementation requires the coordination of experts in multiple teams in the NCBs and the ECB. They have to ensure that the instruments in the CSDB, the SHSDB and AnaCredit are uniquely linked to the corresponding counterparties. They also have to ensure that RIAD contains all the counterparties relevant for the CSDB, the SHSDB and AnaCredit.

What looks very easy in principle has, in practice, some hidden obstacles that complicate the achievement of the final result.

4.1 Challenges

The datasets described in Section 3 were not all designed and implemented at the same time. They were each created for different and/or ad hoc purposes, and only later was it decided to maximise the benefits of the granular information by synchronising them like the parts of a Swiss watch. However, there are some “legacy issues” arising from the previous non-synchronisation.

RIAD was designed as the unique master data repository for information on loans, and is thus fully integrated with the AnaCredit dataset. In other words, all the counterparties involved in credit transactions (borrowers, lenders and guarantors) have their reference data registered in RIAD. Indeed, the reporting of reference data via RIAD is part of the AnaCredit reporting process. However, RIAD only became the master data repository for entities and group structures in the CSDB and the SHSDB after both systems were already up and running.

A lot of work was required to make sure the systems are synchronised. First, for each country, the ECB and the NCB had to agree on which common identifier could be used to link the granular datasets. Moreover, the NCBs had to make sure that RIAD was populated with all the counterparties from the CSDB and the SHSDB. For this purpose, the ECB issued a legal act addressed to the NCBs.³⁹ Furthermore, the entity identifiers needed in order to link the granular databases have to be reported to the CSDB and the SHSDB.

Another complication arises from the use of data from commercial data providers and the related limitations on accessing and sharing such data within the ESCB. While RIAD and AnaCredit are populated with data sourced from the NCAs and NCBs and data in the SHSDB are sourced either directly from the reporting banks or from the NCBs, the data sources for the CSDB are more diverse. The CSDB is partially populated with data from the NCAs and NCBs, but data are also collected from commercial data providers. For contractual reasons, information from commercial data providers cannot be freely shared within the ESCB in order to fully align the databases.

4.2 The ESCB approach

As part of the linking process, a number of preconditions were highlighted to ensure the successful integration of the systems. These included: (i) the use of a unique and stable common identifier, (ii) data provision synchronisation at national and European level, and (iii) the definition of clear workflows and a clear division of responsibilities among stakeholders with strong permanent collaboration and teamwork among all relevant teams. Under this approach, the role of RIAD is crucial. The master data repository is pivotal to ensuring the unique identification of counterparties, hosting multiple identifiers that enable instruments to be matched to the related entities.

In order to be able to link the different datasets, it is necessary, first, that all the relevant entities are registered in RIAD. Once correctly registered, they need to receive at least one identifier that allows matching between RIAD and the other granular databases. This identifier could belong to one of three categories:

- an agreed country-specific identifier (e.g. national identifier);
- a Legal Entity Identifier (LEI);

³⁹ Guideline (EU) 2018/876 of the European Central Bank of 1 June 2018 on the Register of Institutions and Affiliates Data (ECB/2018/16) ([OJ L 154, 18.6.2018, p. 3](#)).

- an International Securities Identification Number (ISIN).

When the datasets are fully linked, all the reference data on debtors and/or creditors are retrieved from RIAD, which is seen as the most up-to-date and reliable source for the ESCB. This in turn reduces the reporting burden on NCBs in the medium-term by eliminating the need to report certain master data attributes (such as sector classification) to multiple systems in parallel.

Following the implementation of several steps to link RIAD with the CSDB and the SHSDB,⁴⁰ a data gap analysis shows a big improvement in the matching of issuers and holders (in the CSDB and the SHSDB) and RIAD entities. This matching takes place using the agreed country-specific identifiers and LEI and ISIN codes. The matching coverage has increased from 32% to 78%⁴¹ in terms of entities (count) and from 59% to 93% in terms of amount outstanding and market capitalisation of outstanding instruments⁴² (AO/MC) in the CSDB. While the count provides an indication of the number of problematic (entity) records in the databases, the AO/MC indicator is volume-weighted and indicates the potential significance of the problem. Particularly marked increases can be seen for some countries where the coverage is close to 100% of AO/MC, although not always yet in terms of number of entities. For other countries, however, the coverage is not so high, even after performing the initial agreed actions.

In addition to the coverage indicators, the data gap analysis also includes a consistency check of the sector classification of all entities that have been matched. The results show that for all matched entities there is already consistent sector classification for more than 94% in terms of entities and more than 96% in terms of AO/MC for alive instruments in the CSDB.

4.3 Limitations and way forward

The ECB will define in cooperation with the NCBs concerned the follow-up actions needed to further close the gaps. The cases where no match was found may mainly be due to legacy issues as described in Section 4.1. Differences in the timeliness of NCB data provision to the CSDB and RIAD can also create (temporary) gaps that are resolved over time (e.g. the entity may already be in the CSDB but not yet sent to the RIAD at the time of the data extraction). In view of this, reporting procedures and reporting timeliness may need to be further examined.

The linking exercise has been greatly complicated by the absence of a global common linking identifier available for all the relevant entities. Business registers, through the use of identifiers, have the main objective of ensuring the unique identification of entities (i.e. avoiding duplication). However, this is not the only goal of a business register; it should also allow (i) the “company history” to be tracked, (ii) the data to be

⁴⁰ AnaCredit is already fully integrated with RIAD.

⁴¹ Taking into consideration only those countries that have performed all required actions. The previous data gap analysis was performed before any action was taken.

⁴² Market capitalisation is the market value of the outstanding instruments, calculated as the price per instrument times the number of instruments outstanding.

aggregated at different levels of consolidation, and (iii) granular information stored in other datasets to be linked. These goals can be achieved through the use of an identifier that potentially satisfies the following characteristics: uniqueness, stability, coverage and “no proprietorship”. The identification of entities, however, is not always an easy task owing to the dynamic life of entrepreneurial business. Corporate events (e.g. mergers, acquisitions, splits and changes of residency) present a challenge for statisticians attempting to track a company’s history. In principle, in the case of mergers and acquisitions, only the identifier of the acquiring company should survive, while the target company identifier should be “closed”⁴³ and only connected to the new one to maintain a link. However, in practical terms it is not always easy to identify which is the acquiring company, and reporting agents often end up creating new identifiers. The concept of a stable and unique identifier should be also extended to group structures. Companies joining and leaving a group should not imply a change in the group identifier if the main group identity does not change. For instance, if a new group holding company acquires the head of the group, this should not imply a change in the group identifier if the main stakeholders behind the group have not changed. Stability and uniqueness are core characteristics to allow the evolution of companies and groups to be tracked over time, and they should be maintained through all corporate events to the maximum extent possible. A current limitation is the absence of a group identifier that allows analyses over time for group structures relevant for ESCB policy issues.

An LEI is, in principle, unique (meaning that the code cannot be reused when an entity ceases to exist), stable (meaning that it cannot be changed over the entire lifetime of the entity) and global (no national proprietorship). Nevertheless, the LEI is still far from being the perfect identifier, because its coverage is very limited compared to the number of legal entities registered world-wide. Even though a high level of coverage has not been reached yet, the use of LEIs has increased dramatically over the years, and so far more than 1.4 million LEIs have been issued by the Global Legal Entity Identifier Foundation (GLEIF).

In the light of this steady growth, an increase in the use of LEIs in all the ESCB datasets, particularly in RIAD,⁴⁴ is envisaged in order to close the linking gap mentioned in Section 4.2. Under current plans, for most, but not all, EU countries, granular databases will be linked using the RIAD code, which is an internal ESCB identifier. This has the drawback that it is more difficult to compare granular ESCB data with granular data from commercial data providers, since the latter do not use an internal ECSB identifier. If the coverage of the LEI were universal, the LEI could be used in the medium-term as the sole identifier to link the granular databases for all countries, since it satisfies the main conditions of a good identifier: uniqueness, stability, coverage and being global. The use of an identifier with such characteristics is helpful in (i) ensuring the unique identification of counterparties, (ii) tracking company histories, (iii) aggregating data at different levels of consolidation, and (iv) linking granular information stored in other datasets.

⁴³ “Closing” an identifier, or any data property in general, means closing the validity range of that property.

⁴⁴ At the time of writing, more than 400,000 entities already have an LEI registered in RIAD.

On the last point, increased use of LEIs would also help create a bridge between credit/debt exposures and derivatives transactions (registered in the EMIR database⁴⁵) and transactions on the money market (registered in the MMSR database⁴⁶). All the counterparties relevant for EMIR and MMSR data collection already have an LEI. Increasing the global coverage of LEIs in RIAD would allow a high level of data standardisation, which would help users to link, analyse and compare different data sources in a fast and possibly automated way.

So far no granular information is collected on the financial assets or debts of households. Indeed, RIAD does not include information on households or natural persons. One reason for this is that the processing and storage of information on natural persons is subject to much stricter confidentiality and privacy rules than information on corporations. Moreover, information on households is not needed to link AnaCredit, the CSDB and the SHSDB, since these databases do not contain information on credit extended to households. However, this does not mean that the ECB does not collect any information on households. The ECB Household Finance and Consumption Survey is aimed at collecting micro-level structural information on euro area households' assets and liabilities, as well as understanding the rationale behind economic decision taken by them. The real estate component is critical in the financial cycle,⁴⁷ and it is closely monitored by macroprudential policymakers in order to avoid potential asset bubbles. For this reason, in addition to capital requirements like the capital conservation buffer (CCoB) and the countercyclical capital buffer (CCyB), which are also aimed at limiting the banking credit supply, national authorities can implement borrower-side tools in order to influence demand for credit, such as loan-to-value (LTV) and debt-service-to-income (DSTI) ratios. The collection of more frequent data on households is paramount in order to provide input for macroprudential models and analyses⁴⁸ aimed at assessing the impact of such borrower-based measures. Moreover, the data will shed more light on the financial strength of households and will help in the derivation of credit risk parameters for households.

⁴⁵ The EMIR database is an ESCB database containing detailed information on derivatives transactions. The data are collected under the European Market Infrastructure Regulation (EMIR).

⁴⁶ The money market statistical reporting (MMSR) dataset is based on transaction-by-transaction data from the 52 largest banks in the euro area in terms of balance sheet size. The data cover transactions on the secured, unsecured, foreign exchange swap and euro overnight index swap money market segments.

⁴⁷ See Constâncio, V., "[Principles of Macroprudential Policy](#)", speech at the ECB-IMF Conference on Macroprudential Policy, Frankfurt am Main, 26 April 2016.

⁴⁸ Gross, M. and Garcia, F.J.P., "[Assessing the efficacy of borrower-based macroprudential policy using an integrated micro-macro model for European households](#)", *Working Paper Series*, No 1881, ECB, 2016.

5 Conclusion

Once full integration of the granular credit datasets is achieved through RIAD, policymakers will have a powerful tool at their disposal to compute exposures and indebtedness at entity and group level. The unique identification of both the borrower/issuer and the lender/holder will allow different risk analyses to be performed, such as the identification of excessive concentrations of investments in asset classes with common features, like business activity, country of risk or company size. Moreover, the unique identification of counterparties is crucial for collateral management. The identification of specific wrong-way risk in the case of collateralised transactions, owing to the excessive use of “own funds”, would benefit the assessment of the quality of the assets pledged as collateral by counterparties in market operations. However, the types of analyses that can be performed are not limited to the ones mentioned. Thanks to the granular nature of the data, users will be able to aggregate the granular information as needed. Many actors within the ESCB are waiting for such a tool to help them tackle challenges by exploiting the power of granular data. In this respect, new challenges could be identified and new reporting requirements may be set in order to collect relevant information for central banking and supervisory decision-making processes.

Interest in the various group structures (financial accounting perimeter, prudential consolidation perimeter and non-controlling ownership) is already increasing, and more demands are being made in terms of data coverage and frequency. New reporting requirements may at first glance look like an increase in the reporting burden and costs. To some extent this may be true, especially at the beginning when the whole reporting process is being established. However, the centralised collection of groups’ reference data will help to reduce the time needed to search and retrieve information, as it will all be available in RIAD, the ESCB’s unique, centralised and continuously updated master data platform. In view of its usefulness, in particular for banking supervision, a possible future expansion of the reporting requirements in RIAD to also cover groups of “connected clients” would be welcome. Authorities streamlining data requests so as to reduce the reporting burden will have major benefits. In particular, reporting master data to a unique, centralised platform will avoid duplication of work and reduce inconsistencies among different datasets.

Granular databases require powerful IT tools to efficiently handle the huge amount of data received, stored and analysed. Specifically, there is a demand for the easy and dynamic visualisation of group structures and the overlaps and differences between the multiple group structures that RIAD can generate. It is also important that IT tools are used to further develop data quality checks to verify not only the consistency but also the plausibility of the data provided. Efforts are already under way to develop such data quality checks within the ESCB.

Finally, there is a need to increase the use of LEIs to allow a high level of data standardisation, thereby helping users to link, analyse and compare different data sources in a fast, automated way.

Abbreviations

AO/MC	amount outstanding and market capitalisation	IFRS 10	International Financial Reporting Standard 10
APP	asset purchase programme	IMF	International Monetary Fund
BIS	Bank for International Settlements	ISIN	International Securities Identification Number
CCoB	capital conservation buffer	LEI	Legal Entity Identifier
CCyB	countercyclical capital buffer	LTV	loan-to-value ratio
CRR	Capital Requirements Regulation	MMSR	money market statistical reporting
CSDB	Centralised Securities Database	NACE	statistical classification of economic activities in the European Community
DSTI	debt-service-to-income ratio	NCA	national competent authority
DQA	data quality assurance	NCB	national central bank
EBA	European Banking Authority	NPL	non-performing loan
ECB	European Central Bank	NUTS	Nomenclature of Territorial Units for Statistics
EMIR	European Market Infrastructure Regulation	PD	probability of default
ESA 2010	European System of Accounts 2010	RIAD	Register of Institutions and Affiliates Data
ESCB	European System of Central Banks	SHSDB	Securities Holdings Statistics Database
EU	European Union	SSM	Single Supervisory Mechanism
GDP	gross domestic product	SWWR	specific wrong-way risk
GLEIF	Global Legal Entity Identifier Foundation		
GWWR	general wrong-way risk		

Acknowledgements

We would like to thank Jean-Marc Israël, Romana Peronaci, Henning Dräger and Jorge Diz Dias for their comments. The paper has also benefited from feedback and suggestions from the Editorial Board of the ECB's Statistics Paper Series.

Antonio Perrella

Bank for International Settlements, Basel, Switzerland; email: antonio.perrella@bis.org

Mr Perrella previously worked at the European Central Bank.

Julia Catz

European Central Bank, Frankfurt am Main, Germany; email: julia.catz@ecb.europa.eu

© European Central Bank, 2020

Postal address 60640 Frankfurt am Main, Germany

Telephone +49 69 1344 0

Website www.ecb.europa.eu

All rights reserved. Any reproduction, publication and reprint in the form of a different publication, whether printed or produced electronically, in whole or in part, is permitted only with the explicit written authorisation of the ECB or the authors.

This paper can be downloaded without charge from the [ECB website](http://www.ecb.europa.eu) or from [RePEc: Research Papers in Economics](http://RePEc.org). Information on all of the papers published in the ECB Statistics Paper Series can be found on the ECB's website.

PDF

ISBN 978-92-899-4256-0, ISSN 2314-9248, doi:10.2866/565922, QB-BF-20-001-EN-N