Connecting the ESCB Business Register with other Supranational Sources

An example of linking up to the GLEIF platform

Marcel Kropp, Sara Thijs, Peter Neudorfer and Sandrine Corvoisier[1]

1 Introduction

The 'Register of Institutions and Affiliates Database' (RIAD) is a business register jointly operated by and accessible to all members of the European System of Central Banks (ESCB), namely the European Central Bank (ECB) and all National Central Banks (NCBs) in the EU, and the National Supervisors of countries participating in the Single Supervisory Mechanism (SSM).

RIAD started as mainly a 'statistical' tool to maintain the list of 'Monetary Financial Institutions' (MFIs), i.e. those institutions whose balance sheets are used by the ECB to produce the monetary aggregates. Throughout the years the statistical populations covered in RIAD grew in line with expansions of the ECB's coverage of the financial sector in its statistical productions to also include Investment Funds (IFs), Financial Vehicle Corporations engaged in securitisation transactions (FVCs), Payment Statistics Relevant Institutions (PSRIs) and Insurance Corporations (ICs)[2]. Consequently, it has become a pivotal infrastructure within the E(S)CB that allows multiple groups of clients to share, jointly access and use reference data for various business purposes. Currently, RIAD holds reference data for around 440,000 organisational units (identifying among them 6,000 group structures) which are relevant for a wide range of the ESCB and SSM business areas[3].

In the wake of the financial crisis but also as a result of increasing complexity and interconnectedness of economies and institutions, it has appeared that aggregated statistics were no longer sufficient per se to analyse economic developments. Dispersion measures and sectoral analyses increasingly require the ability to drill down from aggregates to more micro (institution-level), or even granular (transaction-level) developments. The resulting rise of micro and granular data collections within the Eurosystem and SSM puts RIAD into the spotlight for integrating different datasets. Indeed it consolidates the core function and business model of RIAD as a shared view on institutions and groups through the management of reference data within the ESCB. By doing so, it also calls for the need to define (across domains)

---

[2]   Based on the reporting populations registered in RIAD, the ECB regularly publishes the various lists of financial institutions on its Website.

[3]   Outside statistics, RIAD's importance grew significantly both for stakeholders/business areas such as the ECB's own Market Operations in the area of liquidity and collateral management and the SSM for whom RIAD was chosen as the repository of the 'master data' for the respective supervised entities.

and identify the relevant populations. Even beyond this, RIAD stepwise allows to map different datasets and combine the datasets to allow a more valuable analysis. Prominent resulting initiatives are the Money Market Statistical Reporting (MMSR[4]), for which large credit institutions in the euro area report daily transaction-by-transaction data to the Eurosystem, and the upcoming AnaCredit project[5], which serves the purpose of collecting granular data on individual bank loans in the euro area (and some associated EU countries). The coverage of entities registered in RIAD is estimated to grow to 15 million, mostly non-financial corporations, when including information of parties involved in credit relationships.

Regarding the maintenance of RIAD, while the relevant reference data are usually provided by NCBs based on national sources (including local administrative sources), the ECB as operator of RIAD is also striving to make direct use of supranational sources.

In this context a special business case has recently gained importance, namely when a client would like to retrieve reference data from RIAD with solely the LEI (Legal Entity Identifier) as identifier of the entity. In case the respective LEI is not yet present in RIAD, a (semi-automated) mechanism has been developed that polls the so-called 'Legal Entity Identifier Service' online database provided by the 'Global Legal Entity Identifier Foundation' (GLEIF). Subsequently, the reference data are forwarded to the competent NCBs responsible for managing the data in the country where the entity is residing. The respective NCB is eventually responsible for verifying and integrating the information in RIAD.

Against this background this paper will briefly sketch the main features and data governance of RIAD in Section 2. Section 3 presents the current technical solution for the ad-hoc linkage of RIAD and GLEIF, and Section 4 sketches future plans on how to expand this input channel by also linking up to other external supranational sources. Section 5 concludes.

2   Main features of RIAD and data governance

The data model of RIAD covers an extensive set of attributes[6] which can be classified into four categories of information related to (1) the identification of units and their major characteristics, (2) properties that can be used for stratification, (3) demographic developments including corporate actions, and (4) various types of relationships between entities (e.g. ownership or control). As for each value of an attribute a validity range has to be provided, the RIAD data model allows a full historisation of the data.

The principles of the joint maintenance of reference data in RIAD are laid down in a dedicated framework for data provision and data quality management: On the one hand, for individual organizational units the residency/host country principle applies, i.e. the NCB of the country where the economic activity is performed is in charge. In contrast, NCBs of countries that are the home (legal registration) of a group head can contribute to all organizational units within the respective conglomerate.

---

[4]   See https://www.ecb.europa.eu/stats/financial_markets_and_interest_rates/money_market/html/index.en.html for a brief overview about the MMSR project.

[5]   The data collection is scheduled to start in September 2018. See 'The Analytical Credit Dataset – A magnifying glass for analyzing credit in the euro area' on the ECB's Website for a succinct overview of the AnaCredit project.

[6]   The RIAD data model follows the internationally recommended design of business registers. See 'Business registers – Recommendations manual', Eurostat, 2010.

With a view to ensure a fair and effective governance for the maintenance of the dataset, the local or 'competent' NCBs – in liaison for supervisory data with the respective National Competent Authorities (NCAs) – are the main gateways to RIAD for managing data on individual financial institutions in their jurisdictions (organised in local 'hubs'[7]). According to the current model a 'competent' NCB is not allowed to create an entity or add information to entities outside their own jurisdiction[8]. In addition, the ECB manages reference data for non-EU entities on a reasonable effort basis.

As statisticians in a central bank may not be the only actors responsible for the availability and quality of data, information from various business areas needs to be combined, e.g. statistics, supervision, market operations or payment systems. The RIAD system caters for this by means of 'multiple source management' and accepts data on the same unit or attribute from more than one national source. Out of these 'candidate sources' RIAD constructs the optimal 'authoritative' view of reference data on attribute level. The national data quality managers, having an informed view on the national data sources, are further entitled to decide which specific selection/compounding rule should be applied to the individual attributes by selecting a ranking of the various sources that go into the compounding mechanism.

To sum up, the information input to RIAD is **originally solely based on domestic sources[9]**.

The strict national responsibility can, however, not apply to the construction of cross-border relationships in the context of modelling group structures (conglomerates). More precisely, the type and features of the respective bilateral relationship stemming from the two NCBs (or NCAs) affected are processed in a controlled way. This goes so far that an NCB can even provisionally establish a missing counterparty in another jurisdiction.

3   A first usage of GLEIF data

RIAD is still primarily a product of and for the Eurosystem and the SSM. Although its coverage grew significantly in the past decade primarily for the financial sector, the increasing amount of stakeholders and the need for being a central reference database that can link various granular systems, calls for further efforts to incorporate data from non-local sources into RIAD.

For RIAD a special business case that resulted in a first connection to GLEIF has recently gained importance in the context of the Money Market Statistical Reporting (MMSR), an application that uses the Legal Entity Identifier (LEI) as the sole identifier of an entity[10]. More precisely, the MMSR tries to retrieve reference data from RIAD via queries based on the LEI.

To cater for cases where an LEI code could not be found in RIAD, a process has been developed that ensures for the respective LEI and reference data to be included in RIAD.

---

[7] The organisational structure refers to the NCBs as 'local RIAD hubs' and the ECB's RIAD Team as the 'hub of hubs'.

[8] The respective data provision and DQM responsibilities of NCBs are defined by ECB Regulations and Guidelines that describe the requirements of information to be provided ('mandatory', 'encouraged' or 'optional') as well as the timeliness.

[9] Particularly for entities in the euro area or EU.

[10] This constellation applies in particular to business units that are subject to supervisory legislation, as the LEI is the main identifier to identify counterparties in regulatory reporting.

Figure 1 Workflow to register missing LEIs (or Counterparties) needed for MMSR purposes in RIAD
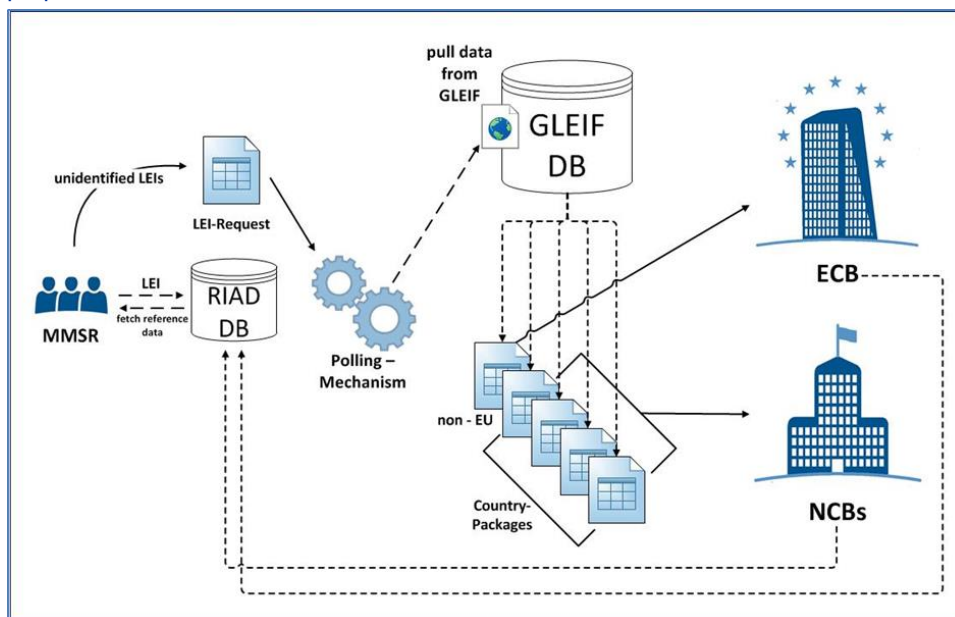
Triggered by a list of LEIs which the MMSR application could not find in RIAD, an external tool (i.e. a module not yet directly integrated into RIAD) pulls reference data from GLEIF's web-based search engine. The downloaded data (i.e. basic reference data describing an entity in GLEIF) is bundled together in country-packages and forwarded to the respective competent NCBs. The ECB remains responsible for cases related to entities resident outside the EU. Based on the counterparty information, NCBs and the ECB match the retrieved information against data already available in RIAD to decide if and how to incorporate the received information into the system. In case the respective counterparty could be identified, only the LEI needs to be added as identifier. If the entity is not yet registered in RIAD, it will be newly entered.

4 Broader usage of non-resident and supranational data

The example of the inclusion of GLEIF's reference data on a case-by-case-basis as described above is a valid illustration that the inclusion of non-local sources can contribute to a more comprehensive coverage and increase the quality of data in RIAD.

In fact, the forthcoming RIAD major release (RIAD 4.0)[11] is providing a new functionality that allows any NCB to add information on individual attributes of (and provisionally create) entities outside their jurisdiction. The data sourced by 'non-competent' NCBs will, in accordance with the principles of multiple source management applied today, be stored under a dedicated source identifying the respective NCB that added the information. The new functionality will be backed by a proper governance framework and complemented by technical compounding rules.

---

[11] Go-live planned for Q1 2018.

A challenge in this new framework will be to avoid the proliferation of duplicated entities in the database when, e.g. an entity that is presumably new in one country has de facto already been registered in the same jurisdiction. As an example, in the context of AnaCredit this is obviously a risk mainly for large borrowers or guarantors that may perform activities in many countries. To limit the risk of such duplications, the forthcoming new release of RIAD will be equipped with an integrated semi-automated procedure ('duplicate identification mechanism') which will leave NCBs (and the ECB) the option to 'manually' decide whether or not an entity proposed by another NCB is in fact a duplicate or not.

In other words, RIAD will, as of release 4.0, **move to a data provision model that explicitly allows for the incorporation of non-resident information**.

A logical next step would be to systematically open up RIAD to process data from supranational sources i.e. data collections that are by design not restricted to manage only reference data referring to entities resident in a single country.

As any data coming from external sources cannot be directly loaded into the system, a controlled inclusion of the data has to be guaranteed. The envisaged solution would be the integration of a dedicated 'staging area' to alleviate the risks of producing massive amounts of duplicates. This module within the RIAD application will allow both loading and reviewing data from supranational sources before actually incorporating them in RIAD. All incoming and subsequently staged data will be run against data already in the database by using the above mentioned duplicate identification algorithm. Subsequently, national data quality managers in each NCB are responsible for further assessment which can result in the data being added under a dedicated source of an existing entity or recognised as a new entity.

Currently four supranational data sets are of particular interest for the RIAD register. Each of these sources reveals distinct features that would be of particular interest for the E(S)CB:

- The **Eurogroup's Register (EGR)**[12] is the core of a network of business registers used for statistical purposes in European Union (EU) Member States on multinational enterprise groups. The broad coverage of entries would be of relevance for European institutions such as the ECB or NCBs and their business processes. Besides a small number of core reference data the main interest for RIAD would be relationships information, which can be expected to cover a wide (geographical) range of group members. One limitation is the timeliness and (annual) frequency of updates. Furthermore this statistical register cannot serve non-statistical purposes (although a large part of the information is in the public domain). Hence, the link between the EGR and RIAD only relates to the sharing of commercial data.

- **Global Legal Identifier Foundation data (GLEIF)**[13]. Based on the self-registration of market participants the so-called Level 1 data of the GLEIF cover a set of core reference data describing legal entities that perform various types of financial transactions. This is supplemented by Level 2 information that records relationships to (controlling) direct and/or ultimate parent institutions identified on the basis of consolidation rules. With a current focus on some regulated markets, GLEIF is expected to increase its coverage in the future. A major asset is that changes in data in principle have to be reported at the time a change occurs, which includes the likelihood that data are up-to-date.

---

[12] See http://ec.europa.eu/eurostat/statistics-explained/index.php/EuroGroups_register.

[13] See https://www.gleif.org/en/.

- Ownership or voting rights information originating from the **EU's Transparency Directive**[14]**.** The European Securities Market Authority (ESMA) is currently setting up a public European Electronic Access Point (EEAP) containing notifications of the acquisition or disposal of major holdings (and related changes of voting rights) of shares that are admitted to trading on a regulated market. The main focus would be (accurate) ownership data. The perimeter of regulated markets is of significant interest, the annual frequency on the other hand is a shortcoming.

- Group data included in the **FINREP/COREP data set**[15]. The Implementing Technical Standards (ITS) on supervisory reporting financial information (FINREP) covers all subsidiaries (even if they are not fully or proportionally consolidated), joint ventures & associates subsidiaries of supervised financial entities (see FINREP Template F_40.01). Again emphasis would lie on the identification of group membership information rather than on detailed description of the individual group participants. While the information would be available on a quarterly basis, it only covers the prudential supervision perimeter.

Because all supranational sources considered as potential candidates (EGR, GLEIF, EU's Transparency Directive, FINREP/COREP) also provide valuable relationship information, the envisaged procedure for the inclusion of any of the data sources into RIAD would cover two distinct flows: one exclusively on legal units and another one on relationships or data of the selected enterprise groups, respectively.

If the above sketched efforts will be successful, RIAD may consequently and ultimately move to **a model of continuous incorporation of broader sets of supranational sources.**

5   Conclusion

Since its inception the scope of the 'Register of Institutions and Affiliates Database' (RIAD) steered by the ECB and maintained on basis of contributions from all EU NCBs and from the SSM has been continuously expanded. Starting from a model that was solely based on domestic sources (every country is responsible for its own data and only has various national sources), a first loosening of this principle was to allow NCBs to also register data on non-resident entities. In this context, the forthcoming RIAD 4.0 system will complement this approach with an integrated semi-automated procedure ('duplicate identification mechanism') that leaves NCBs the option to 'manually' decide whether or not an entity proposed by another NCB is in fact a duplicate or not. The logical next expansion step would be to also allow the processing of various supranational datasets.

The biggest challenge of scaling up a jointly maintained database is defining a comprehensive governance framework which guarantees a controlled injection of supranational data without compromising on data quality.

---

[14]   Directive 2013/50/EU has amended the Transparency Directive (2004/109/EC).

[15]   See https://www.eba.europa.eu/regulation-and-policy/supervisory-reporting.