

Session 2: Quality of the statistical business register

Managing the quality of the ECB's enhanced 'Register of Institutions and Affiliates Database' (RIAD)

Peter Neudorfer, ECB¹

Background

For one-and-a-half decade the European Central Bank (ECB) operates the 'Register of Institutions and Affiliates Database' (RIAD), an application accessible to all members of the European System of Central Banks (ESCB) that is designed to hold reference data of (financial) organisational units relevant for statistical departments of European Union National Central Banks and other business areas, predominantly market operations. The RIAD system, as it was earlier set up and with its new features, allows the ECB to publish lists of Monetary Financial Institutions (MFIs) which is important information for the public at large, and in particular for the financial industry for the classification of their counterparts in reported transactions and positions. Lists of Investment Funds and Financial Vehicle Corporations (FVCs), the vehicles involved in securitisation transactions, are also published.

As its forerunner version, the recently completely refurbished business register comprises reference data on individual MFIs, Investment Funds and FVCs in the EU.² The main innovations of the 'enhanced RIAD' are its ability to cover individual reference information on more sub-sectors of the financial sector (S.12 in ESA 2010), an expanded list of data attributes related to these institutions, additional (access and update) functionalities and the capability to receive and process information on (ownership) relationships between organisational units (as prerequisite for the analysis of various types of banking groups and financial conglomerates).

RIAD consists of two modules - the Transactional Subsystem and the Data Warehouse subsystem (Figure 1). This architecture allows qualified users to add, manage and clean reference data in an environment that holds live, unstable data in the first module, while the second holds frozen, regularly updated snapshots on which reports and any kind of analysis would be based.

- On the input side the standard data acquisition takes the form of batch data transmissions from local NCB systems to the RIAD Transactional Subsystem, which trigger automated acknowledgement messages and (regular) dissemination of reference data collected to all ESCB members. In parallel any data input or update can also be carried out via an online facility. Furthermore RIAD receives daily data from the ECB's Eligible Asserts Database (EADB) and (on a low frequency) data from commercial sources channelled via Eurostat's Euro-Groups Register

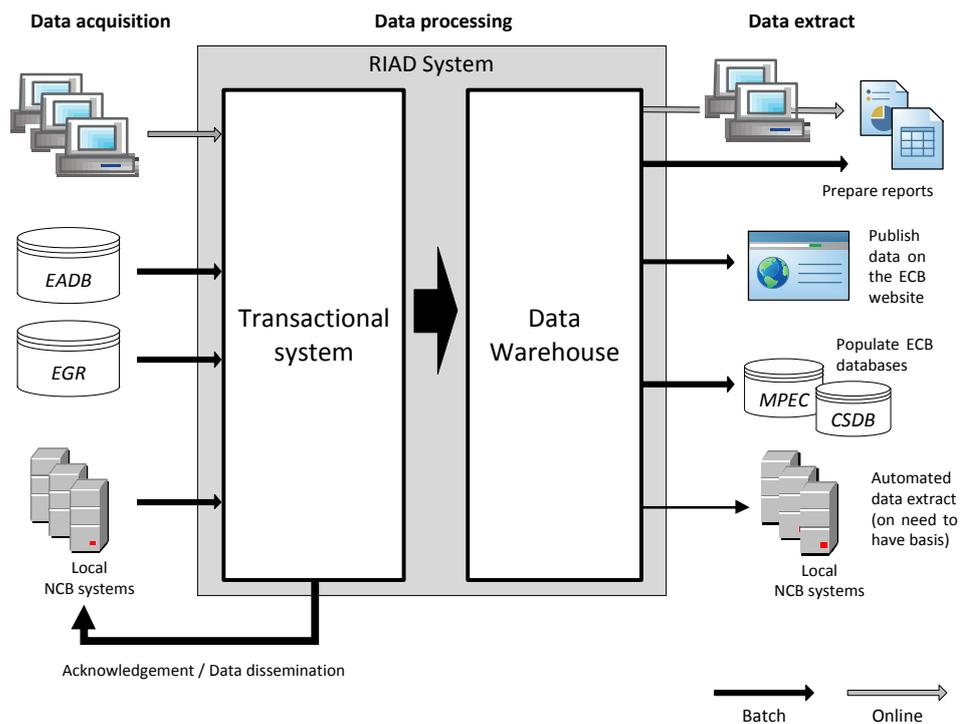
¹ The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the European Central Bank. Comments provided by Jean-Marc Israel, Lasse Nordquist, Patrick Sandars and Olgerd Unger have been much appreciated..

² At end-2012 there were some 7,100 MFIs, 49,500 investment funds (other than Money Market Funds) and 3,000 FVCs on the relevant lists of institutions resident in the euro area. In all EU Member States, the numbers at end-2012 total about 63,000 (excluding investment funds and FVCs resident in the United Kingdom).

(EGR). It is worth noting that the data model of RIAD is aligned with that of the EGR to enable easy data exchanges, when appropriate. While the EGR focuses on non-financial multinational groups of enterprises, the RIAD system will cover (a large part of) the financial sector, regarding both individual enterprises and groups composition. Data from the RIAD can be shared with the EGR.

- The data flow on the output side is based on the Data Warehouse module, which is in turn refreshed by the latest updates from the Transactional System on a daily basis. Typically individual end-users access online the former and create or run individual reports within a dedicated On-Line Analytical Processing tool. The same database is used to publish data on the ECB website and to update internal applications, such as the not public list of Monetary Policy Eligible Counterparties (MPECs) or the issuer information in the Centralised Securities Database (CSDB).

Figure 1: High level architecture of the enhanced RIAD system



The structure of the data model adopted in RIAD follows the general design of business registers and comprises four basic categories of attributes describing an organisational unit and its relationships:

- (A) Different types of information to identify a unit, covering ‘identifier’ codes as well as other descriptive variables such as name or address.
- (B) Stratification variables such as industry activity, size or institutional sector are usually employed for selecting fields of enquiry and taking samples.
- (C) Demographic features, essentially ‘date of creation’ and ‘date of closure’ and information on mergers or splits.
- (D) Attributes describing the relationships between units, which are the necessary building blocks to construct group structures (in a bottom-up approach).

For each variable validity ranges of individual attribute values can be recorded, eventually allowing retrieving snapshots of reference data by freely choosing dedicated points of reference dates.

Originally designed for the purpose of statistics and market operations areas in National Central Banks and the ECB, RIAD is now also serving the business needs of financial stability, and payment systems departments, and may also support the development and regular reference information necessary for micro-prudential supervision, once the ECB is in charge. While currently only EU National Central Banks are connected to the system, its architecture explicitly foresees that national or supra-national organisations outside the ESCB may be linked, such as Eurostat, National Statistical Institutes as well as the European and National Supervisory Authorities.

To sum up, the enhanced RIAD system allows different stakeholders to provide data (on all or subsets of entities, covering all or a subsets of attributes) to the database, perform the data quality management and (subsequently) deliver the up-to-date information to end-users (which may be individuals or systems).

Data Quality Management framework and principles

The Data Quality Management (DQM) framework of RIAD aims at following international standards and is as much as possible aligned with the “Data Quality Assessment Framework (DQAF)” of the International Monetary Fund (IMF).³ In order to be usable for RIAD, the macroeconomic background of the DQAF has at the operational level been adapted to fit in the context of a data collection and publication system that is based on reference data of individual statistical entities, i.e. to be useable at a micro level. Among other aspects the following quality dimension are covered and monitored in RIAD:

³ The IMF DQAF comprises 5 “dimensions”: integrity, methodological soundness, accuracy and reliability, serviceability and accessibility.

- Completeness - Completeness refers to the set of information captured and stored. This dimension ensures that the all requested information is covered in RIAD. An example is whether an entity defined in any relevant reporting populations is identified therein. Also, a completeness check is whether an attribute is served and updated, or not.
- Exactness - The exactness dimension ensures that the information provided by the sources is accurately loaded and stored within the RIAD system. While the “true” value may eventually need to be verified in separate steps, doubtful cases may for instance be detected by outlier checks and initially checked with authorised sources.
- Consistency - The consistency dimension ensures that the information stored does not entail duplicates and fulfils the business representation. This includes for instance the possible double-counting of entities, as well as cross checks between attributes of given entities such as start date and end date, or between sector classification and reporting requirements.
- Validity - The validity dimension ensures that the information provided is correct in the context of what it represents. Examples are the compliance of attribute values with specific domains (such as ISO code lists) or the simple check that numeric fields indeed include numeric values.

The underlying data provision and quality management process are organised as follows:

- National Central Banks act as main gateways to RIAD for all available data on individual financial institutions, based on the technical platform provide by the ESCB Net.
- The data provision and DQM responsibilities of NCBs are defined by respective ECB legal acts (regulations and guidelines); this includes in particular the requirements for specific classed of information (such as *mandatory*, *encouraged* or *optional*).⁴
- In all aspects of data provision and data quality management the *residency/host country* principle applies for individual organisational units and the *national /home country* principle for financial groups’ composition according to the corresponding leading home supervisor.

Multi source and confidentiality management

An essential aspect of the RIAD DQM framework is to improve quality of the information - such as ***completeness***, ***consistency*** or ***exactness*** - via the technical facility to acquire information from relevant, possibly more than one, source(s). Assuming that specific data providers have information of superior quality for specific (sub)sets of information (at different points in time), the overall quality of the data collected and published via RIAD can be increased by adequately combining these inputs.

While RIAD has so far solely been fed by statisticians in National Central Banks, the new framework explicitly allows that in individual countries organisational models may implement a split of responsibility between departments within NCBs, e.g. statistics, market operations, financial stability or (banking and/or insurance) supervision that may in turn appear as separate sources.

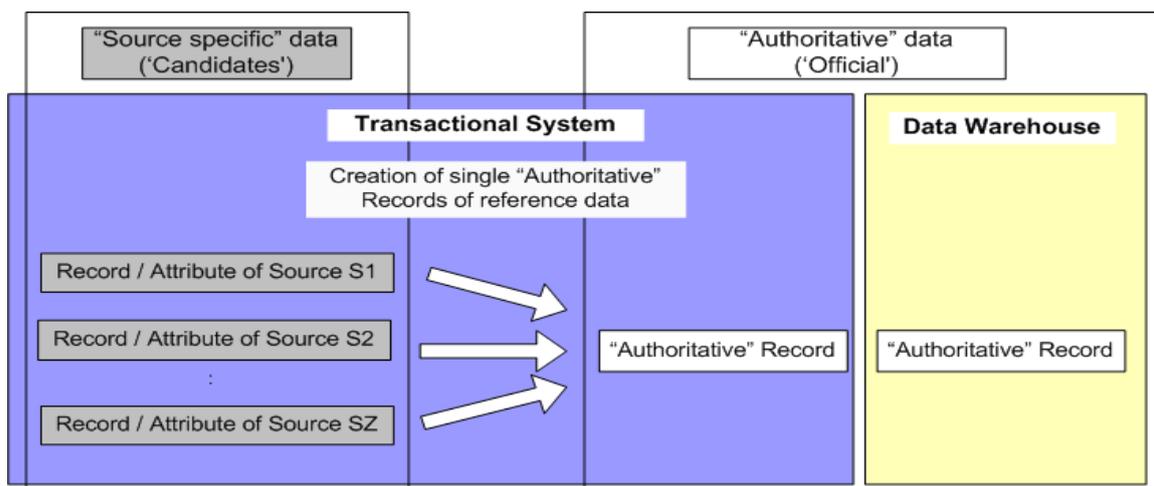
⁴ See the Guideline ECB/2007/9 on monetary, financial institutions and markets statistics as amended <http://www.ecb.europa.eu/ecb/legal/pdf/02007o0009-20120914-en.pdf> (update forthcoming by end 2013)

Figure 2 illustrates how the input from potentially multiple sources (or ‘candidates’) will be condensed to a single ‘authoritative’ set of reference data (compound). Consequently the following principles of data representation in RIAD apply:

- Potentially several instances of “source specific data” (candidates) can exist in the RIAD Transactional System.
- Only a single instance of “authoritative data” (official) exist in the Transactional System; however some information may be hidden for specific users due to confidentiality constraints.
- Only a single instance of “authoritative data” (official) can exist in the Data Warehouse; however, some information may be hidden for specific users due to confidentiality constraints.

In case multiple data on the same unit and feature for the same point in time would be available from more than one source, prioritisation among data providers will prevent overlaps and simple overwriting. As a result the system stores for each value assigned to an individual attribute in RIAD the source that provided the information.

Figure 2: Data processing in the ‘Enhanced RIAD’ System



In parallel the DQM framework of RIAD also aims at improving the **accessibility** dimension by a specific functionality to manage the confidentiality of the information processed and published. To this end RIAD stores for each value assigned to an individual attribute a confidentiality status as assigned by the respective data provider. The confidentiality status of an attribute is to be carried throughout the entire process and eventually determines the accessibility of information in the two sub-systems according to specific user profiles.

According to data quality management principles set out above each Central Bank is responsible for the quality of reference information related to individual organisational units recorded in its jurisdiction. As the delineation of 'authoritative' reference data has to reflect both the 'source' and 'confidentiality' information for the individual attribute values provided in the data acquisition step both 'source' and 'confidentiality' need to be processed simultaneously.

To this end the data processing algorithm in RIAD provides different selection / compounding algorithms and national data quality managers can decide ("switch on"), which specific selection / compounding rules for individual attributes will be executed:

- In cases where 'multiple' information (i.e. data from more than one source) exist for a single attribute and the values are not identical, selection (compounding) rules can be applied to derive the authoritative value, based on a predefined *hierarchy of sources*.
- In analogy also for the 'confidentiality' of an attribute value only a single confidentiality status will be carried through to the authoritative representation, again based on a *confidentiality hierarchy*. The result is an exact classification of accessibility, which can be compared with the privileges of specific user groups, i.e. determining access to / exclusion from specific class of protected data.

The rationale behind this flexible system is that each Central Bank is supposed to have the best view on the national situation, i.e. knows sources for each sector of financial institutions and for each attribute that can be used, which quality level can be assigned to them and which level of confidentiality would apply.

Next steps

The new enhanced RIAD system went live in May 2013. It still receives the same reference data for the same entities as in the previous decade as set out in the respective ECB Guideline currently in force⁴. The forthcoming update will in 2014 stepwise enforce to cover the complete set of enhanced properties foreseen in the new data model. National reporting systems will need to interact with the new RIAD platform and the new entities and attributes therein.

At the end of this transitional period RIAD is supposed to acquire and combine the complete new rich data set from reporting agents or selected commercial sources, under the checking and validation of National Central Banks. Concerning the expansion of the financial sector, RIAD will soon also cover insurance corporations.

In the steady state it is expected that the enhanced RIAD system, with its multiple source management, on-line accessibility and recording of various types of links among organisational units becomes a high quality register of financial corporations and groups for a multitude of users in the ESCB and associated national and European authorities.