Register of Institutions and Affiliates Database (RIAD) – Modelling for flexibility

Michele Romanelli*
Banca d'Italia, Rome, Italy – michele.romanelli@bancaditalia.it

Abstract

A complete, accurate, up-to-date register of business entities is a cornerstone component of a mature information system. The significant (administrative) costs necessary to manage its content could be positively balanced by benefits if such component is shared among different applications and use cases, so to put in common the efforts towards timeliness, completeness and accuracy of its data and ensure a reliable and trustable "single source of truth". Conceiving and managing a register as a shared service puts, though, additional challenges, in particular to inject the necessary flexibility to adapt and shape its content to the multiple needs of the different stakeholders. A proper modelling of the register's data structures and its core processes is, in turn, crucial to ensure such flexibility.

The paper presents few design solutions adopted for the 'Register of Institutions and Affiliates Database' (RIAD) jointly managed by all members of the European System of Central Banks (ESCB) to handle the reference data of various organisation units supporting several E(S)CB institutional functions. Following a short description of RIAD and its evolutionary path (Introduction), the main requirements towards the flexibility are presented (section 2) together with the description of the technical solutions put in place to address the data modelling (section 3) and the historicity management (section 4). The last section presents some conclusion.

Keywords: Register of business entities; Data modelling; Flexibility; Historicity management.

1. Introduction

The Register of Institutions and Affiliates Database (RIAD) is a central repository of reference data about entities operated by European Central Bank (ECB) and jointly managed by all members of the European System of Central Banks (ESCB) under the system ownership of the ESCB Statistics Committee (STC). RIAD holds attributes on individual organisational units as well as various types of relationships between them aiming at providing the frame for statistical reports on financial institutions in the EU and/or euro area.

Over the years, the RIAD system has undergone a series of subsequent expansions and enhancements aimed at supporting different ESCB processes and new user needs. Indeed, it started as a pure 'statistical' tool to maintain and regularly publish the euro area as well as entire EU list of Monetary and Financial Institutions (MFIs, 2002). This mission was steadily expanded to other euro area (or EU) list of financial institutions: Investments Funds (2008), Financial Vehicle Corporations engaged in securitization (2008), Post Offices and Giro Institutions (2015) and, recently, Insurance Corporations (2017).

On the other hand, during the last years the concept of data compilation on the basis of micro data has spread over many business areas inside the ECB and the NCBs – beyond the typical production of statistics. Consequently a strong demand for reference data describing various types of business units can be observed. The landscape of stakeholders and business cases has consequently been steadily enlarging (Neudorfer, P., 2016).

For example RIAD is highly used in the 'Market Operations' liquidity and collateral management's area, mainly to assist in the identification of relevant counterparties. In this context a special module implemented in RIAD allows to scan the networks of conglomerates in order to identify specific connections between issuers and holders of assets acknowledged as collateral in monetary policy (so-called 'close links').

Similarly the area of Risk Analyses and Management is increasingly using reference data managed and stored in RIAD to match various counterparties which may, directly or indirectly, generate various types of risks for the Eurosystem (e.g. debtors behind non-marketable instruments).

The Single Supervisory Mechanism (SSM) is also stepwise adopting RIAD for storing and disseminating reference data on supervised institutions (e.g. the metadata on Supervisory Reporting requirements[1]) and their group structures.

In recent times the increasing demand for integrated statistics across countries and markets is leading the system owner to leverage once more RIAD as master reference data repository about counterparties of interest. Prominent examples for this trend are the 'Money Market Survey'[2] and the forthcoming Analytic Credit data collection on lenders and borrowers behind loans (AnaCredit) or on the issuer and holders of securities[3].

In short, the system has progressively become a pivotal platform to store and distribute reference data on corporations relevant for various ESCB business processes, also beyond statistics, and for the Single Supervision Mechanism (SSM). From the technical perspective, the initial system (2002) was replaced by a significantly enhanced version (2013) and is now undertaking a major expansion (go live by 2018) to serve as counterparty repository for the AnaCredit system.

2. The requirements towards the flexibility

In 2011, pushed by the need for the (existing) RIAD 3.1 technological renewal, the STC launched the 'RIAD 3.2 enhancement' project with the aim inter alia of "*…enhancing the functionality and usability of RIAD* beyond the statistics *also for market operations and financial stability purposes*". The ultimate goal was to transform RIAD into a backbone to support the ESCB and SSM granular datasets, a necessary condition for effective data integration. In practice, the system has been evolved from a simple list keeper for relevant EU entities to a rich toolset allowing:

i) acquisition of reference data from multiple, potentially conflicting sources and derivation of authoritative views, based on various compounding criteria;
ii) identification of similarities of counterparties and de-duplication of double entries (Central Identification Service);
iii) derivation of various conglomerate data structures (e.g. groups structures);
iv) identification of actual and potential close links between counterparties involved in monetary policy activities.

Such goal was translated into a set of quite ambitious requirements:

1. Expanding the data model to include additional attributes and to enlarge the reference population aimed at fulfilling the needs of the various user groups;

2. Introduce a parsimonious data transmission format so to limit the data acquisition only to the attributes whose value changes;

---

[1] See EBA 'Implementing Technical Standards Amending Commission Implementing Regulation (EU) No 680/2014 on Supervisory Reporting of institutions'.

[2] Regulation of the ECB concerning statistics on the money markets (ECB/2014/48)

[3] Regulation of the ECB on the collection of granular credit and credit risk data (ECB/2016/13) AnaCredit and Regulation (EU) No 1011/2012 of the ECB concerning statistics on holdings of securities

3. design the system and its data model for flexibility so allowing that further attributes expansions could possibly be implemented with minor changes in the software application;
4. provide for each individual piece of information the possibility to describe the data source and the specific confidentiality level;
5. keep and manage the full historicity of the reference data.

The system has been implemented as two separate, although interacting modules: a Transactional module, in charge of collecting data from ESCB National Central Banks (NCBs) and providing data quality management functionalities and an Analytical module offering functionalities for querying, reporting and extraction of reference data. In order to fulfil the above challenging requirements, the system has been built on top of metadata-driven solution platforms specifically designed to support the generic data collection and usage of statistical data[4]. The way the above requirements have been translated into the IT solutions is discussed in the following sections.

3. RIAD 3.2 enhancement: a data model designed for flexibility.

The RIAD 3.2 data-model requirements have been grounded on the lessons learned on the existing RIAD 3.1 version, with the aim to overcome the weaknesses therein experienced.

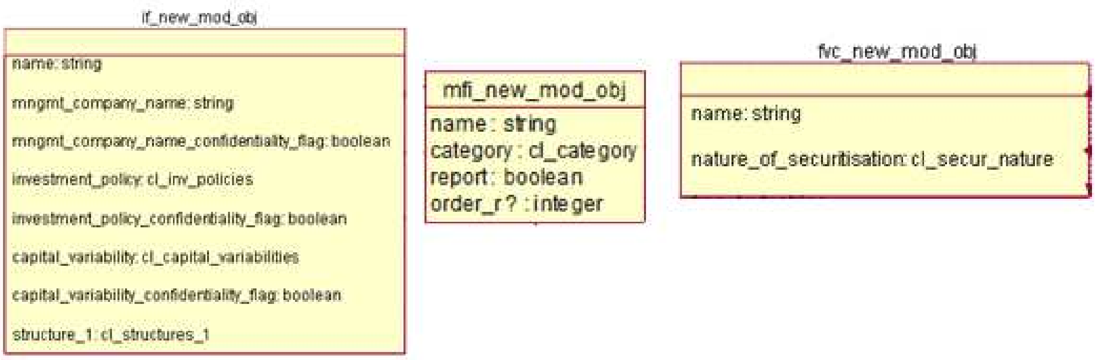

Figure 1 RIAD 3.1 core data model

Indeed, the RIAD 3.1 data model was approached as separate, specific datasets organised along the institutional sectors as identified in National Accounts, one for each list managed within RIAD (figure 1 above)[5]. In addition, some metadata attributes, like e.g. the confidentiality, where not implemented in consistent way across the datasets. The introduction of a new sector of entities required a major change in the system in order to design the new data set structure and implement all the related management functions. The enrichment of a list with new attributes required as well a similar major change. In addition, the lists were managed as snapshots, collected at predefined frequencies so that the full complete lists were required to be provided by NCBs at each deadline[6].

[4] The Transactional module is based on the Banca d'Italia INFOSTAT platform, a bespoke IT system for statistical processing (Di Giovanni, F., Piazza, D., 2009), based on a robust and well founded information model, called "Matrix" (Del Vecchio, V., Di Giovanni, F., Pambianco, S. 2007 and Del Vecchio, V. 1997). The Data warehouse module is based on off-the-shelf products for U2A query & reporting plus web services for A2A data access.

[5] For instance, MFI, IF and FVC refers to Monetary and Financial Institutions, Investment Funds and Financial Vehicle Corporations list.

[6] The MFIs' list was an exception as only changes were collected on daily frequency.

Also the database logical design suffered the same weaknesses with separate, custom tables one for each list. From this perspective, even a change in the economic sector of an entity implied some complex, error prone activities[7].

The envisaged multipurpose usage of RIAD 3.2 postulated the need to handle the system in a much more flexible and agile way according to the requirements depicted in section 2.

Based on the “Matrix” information model, the RIAD 3.2 model has been designed with a new approach: instead of thinking in terms of datasets along institutional (sub-) sectors, the generic “RIAD institution” has been modelled as an “organisational unit” having a set of “properties” and linked to other organisational units through “relationships”. In turn, each “property” and each “relationship” has been modelled as a mathematical function (also called data cube) having its own multidimensional structure (dimensions, measures and attributes). The general structure for properties and relationships is the following:

Property data cubes

CubeID	Dimensions				Measure	Attribute	
property name	source	orgUnitID	validFrom	ValidTo	property value	Confidentiality	updateTstmp

Relationship data cube

CubeID	Dimensions					Measure	Attribute	
relationship name	source	orgUnitFrom	orgUnitTo	validFrom	ValidTo	relationship value	Confidentiality	updateTstmp

Figure 2 - Data model of RIAD 3.2 (excerpt)

Also the data base logical model was simplified with three main tables: the organisational unit, the properties and the relationships:

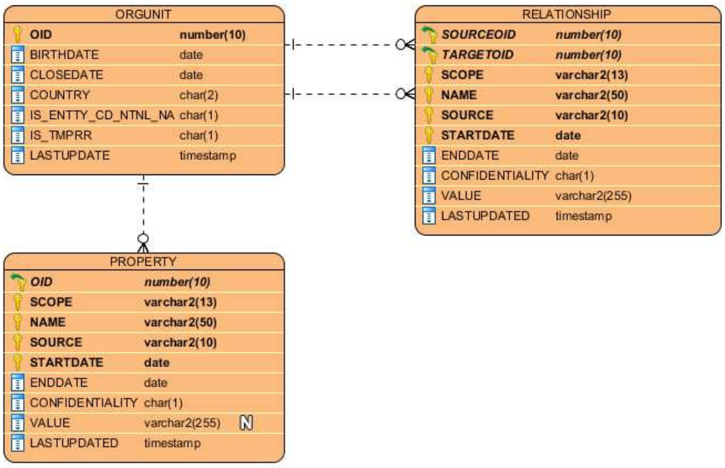

Figure 3 - RIAD 3.2 (core) data model

The above approach added a significant flexibility on the system as it was able to capture all the main needs of the users:

- thanks to the active data dictionary supporting the system, introducing a new property/relationships, previously requiring a major software change, has become as simple as adding a new data cube definition in the dictionary, an effort comparable with a code-list

[7] For example, the change of an investment fund into a money market fund required (i) the deletion of the fund from the IF list; (ii) the insertion of the same fund in the MFI list and (iii) the consequent adaptation of the data structure of the entity.

enrichment. Also from the database perspective, the addition of a new property/relationship translates into adding rows into an existing table[8]. As a result, the new entity lists introduced from 2014 onwards (i.e. the Post Offices and Giro Institutions and Insurance Corporations) required a development effort of a few working days, compared to the weeks or months necessary for the introduction of the list of Investment Funds and Financial Vehicle Corporations in the old system. The number of properties/relationships has been enlarged from the 15 available in RIAD 3.1 to some 90 available today while the IT effort necessary to support such enlargement has been negligible.

- the definition of a data cube for each piece of information allows various contributors to provide only the attributes under their responsibility, instead than the full picture of an organisational unit as was requested in the previous version of the system (parsimonious data format);
- also the management of the multiple sources as well as the confidentiality and the validity interval for each property/relationship has been harmonised (*source*, *validFrom*, *validTo* dimensions and *confidentiality* and *updateTmstmp* attributes have been included in each data cube);
- the adoption of a robust and formal information model allowed the transposition of the RIAD data model into two reference technologies like SDMX and XBRL, enabling the contributing NCBs to better integrate the RIAD data into the existing frameworks and toolsets for statistical and supervisory data management.

4. Historicity handling

The RIAD system is capable to keep the validity time interval for each property and relationship value, implying a full historicity of the organisational units across their lifespan. Moreover, the system is also requested to support the amendment to the history at any point in time.

Although this is achieved at model level through the usage of *validFrom-validTo* attributes, the technical challenges arising from the requirements are significant. Indeed, while in a "snapshot" approach, the new data submitted by the contributors override the existing ones, the historicity management requires "merging" the new data with the old ones. For example, a change into a property value would imply two actions: (i) closing the validity of the previous record and (ii) inserting the new one; therefore there is a need to unambiguously identify the record to be closed and ensure the mutual consistency between the different time intervals (of the old and the new ones).

Operationally, the above actions could be complex and error prone, in particular if NCBs contributing to the central reference data base would be requested to keep their local reference databases in sync with the RIAD database (as it was the case in RIAD 3.1).

A different approach, implemented within RIAD 3.2, is based on the assumption that new data submitted by any contributor is, by definition, more up-to-date than any other data so far recorded into the system and originated from the same contributor; therefore the new data could safely be used to "override" the current database content. The consequence is significant as (i) the contributor is not requested to unambiguously identify the record(s) to be closed – the complex part in the historicity management – and (ii) the system doesn't need to check the mutual consistency between the existing information and the incoming one. Because the incoming information is assumed to be correct, the system applies all the necessary amendments to the existing information so to allow the new one entering the system while keeping the overall data consistency.

Figure 4 illustrates this behaviour on the example of the property "Address".

[8] Contrast this with the previous approach where a change in the physical database structure was required.

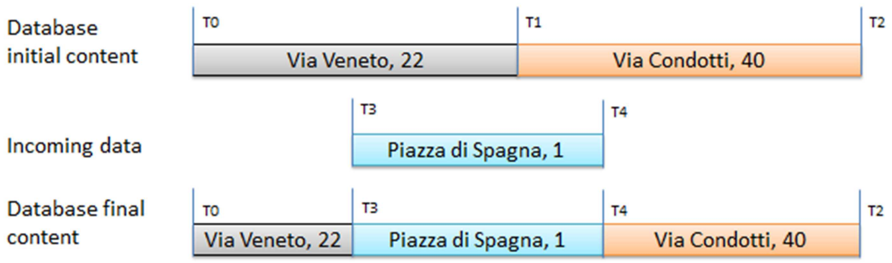

Figure 4 - example of data update taking into account the historicity

In the initial configuration, the address occurs two times: “Via Veneto, 22” in the time interval T0-T1 and “Via Condotti, 40” in the time interval T1-T2. Then a contributor provides a new address “Piazza di Spagna, 1” valid in the time interval T2-T3. The system executes all the necessary actions to get to the final configuration, namely: (i) change “Via Veneto,22” validTo date from T1 to T3; (ii) change “via Condotti, 44” validFrom date from T1 to T4; (iii) insert the incoming record.

5. Conclusions

The management of a register of business units is by nature a complex activity. In the case of RIAD, such complexity is enhanced by both the multipurpose usage and some challenging requirements.

In this context a key success factor for the IT system is to offer the necessary flexibility to shape the system behaviour according to the needs of different user groups and to avoid introducing additional complexity on top of the one intrinsic to the business activity.

While the two technical aspects discussed in this paper may appear small, their impact on the flexibility and simplicity has proven to be extremely effective, contributing among others significantly to the acceptance of the system by various user groups.

Eventually these features also triggered the decision to use RIAD as the authoritative repository on all counterparties engaged in credit activities, as requested by the AnaCredit regulation. As a consequence, a significant capacity expansion project (RIAD 4.0) has been launched which – while confirming the validity of the data model and historicity management - will prepare the IT platform to support - as of 2018 - the storage and handling of some 17 million entities in RIAD.

References

Neudorfer, P.(2016). Multiple usage of the ESCB ‘Register of Institutions and Affiliates Database’ (RIAD) – features and challenges. 25th Meeting of the Wiesbaden Group on Business Registers. http://www.stat.go.jp/english/info/meetings/wiesbaden2016/pdf/2-02_ecb_pap.pdf

Di Giovanni,F., Piazza,D.(2009). Processing and managing statistical data: a National Central Bank experience. International Statistical Conference, Prague. https://www.czso.cz/staticke/conference2009/proceedings/data/process/piazza_paper.pdf

Del Vecchio,V., Di Giovanni,F., Pambianco,S.(2007): The Matrix Model – Unified model for statistical data representation and processing, Banca d’Italia. http://www.bancaditalia.it/statistiche/raccolta-dati/sistema-informativo-statistico/modellazione/matrixmod.pdf

Del Vecchio, V.,(1997): Statistical Data and Concepts Representation, Banca d’Italia. http://www.bancaditalia.it/statistiche/raccolta-dati/sistema-informativo-statistico/modellazione/StatisticalDataAndConceptsRepresentation.pdf