



## The challenges of standardisation and aggregation of EMIR data in Europe: Six trade repositories and 28 countries<sup>1</sup>

Linda Fache Rousová\*

[Linda.Fache\\_Rousova@ecb.europa.eu](mailto:Linda.Fache_Rousova@ecb.europa.eu)

Małgorzata Osiewicz

[Malgorzata.Osiewicz@ecb.europa.eu](mailto:Malgorzata.Osiewicz@ecb.europa.eu)

Grzegorz Skrzypczyński

[Grzegorz.Skrzypczynski@ecb.europa.eu](mailto:Grzegorz.Skrzypczynski@ecb.europa.eu)

European Central Bank, Frankfurt, Germany<sup>2</sup>

### Abstract

Since February 2014, the European Market Infrastructure Regulation (EMIR) requires all counterparties in the European Union (EU) entering into a derivative contract to report its details to one of the six authorised trade repositories. The decentralized and heterogeneous landscape for this reporting obligation poses significant challenges to the regulators accessing and analysing the data. The situation in Europe seems unique, as even basic concepts such as the definition of a derivative contract are not well developed. Nonetheless, when it comes to the global aggregation of these data, similar difficulties likely exist everywhere, so other jurisdictions could benefit from the European experience. This paper elaborates on these challenges, putting particular emphasis on the lack of common standards, data quality caveats and the need to match the data from the different trade repositories.

**Keywords:** derivatives; European market infrastructure regulation (EMIR); trade repositories; harmonisation.

### 1. Introduction

According to European Market Infrastructure Regulation (EMIR), since 12th February 2014 all counterparties located in the European Union (EU)<sup>3</sup> that enter into a derivative contract have to report the details of the contract to one of six trade repositories (TR) authorised under EMIR. These are (i) CME Trade Repository Ltd. (CME), (ii) DTCC Derivatives Repository Ltd. (DDRL), (iii) ICE Trade Vault Europe Ltd. (ICE), (iv) Krajowy Depozyt Papierów Wartościowych S.A. (KDPW), (v) Regis-TR S.A. (Regis-TR), and (vi) UnaVista Limited (UnaVista). As a result, the data are currently scattered among the six TRs, which embraced different technical solutions for storing, providing and representing the data. Moreover, the data are not standardised and suffer from serious drawbacks in terms of quality.

---

<sup>1</sup> The authors thank to Carlos Sánchez Muñoz, Harun Mirza and Peter Hoffman for their useful comments but the authors remain solely responsible for any remaining errors or omissions.

<sup>2</sup> The views expressed in this paper are solely those of the authors and do not necessarily reflect the opinion of the European Central Bank.

<sup>3</sup> This paper focusses on the implementation of EMIR in the EU. EMIR is however an EU legal act marked as European Economic Area (EEA) relevant and currently under consideration for incorporation into the EEA Agreement by Iceland, Liechtenstein and Norway.

This decentralized and heterogeneous landscape poses significant challenges to more than 50 regulators in Europe accessing, managing and analysing the EMIR data from the six TRs. Moreover, high data confidentiality causes additional difficulty, restricting the exchange of the knowledge gained by individual regulators in handling the dataset.

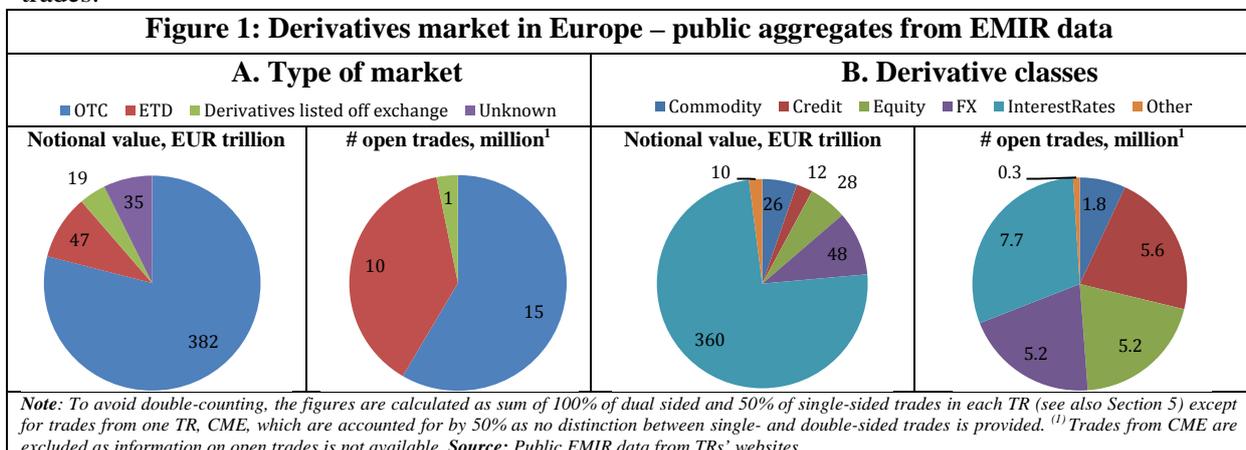
In this paper, we first describe the main features of the EMIR data (Section 2). Drawing on the experience gathered in the Directorate General Statistics of the European Central Bank (ECB), we elaborate on the challenges of these data, putting particular emphasis on the lack of common standards and the need to match the data reported to the different TRs (Section 3). Furthermore, Section 4 focusses on the various caveats in the EMIR data quality from the micro-data perspective; while Section 5 takes the macro-data view and presents the comparison of the aggregate EMIR over-the-counter (OTC) credit derivatives data with similar data collected through the semi-annual surveys conducted by the Bank of International Settlement (BIS). Finally, Section 6 concludes.

## 2. EMIR data in general

EMIR is a far-reaching reform of the derivatives market in Europe and introduces, inter-alia, a reporting obligation to all counterparties located in EU, which trade a derivative contract. The reporting obligation applies to all types of derivatives contracts – both OTC and exchange-traded derivatives (ETD), on all main five derivatives classes (credit, commodity, equity, interest rates and foreign exchange), including trades cleared via Central Clearing Counterparties (CCPs).

Around 85 data fields are to be reported for each transaction and they are divided into two groups. The first group contains information on the counterparties involved, which usually remain static over life cycle of a transaction. The second group provides details on the characteristics of the contract (e.g. type of derivative, underlying, prices, amount outstanding), how/on which venue the contract was executed and/or cleared, valuation and collateral, and life-cycle events (e.g. new contract, modification, termination). The high granularity in principle allows for the derivation of positions and aggregate data but this is not without caveats in practice owing to the lack of standards to report the data to the TRs (which for instance impede matching the two legs of each contract) and the different features of the outputs/files provided by the six TRs (see also Section 3).

Such a wide-scaled and detailed reporting implies huge data volumes. Over the first year of reporting, almost 10 billions of records were received and processed by the six TRs in Europe according to European Securities and Markets Authority (ESMA), the supervisor of the TRs (ESMA, 2015a). In addition, Figure 1 shows that the notional values and number of outstanding trades on a given date (reaching almost EUR 500 trillion and over 25 million respectively) are also substantial. In terms of notional values, OTC trades account for around 80% of the derivatives market in the EU (Panel A) and interest rates derivatives with around 75% represent by far the most important derivative class (Panel B). However, as both OTC trades and interest rates derivatives tend to be associated with relatively large notional values, the distribution is much more balanced, when considering the number of open trades.



### 3. Challenge of data scattered across six different TRs

The EMIR legislation and TR supervision do not currently provide a sufficiently detailed and coherent framework to provide regulatory authorities in the EU with high quality data about the derivatives market. First and foremost, since the EMIR secondary legislation leaves significant room for interpretation of the reported fields, the data are not standardised at the input point, when entered into TRs by reporting agents. In particular, the EMIR reporting relies on concepts such as the Unique Trade Identifier (UTI) and Unique Product Identifier (UPI), which are not yet developed; several reported fields are not accompanied by code-lists (lists of allowed values) and the interpretation of their content raises questions. Second, the EMIR framework does not provide any detailed guidance on how the six TRs shall treat, structure and present the collected data. For instance, even the variable names, the number and structure of the data files provided to regulators vary from TR to TR. Another level of complexity in EMIR is added by the differences in certain aspects of reporting across EU member states with the remarkable example of the definition of a derivative contract, and consequently what type of transactions should be reported (see e.g. Maxwell, 2014).

A peculiar point in EMIR reporting is the so called 'double reporting obligation', i.e. both counterparties to a derivative transaction have the reporting obligation, if they are located in the EU (while in other jurisdictions one counterparty is usually obliged to report on behalf of both sides to the trade). As a result, trades are frequently reported twice under EMIR (i.e. always when the two counterparts are EU resident), possibly to two different TRs. Hence, any meaningful data aggregation requires the reconciliation of the information between the duplicated trades, which shall in principle rely on the use of a Unique Trade Identifier (UTI), but its definition on a global level is still under development<sup>4</sup>. In the meantime, ESMA put forward temporary guidelines how an interim UTI should be generated (see ESMA, 2015b).

Our results show that the pairing rate is particularly low in case of trades reported to different TRs. Moreover, even if the two legs are paired using an interim UTI, information in the other data fields submitted by the two counterparties very often do not match, which raises the question which of the two to keep in the final database with de-duplicated trades. Even for trades reported to the same TR, there can be significant discrepancies for variables such as execution timestamp, price per contract or notional value. Buyer/seller field proved to be particularly challenging to report, as for many types of contracts (e.g. interest rates swaps) the distinction between the buying and selling party is not applicable or arbitrary.

Reconciling and aggregating data across the different TRs may be easier in certain market segments, especially when reporting of a particular asset class is concentrated in one TR. This is the case for OTC credit derivatives, which are reported in around 95% (in terms of notional values) to DDRL (see Figure 2). Similarly, around 75% or more of all the other OTC derivatives classes except for commodity derivatives are also reported to this TR.

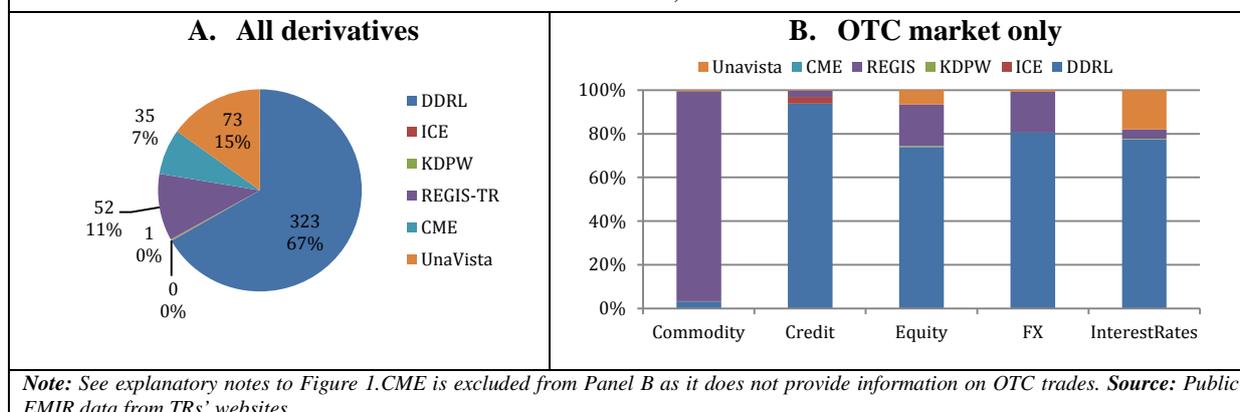
### 4. Data quality caveats

To monitor the quality of EMIR data, we have developed several checks on the micro-data accessible to the ECB. In particular, we regularly check the number of missing values and the use of the key identifiers such as Legal Entity Identifiers (LEI), UTI and UPI as well as International Securities Identification Numbers (ISINs).

---

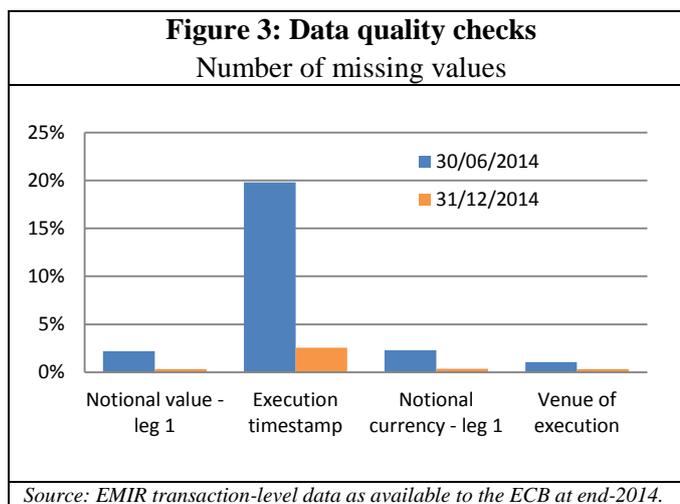
<sup>4</sup> The Financial Stability Board (FSB) has asked the Committee on Payments and Market Infrastructures (CPMI) and the International Organization of Securities Commissions (IOSCO) to develop global guidance on harmonisation of data elements that are reported to trade repositories, in particular the UTI and the UPI. It is envisaged that the sub-structure created for this purpose, the CPMI-IOSCO Harmonisation Group, will publish final guidelines on the UTI by the end of 2015.

**Figure 2: Importance of the six TRs in Europe**  
Notional values at end-March 2015, EUR trillion and % of total



The results suggest that the current quality of EMIR data is relatively low, but that the situation is gradually improving, owing to large extent to ESMA's on-going efforts to improve EMIR data quality.<sup>5</sup> For instance, Figure 3 shows that the number of missing values for some fields has significantly decreased over the second half of 2014. Similarly, as the same UTI is expected to appear only once or twice in the data on outstanding trades on a given date, we monitor the cases of triplicated or more often repeated UTIs and found out that those cases dropped from more than 30,000 in the initial reporting stage to about 300 at the end of 2014 (based on data available to the ECB).

We also match the reported counterparty identifiers with those in the official LEI database<sup>6</sup> to examine to what extent the correct LEIs are used. The match is quite successful for the field of the reporting counterparty (in 99% of trades) but remains lower for the field of the other counterparty (around 75%) as the reporting counterparty may not know the LEI of the other entity. The number of distinct LEIs in the data has been significantly increasing since the reporting start (e.g. tripled for the reporting counterparty between mid- and end-2014), as more entities apply for LEI and join the reporting.



## 5. Comparison of EMIR OTC credit data with BIS CDS survey data on OTC derivatives

Beyond the checks conducted on micro-level, we also compare aggregated EMIR data with those obtained from the established semi-annual OTC derivatives surveys conducted by the Bank of International Settlement (BIS). We focus on OTC credit derivatives, as we expect the data for this

<sup>5</sup> ESMA, the supervisor of TRs, aims at improving EMIR data quality through a data quality action plan, which foresees that TR run consistent data validations at the data submitted by reporting agents. The Level 1 validations were already put in place in December 2014, while Level 2 validations are expected to be introduced in the course of 2015 (see ESMA, 2015a for more details).

<sup>6</sup> This list can be downloaded from the website of the Global LEI Foundation: <https://www.gleif.org/en>

asset class to be of somewhat better quality, owing to the high concentration of these data in one TR (DDRL) and to the prior experience of market participants with the voluntary reporting of CDS to this TR since 2008. Due to the double-reporting regime, mentioned before, the transactions are divided into the categories of dual-sided and single-sided trades. We derive from the public EMIR data lower-, middle- and upper-bound estimates for the EU aggregates using different assumptions (see Table 1 for an overview and the underlying assumptions).

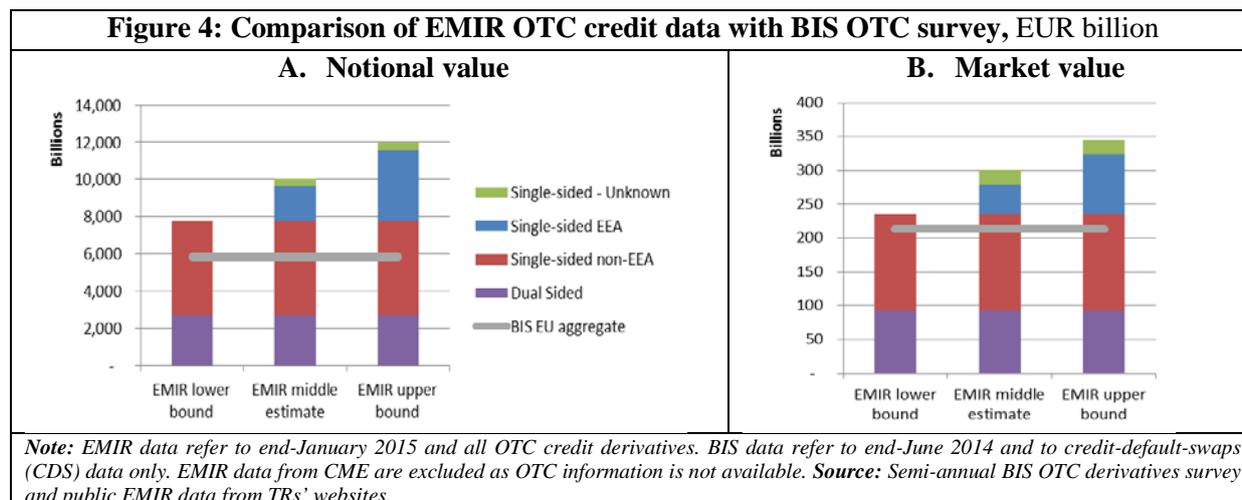
**Table 1: Types of transactions and estimates for EU aggregates**

Category	Explanation	Amounts included in the computation of the EU aggregates		
		Lower bound	Middle estimate	Upper bound
Dual sided	Both legs of transaction in the same TR	As in the middle estimate	100% - as the amounts refer to one transaction without double-counting	As in the middle estimate
Single-sided non-EEA	Both legs are known: one counterparty belongs to EEA, another to non-EEA	As in the middle estimate	50% - to align the results with BIS methodology, where the transactions between EU and non-EU counterparty are halved	As in the middle estimate
Single-sided EEA	Both counterparties belong to EEA, but the transaction cannot be reconciled with any other transaction in the same TR	0% - disregarded due to doubts about the data quality (since it should be matched in theory)	50% - assuming that the amount is duplicated (another leg is reported but cannot be paired)	100% - assuming that the amount is NOT duplicated (another leg is not reported and for that reason cannot be paired)
Single-sided unknown	One counterparty belongs to EEA, another is unknown and the transaction cannot be reconciled with any other transaction in the same TR	0% - disregarded due to doubts about the data quality (since another counterparty is not known)	50% - assuming one of the two cases above for single-sided transactions	As in the middle estimate

Note: EEA refers to European Economic Area. See also footnote 3.

Looking at the EU aggregated data based on EMIR and the results from the BIS OTC survey (Figure 4), the figures are of similar magnitude, although the BIS figure is somewhat below the lower bound for EMIR data, both in terms of notional and market values. The difference can be partially explained by: (i) geographical coverage (all entities in EMIR, only largest dealers and selected countries in BIS); (ii) product coverage (all credit derivatives in EMIR, only CDS for BIS); (iii) reporting basis (residency basis for EMIR, consolidated basis for BIS); and (iv) different reference periods (end-January 2014 for EMIR and end-June 2014 for BIS).

**Figure 4: Comparison of EMIR OTC credit data with BIS OTC survey, EUR billion**



While the higher EMIR coverage with respect to geographical and product coverage (points i and ii) are consistent with the results, these conceptual differences should not have a significant impact on the comparability of the two statistics given the high concentration in the CDS market (i.e. the largest

dealers cover most of the CDS trades) and the fact that CDS products constitute the lion share in the credit derivatives asset class. On the other hand, the effect of differences in the reporting basis (point iii) – consolidated reporting for the banking groups in BIS versus inclusion of only EU resident entities under EMIR – is more difficult to quantify without access to the data reported by the overseas subsidiaries of the EU entities. Regarding the last point on the reference period, our analysis will be updated with the next release of the BIS OTC survey to align both statistics. The long time lag of the BIS publication illustrates the superiority of the EMIR data with respect to the timeliness.

## 6. Conclusions and way forward

The introduction of the daily reporting obligation for the derivatives market on the transaction-by-transaction basis is a challenging undertaking by the global authorities, due to its enormous scale and complexity of the financial products. We have described some of the biggest challenges faced in Europe in the first stage of the data reporting. In particular, the experience shows that clear and detailed guidance on the fields to be reported to the TRs (e.g. code-lists, formats) can enormously improve the consistency and quality of the collected data. This is the case in Europe, when the continuous clarifications via ESMA's Questions and Answers document (ESMA, 2015b), and in particular the introduction and the enforcement of the validation rules on each reported transaction, help gradually improve the quality.

However, in order to make the dataset suitable for a wide range of analytical studies, further harmonisation of reporting is needed. There are currently two work streams which should greatly contribute to further improvements: i) ESMA's planned update of reporting standards and ii) the CPMI-IOSCO work on the global guidelines on the harmonisation of the derivatives reporting, including the global identifiers such as UTI and UPI. Moreover, detailed guidance to TRs on the final data provided to authorities is another key step to enable authorities to aggregate and analyse the data across TRs.

Drawing from the European experience will be useful during the process of the global data aggregation, given that the double-reporting obligation resembles the situation at the time of trying to match cross-border trades, where the two counterparties report the transaction to two different TRs. The challenge of the reconciliation of two sides of the trade highlights the importance of fully standardised rules, not only within jurisdictions, but also globally. In this respect one of the most crucial elements is the globally applicable and unique UTI as well as clear rules about its generation. The development of the UPI and harmonization of other data elements are further key steps to obtain data of high quality.

## References

**European Parliament and Council (2012).** Regulation (EU) No 648/2012 of the European Parliament and of the Council of 4 July 2012 on OTC derivatives, central counterparties and trade repositories, Official Journal of the European Union (available [here](#)).

**European Securities and Markets Authority (2015a).** ESMA supervision of Credit Rating Agencies and Trade Repositories, Annual report 2014 and work plan, 16 February 2015 (available [here](#)).

**European Securities and Markets Authority (2015b).** Questions and Answers, Implementation of the Regulation (EU) No 648/2012 on OTC derivatives, central counterparties and trade repositories (EMIR), 31 March 2015 (available [here](#)).

**F. Maxwell (2014).** Esma trying to avoid commodities rerun of forex spat. Risk magazine, 30 Sep 2014 (available [here](#)).