FINBOURNE

Propellant.digital | ADAMANTIA

# UK and EU
# Fixed Income

## Data Quality Project
## Initial Findings and Outline Report

19 September 2024

## Executive summary

**ESMA recently stated that "data quality and data-use function in tandem, with one naturally fueling the other, in a virtuous cycle."[1] Following on from previous work that both FINBOURNE and Adamantia have published, we decided to work together with Propellant Digital to present observations on the "critical role"[2] that data plays in the fixed income markets.**

We have been supported in the formulation of this report by AFME, BVI and EFAMA to bring key stakeholders' perspectives on prioritising actions to improve the data quality cycle.

In this report, we collated work that FINBOURNE completed through analysis of over 153 million transaction records - 16.3 million of which were fixed income transactions - for the period 1 January 2023 to 31 March 2024, as well as 12 million FIRDS records.

We have assumed, based on the latest regulation, that any CTP will receive the same data that FINBOURNE has been accessing:

" ...however, to avoid an undue burden on data contributors, it is appropriate to align, to the extent possible, the requirement that data contributors make data public with the requirement to contribute data to the CTP."

We noted that ESMA recently (in their Data Report[3]) outlined the Data Quality Indicators ("DQI") checks they complete for MIFIR (bonds).

**We present this analysis in the form of 8 topics and, using feedback from the stakeholders, determined the following 3 priorities for EU and UK policy makers and regulators:**

1. There need to be better defined standards, alongside improved oversight of actual market practice, to ensure the transparency data supplied is in a **comparable, complete** and **correct** state;

2. There could easily be a **full audit** of the current FIRDS and FITRS databases to ensure accuracy and consistency of the records in advance of any CTP go-live. The opportunity exists to ensure these databases, once audited, become a **'golden source'** of reference data for the market and the public going forward;

3. There need to be clearer steps **to further improve data quality** - additions to the (current) DQIs, publication of reports to demonstrate performance against those DQIs and continual improvement cycles.

While other issues are highlighted and explained in this report, we collectively believe that, if EU and UK policy makers and regulators address these 3 priority issues, it will meaningfully improve data quality and "move the dial" before any bond CTP comes into view.

---

**1** ESMA 2023 Report on Quality and Use of Data 11 April 2024 ESMA12-1209242288-852
2023 Report on Quality and Use of Data (europa.eu) *page 5*

**2** *see above*

**3** *see above*

## Introduction

The revision of the Market in Financial Instruments Regulation ("MiFIR") package includes the creation of a Consolidated Tape ("CT") to enhance the development, attractiveness, and competitiveness of European capital markets. It will provide a consolidated view of the available liquidity of traded securities in the EU by aggregating core market data into a centralized infrastructure, accessible to all types of investors at reasonable cost.  Data quality is critical to develop consistency and a reliable CT.[4]

We have identified three fundamental pillars of data quality: transaction data consistency, reference data integrity and policy frameworks. Each of these elements plays a key role in maintaining high data quality standards, which is why our study focuses on both the transaction and reference data.

Transaction data is at the core of daily financial operations and must meet stringent consistency criteria to ensure accurate recording. Reliable transaction data supports financial reporting and decision-making, reduces errors, and enhances trust in financial systems. For trades executed outside of trading venues ("OTC" transactions), Approved Publication Arrangements ("APAs") are the entities responsible for collecting, processing, and publishing trade data.

Reference data integrity involve managing essential information such as the security identification, classification and issuance data. This data is gathered in the Financial Instruments Reference Data System ("FIRDS") and the Financial Instruments Transparency System ("FITRS") databases under ESMA's regulatory framework. FIRDS and FITRS collect data from trading venues, including Regulated Markets, Multilateral Trading Facilities ("MTFs"),
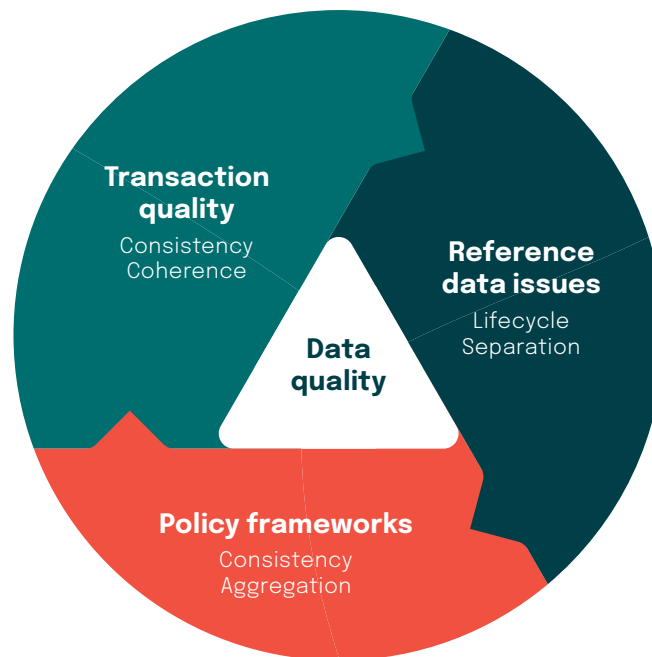
and Organized Trading Facilities ("OTFs"); and APAs, ensuring unique identification through identifiers like the International Securities Identification Number ("ISIN"). The data must be submitted in line with ESMA's standards and formats.

Effective management of this data throughout its lifecycle is crucial to avoiding discrepancies and ensuring all stakeholders have access to accurate and up-to-date information. Thus, policy frameworks are important as they provide guidelines and standards to help organizations maintain data integrity and comply with regulatory requirements.

This document delves into both of these data types, examining the challenges and best practices associated with maintaining high-quality data in the financial sector.



Transaction quality
Consistency
Coherence

Reference data issues
Lifecycle
Separation

Data quality

Policy frameworks
Consistency
Aggregation

---

**4**  "Due to the varying quality of data, it is difficult for market participants to compare such data, which deprives data consolidation of much of its added value."
*Paragraph 19*  REGULATION (EU) 2024/791 (February 2024)

# Transaction data issues

## 1. Replication needs 'de-replication'

We have accessed data from 7 data groups (i.e. venues that included APA, MTFs and OTFs) to provide a dataset that represents some 80 - 85% of Fixed Income records.

In terms of capturing the transaction data, some trading venues and APAs required constant monitoring to record the data (i.e. most venues provide transaction data as a 'discrete' file updated every 1 - 2 minutes that contains transactions on a 15-minute delayed basis).

However, these 'files' continually aggregate previously reported transaction records - along with new transactions data - thus creating replicas in the data available through the day. This requires a significant 'de-replication' effort to retrieve the correct view of transaction data - we distilled some 153 million records to create 16.3 million fixed income records - representing 11% of the total – for the 15-month period.

The total records vs. de-replicated records per day is presented in the graph below:

**Graph 1** Effects of replication
**Total vs de-replicated record counts - by month**



Note: Other data groups, not displayed in our graphic, also report duplicated transactions.
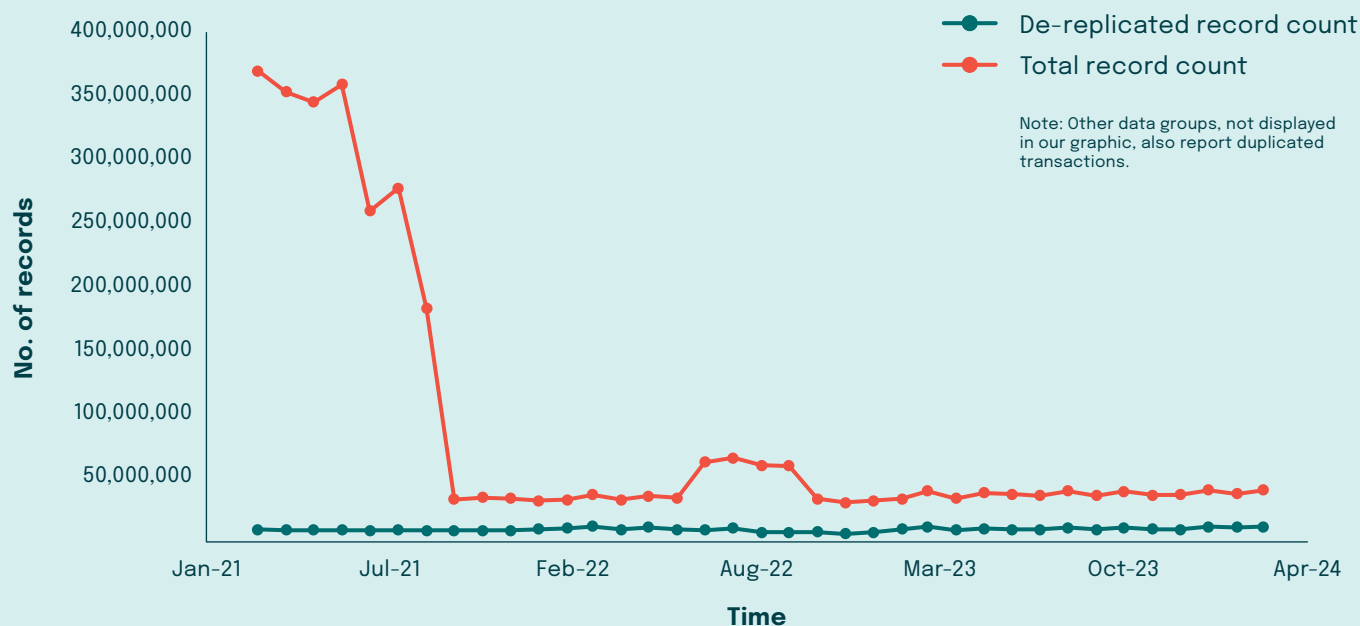
Two venues in particular accounted for most of the replicated information with each release of transaction records including all the previously reported transactions over the past 48 hours (though limited to the last 1,000 records) - rather than simply adding the incremental transactions, a so-called 'delta' file.

We can see that the earlier data (2021) originally had no restriction on the number of records being reported and when the 1,000 records (maximum) was introduced that there was a dramatic decrease in the number of replications.

This replication creates a significant barrier to simply aggregating transaction data and being able to create usable functions.

It also creates significant technical issues, in that software needs to be deployed to:

- Identify the replicated transactions – using the Transaction ID

- Separate the replicated transactions from the actual (underlying) transactions

- Ensure record keeping can be maintained appropriately i.e. the 'de-duplicated' files can be retained

- Ensure there is sufficient 'space' for the storage of data

**Proposal**

Ensure a consistent standard and market for transmission protocols to avoid the need for complex technology to 'de-replicate' trades.

## 2. Missing fields

While reporting transactions, certain specific, basic fields are required to be filled out to ensure that all necessary information is provided under RTS 2 - Price, Notional Amount, currencies in which the transaction is traded (Price and Notional Currency) and Trade date/time.

However, one of the key issues identified is missing data in fields that should, in almost all cases, be populated. This omission can hinder the accuracy and completeness of the data since missing fields lead to incomplete data sets.

In terms of the 16.3 million records, we find the following patterns of missing fields:

**Table 1** Missing fields (heatmap)

| Detailed fields | Q1 2023 | Q2 2023 | Q3 2023 | Q4 2023 | Q1 2024 |
|---|---|---|---|---|---|
| Total Records | 3,001,497 | 3,002,990 | 2,981,611 | 3,288,580 | 3,984,424 |
| Price missing | 0.08% | 0.06% | 0.04% | 0.04% | 0.03% |
| Notional Amount missing | 12.5% | 14.4% | 13.9% | 13.9% | 10.2% |
| Trade date/time missing | 9.0% | 10.1% | 9.9% | 9.8% | 8.4% |
| Price currency missing | 39.1% | 40.5% | 38.7% | 52.2% | 62.9% |
| Notional currency missing | 0.7% | 0.8% | 1.1% | 1.8% | 2.2% |
| Quantity missing | 18.4% | 21.0% | 20.2% | 22.6% | 25.2% |
| Price = '0' | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Notional Amount = '0' | 0.4% | 0.4% | 0.0% | 0.0% | 0.0% |
| Quantity = '0' | 0.5% | 0.4% | 0.0% | 0.0% | 12.5%** |
| Price reported is negative | 0.04% | 0.04% | 0.03% | 0.02% | 0.02% |

*this does include volume deferrals but there are also unflagged records with no 'Notional Amount' reported
*490,634 'Quantity = 0' records (98% of the total) comes from a single venue - potentially a change to process issue following the MiFID II guidance update[5].

**Proposal**

Simple DQI checks and a formal (and transparent) system of remediation and improvement is required alongside periodic guidance or practice updates.

## 3. Irregular transactions

The notional amount and the issuance are among the required fields when reporting transaction data.

One issue previously highlighted was 'irregular' or 'outsized' records for fixed income assets – where the notional amount was multiple times the outstanding issuance size.

Those records are incoherent and are therefore not relevant because they distort the true picture of market activities.

In this regard, when we assessed our latest data series, we found:

**Table 2** Irregular trades

| 'Irregular' transactions (notional > FIRDS records issuance) | Number/amount | % total |
|---|---|---|
| Total records | 68,655 | 0.4% |
| Volume (€) | 3,797,248,119,261 | 6.3% |

When we examine these transactions in more detail, we can see the following transactions represent the 'most irregular' transactions:

**Table 3** Examples of irregular trades (notional > FIRDS issuance)

| Publication date | ISIN | Notional Amount | Times issuance (x)* |
|---|---|---|---|
| 10 January 2023 | XS1891127851 | $38,750,000,000 | 8,811 |
| 11 August 2023 | XS2296644094 | $400,000,000,000 | 2,738 |
| 17 July 2023 | XS1890756189 | $69,300,000,000 | 502 |
| 29 August 2023 | XS2297687431 | $52,100,000,000 | 435 |
| 02 May 2023 | IDG000020801 | IDR 1,424,765,773,304,200 | 9 |

*based on estimated FX rates

### Proposal

Any data quality assurance process needs clarification of the process to be followed and procedural framework related to responsibilities regarding enforcement.

## 4. Price and notional notation

New guidance for the application of MiFiD and MiFIR were issued by ESMA to ensure that financial reporting practices are consistent and standardized across the market earlier this year[6].

The guidance suggested that price notations for bonds should be in percentage format (PERC) (with parameters suggested for the use of alternative notations such as YIEL (yield), MONE (monetary amount) and BAPO (basis point)[7]. Additionally, the guidance sought to clarify the definition of Notional Amount – the "face value" of a bond is the price that the issuer pays at the time of maturity, this is also referred to as "par value."[8]

---

**6** ESMA74-2134169708-6870 Manual on post-trade transparency (europa.eu)

**7** *above page 94*

**8** *above page 92*

However, a review of fixed income reveals the practice remains at levels seen prior to the RTS change:

**Table 4** non-'PERC' Price Notation

| Bonds with non-PERC price notation | 1 January – 31 December 2023 | 1 January – 31 March 2024 |
|---|---|---|
| % total records | 3.5% | 3.2% |
| % distinct ISINs | 36.8% | 31.1% |

We can identify to specific, selected examples:

**Table 5** non-'PERC' Price Notation

| Venue* | Trade/publication date | ISIN | Price | Notional | Price notation |
|---|---|---|---|---|---|
| #1 | 09/01/2024 | DE000NRW0N26 | 105.118 | 54,000,000 | **MONE** |
| #2 | 18/03/2024 | DE000NRW0N26 | 106.08 | 500,000 | PERC |
| #1 | 25/03/2024 | FR0013393774 | 99.41 | 1,988.2 | **MONE** |
| #2 | 14/03/2024 | FR0013393774 | 99.381 | 25,000 | PERC |
| #1 | 23/04/2024 | US871829BS59 | 80.475 | 60,000 | **BAPO** |
| #2 | 06/02/2024 | US871829BS59 | 104.519 | | PERC |
| #1 | 16/01/2024 | US871829BS59 | 82.18728 | 744,632.25 | **SPRD** |
| #1 | 26/03/2024 | US87264AAT25 | 1,754.2 | 5,000 | **BAPO** |
| #1 | 12/03/2024 | US87264AAT25 | 62.204 | 100,000 | **BAPO** |
| #1 | 27/02/2024 | US87264AAT25 | 48.43 | 30,000 | **BAPO** |
| #1 | 13/02/2024 | US87264AAT25 | 57.086 | 54,000 | **BAPO** |
| #2 | 16/04/2024 | US87264AAT25 | 100.079 | 250,000 | PERC |
| #2 | 09/04/2024 | US87264AAT25 | 100 | 25,000 | PERC |
| #2 | 02/04/2024 | US87264AAT25 | 99.888 | 6,000 | PERC |

*the numbers assigned to the venues are consistent across these examples

From these examples, we can glean three issues (as identified in the table above):

- Certain venues (the same venues or venue groups are identified numerically in the table) continue to apply their own 'practices' to certain bonds

- Even where the price notation is 'MONE', the actual prices supplied seem, in many cases, to reflect actual percentages

- While some of the Notional Amounts reflect the "par value", others patently do not

A combination of the continuation of these practices by different venues means that effectively, even where the technical feat of consolidating and normalising the date takes place, the information cannot be applied to a continuous data stream as the underlying data points aren't either accurate or comparable.

**Proposal**

Additional DQIs should be added to identify these outliers – initially, on a regular, periodic basis to remediate followed by a real-time, streamed oversight.

# Reference data issues

One of the core elements necessary to make transaction data usable is the ability to link that data into the relevant reference data databases.  However, the reference data needs to be uniform in order to allow users to access a 'golden' source.

## Background

The **Financial Instruments Reference Data System** ("**FIRDS**") provides reference data for all MiFID instruments that are Traded on a Trading Venue ("**TOTV**" and "**TV**") in the EU. Each TV has to submit the relevant reference data when a bond is traded on that venue. ESMA collects data from trading venues and National Competent Authorities ("**NCAs**") and makes it available on its website in accordance with MiFIR requirements[9].

FIRDS has **1,369,611,314** total records and **79,963,938** distinct ISINs (inception – to-date).  Of those, 62,018,868 are debt instrument records while there are 4,401,544 distinct ISINs for debt instruments.

The **Financial Instruments Transparency System** ("**FITRS**") publishes reference data, liquidity, waiver thresholds and quantitative data to help firms assess their trade reporting although the recent MIFIR Consultation Package on RTS 23 proposes changes to the existing FITRS/FIRDS data fields[10].

FITRS has **2,614,826,226** total records relating to **60,357,855** distinct ISINs (inception – to-date).  Of those 55,237,437 are 'BOND' records with only 277,034 'BOND' distinct ISINs.

There are **492 venues** that have submitted records to the FIRDS and FITRS databases since their inception in 2017 and 2018 respectively. The reference data - generally source on a bond by bond (ISIN) basis - contained in the FIRDS and FITRS reporting includes several essential reference components:

- The Legal Entity Identifiers (**LEI**) - the LEI is a 20-character code that clearly and uniquely identifies legal entities participating in financial transactions.

- The classification of the bond (**CFI**) - the CFI is a 6-letter syntax used in the financial services industry for the systematic classification and description of the structure and function of financial instruments

- The Total Issuance Notional Amount (**TINA**); and

- The **maturity date** of the bond

It is important to note that there is no single 'golden' record rather a series of submissions from each TV where a bond has been traded.  However, this data should be consistent regardless of the venue.

---

## 5. Revisions

We know that venues realise that there are mistakes made in the data. Each time they amend a FIRDS record, they create a 'revised' record. We have examined the revisions (modifications) applied to FIRDS records (by ISIN).

We found high levels of revisions to the FIRDS records.

This is relevant as most users access the data presented on a day-to-day, 'as-is' basis and cannot easily see what has changed from prior records leading to potential confusion or inconsistencies for users.

For some venues, the number of revisions can reach up to 111 revisions per ISIN:

**Graph 2** Revisions per venue
**Number of revisions per ISIN (average)**
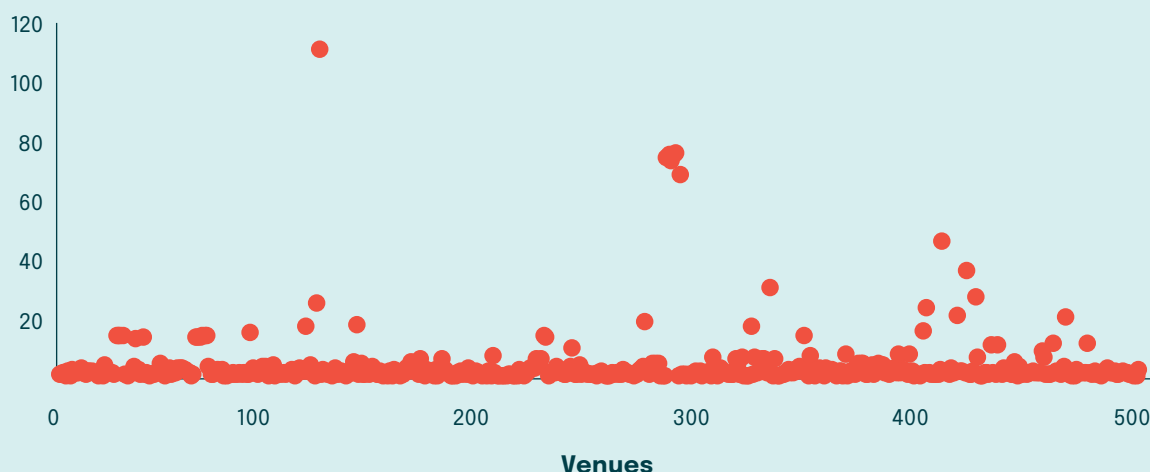


## Proposal

Examination of the main causes of revisions and engagement with the venues around a new protocol.

Notwithstanding the revision of records and activity of the venues, we still find that there are fundamental errors in the FIRDS/FITRS datasets.

## 6. Overall classification

The recent transparency Consultation Paper[10] proposes that bonds are organised in 3 sub-asset categories, namely:

- 'sovereign and other public bonds'

- 'corporate, convertible and other bonds; and

- 'covered bonds'

As we noted above, the categorisation of bonds is via a CFI code that is available in both FIRDS and FITRS. In FIRDS there are a total of over 300 classifications of debt instruments. In FITRS, in terms of bonds there are 6 (which broadly mimic the proposals in the CP).

---

**10** ESMA74-2134169708-7241_CP_Package_on_the_MiFIR_Review_-_RTS_2__RCB_
and_Reference_Data.pdf (europa.eu)

## FINBOURNE | Propellant.digital | ADAMANTIA

**Table 6** FIRDS versus FITRS CFI 'bond' classifications

| FIRDS | | | FITRS | |
|---|---|---|---|---|
| **CFI schema** | **Second CFI letter** | **Remaining 4 CFI letters (# of attributes)** | **Categories** | **Sub-categories** |
| Debt | Convertible | 30 | Bond | Convertible Bond |
| | Bond | 31 | | Sovereign Bond |
| | Bond with warrants attached | 30 | | Corporate Bond |
| | Medium Term Notes | 30 | | Other Public Bond |
| | Money Market Instruments | 19 | | Covered Bond |
| | Structured (capital protection) | 21 | | Other Bond |
| | Structured (no protection) | 24 | | |
| | Mortgage-backed Security | 29 | | |
| | Asset-backed security | 29 | SFPS | |
| | Municipal bonds | 29 | ETC | |
| | Depository Receipts | 35 | ETN | |
| | Other | 9 | | |

However, a cross-referencing of these categorisations displays the challenge that lies at the heart of the data quality issue - namely, different interpretations of the same rules:

**Table 7** Selected examples of cross-referencing of FIRDS and FITRS categorisations

| FIRDS categorisation | FITRS bond categorisation | | | | | |
|---|---|---|---|---|---|---|
| | **Convertible bond** | Corporate bond | Covered bond | Other bonds | **Other Public bond** | Sovereign bonds |
| Bonds | **19%** | **32%** | **51%** | **3%** | **32%** | **60%** |
| **Convertible bonds** | **60%** | n/m | - | n/m | - | - |
| Bonds with warrants attached | - | n/m | - | n/m | - | - |
| Medium-term notes | 1% | **33%** | **22%** | **10%** | **28%** | **18%** |
| Money market instruments | - | 6% | n/m | n/m | n/m | 1% |
| Asset-backed securities* | 1% | 1% | 1% | n/m | n/m | |
| **Municipal bonds** | - | n/m | | n/m | **26%** | 5% |
| Others (miscellaneous)* | n/m | n/m | n/m | 1% | n/m | 2% |
| Structured instruments (without capital protection) | - | 1% | 2% | 4% | - | - |
| No FITRS **BOND** record | **19%*** | **27%*** | **24%*** | **80%*** | **13%** | **15%** |
| Total number of ISINs | 1,924 | 140,798 | 6,147 | 37,813 | 9,926 | 9,352 |

*a large number (some 75%) of ABS/MBS bonds appear in the 'SFPS' category of FITRS rather than 'BONDS'

When comparing, for example, convertible bonds, we find only a 60% match between the two existing databases.

Some ISINs labelled as 'convertible bond' in FITRS are labelled merely as 'bonds' in FIRDS, and nearly 20% of them are not present in FIRDS at all.

When comparing, for example, convertible bonds, we find only a 60% match between the two existing databases.

Some ISINs labelled as 'convertible bond' in FITRS are labelled merely as 'bonds' in FIRDS, and nearly 20% of them are not present in FIRDS at all.

When we look at examples, in more detail, we can see the level of inconsistency in FITRS and the challenges of cross-referencing categories with FIRDS:

**Table 8** Selected examples of divergence between FITRS itself and cross-referencing to FIRDS

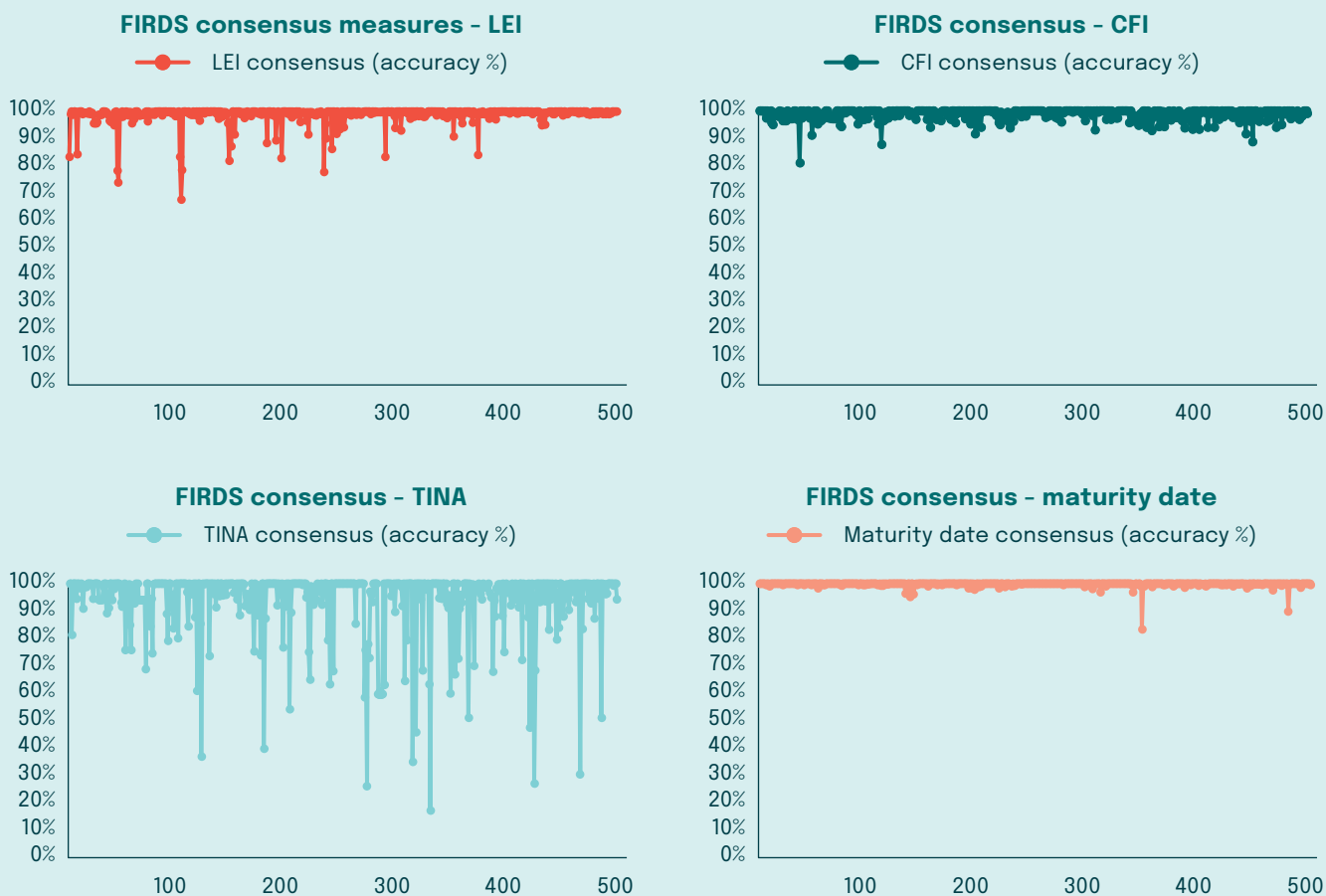| Issuer name | FITRS category | FIRDS CFI category |
|---|---|---|
| Harvest CLO XIV | Convertible bond | **DAVXBR** (a registered asset backed security with a variable rate and amortisation plan with a call feature) |
| Harvest CLO XIX | Corporate bond | **DAVXBR** (a registered asset backed security with a variable rate and amortisation plan with a call feature) |
| Harvest CLO XV | Corporate bond | **DAVXFR** (a registered asset backed security with a fixed rate and a fixed maturity feature) |
| La Banque Postale MTN 2029 | Convertible bond | **DTFNCB** (a bearer, fixed rate medium term senior note with fixed maturity) |
| La Banque Postale MTN 2034 | Public bond | **DTFNFB** (a bearer, fixed rate medium term senior, fixed maturity note with a put feature) |
| La Banque Postale MTN 2031 | Corporate bond | **DTVSFB** (a bearer, variable rate medium term secured note with a fixed maturity) |
| Morgan Stanley EMTN 2028 | Convertible bond | **DTZNFR** (a registered zero rate/discounted, senior medium term note with a fixed maturity) |
| Morgan Stanley EMTN 2025 | Corporate bond | **DTVNFR** (a registered variable rate, senior medium term note with a fixed maturity) |
| Neste Corporation EMTN 2028 | Convertible bond | **DTFNGB** (a bearer, fixed rate medium term senior note with fixed maturity and call feature) |
| Neste Corporation EMTN 2025 | Corporate bond | **DTFNGB** (a bearer, fixed rate medium term senior note with fixed maturity and call feature) |
| HSBC Bank PLC MTN 2033 | Convertible bond | **DTZXFB** (a bearer zero rate/discounted, fixed maturity medium term note) |
| HSBC Bank PLC MTN 2047 | Corporate bond | **DTZUGB** (a bearer zero rate/discounted, unsecured medium term note with a fixed maturity and call feature) |

## 7. Consensus ... or not

It is difficult to access a central 'golden source' when, as we noted above, each record is submitted individually by a TV when the security is traded on its venue. FINBOURNE cross-referenced the FIRDS records - by ISIN - to develop a 'consensus' value. FINBOURNE identified the most frequently occurring LEI for each ISIN across all FIRDS records and deemed this the 'correct' (consensus) data. This modal frequency - the one that appears most often - is

benchmarked against each of the individual ISIN on each individual TV record on FIRDS to display this 'consensus'.

When we had examined the consistency of the records submitted by venues to FIRDS for the relevant time period across the key data fields - LEI, the CFI, the maturity date and the 'TINA' fields - we could observe the following:

**Graph 3** 'Consensus' measure of the four key reference datapoints by TV (alphabetic)

### FIRDS consensus measures – LEI
━●━ LEI consensus (accuracy %)



### FIRDS consensus – CFI
━●━ CFI consensus (accuracy %)



### FIRDS consensus – TINA
━●━ TINA consensus (accuracy %)



### FIRDS consensus – maturity date
━●━ Maturity date consensus (accuracy %)



We can see that there are various degrees of 'consensus' around the different fields. While some venues have a 100% accuracy rate, others deliver information that is only partially correct. Besides, the TINA field shows the lowest levels of adherence to the 'consensus' figure (note this does not infer the issuance amount is correct, rather than it is

the mode - see below) and creates a serious issue where issuance size is used as the basis for any deferral regime.

When we look at the venues (across all asset classes), we can see that there is a small number of large TVs that represent a large number of recorded ISINs:

**Table 9** Ranking of TV consensus data

| Venue | % total ISIN | LEI consensus | CFI consensus | TINA consensus | Maturity date consensus |
|---|---|---|---|---|---|
| Top 5 TVs | 77.40% | 99.95% | 99.13% | 99.99% | 99.96% |
| Next 100 TV | 21.70% | 98.64% | 99.49% | 99.14% | 99.92% |
| Remaining 386 TV | 0.90% | 98.80% | 98.38% | 94.05% | 99.78% |

## Proposal

A data cross-referencing and cleansing exercise should be completed prior to any bond CT 'go-live' to ensure a 'golden source of data.

## 8. TINA ... consistently wrong

The Total Issued Notional Amount ('TINA') is a key record that FIRDS maintains.  The issue of a dynamic outstanding issuance amount is a core part of the fixed income market - with bond programmes, especially in government bonds, being 'tapped' through their lifecycle.

To focus on this issue, we highlighted inconsistencies in the FIRDS records for the French 2.50% OAT maturing on May 30, 2030 (Obligation Assimilable du Trésor) from the official French Treasury issuance records[11].  Although this bond is recorded on 62 venues currently, we tracked the 12 venues that recorded the original issuance to measure the consistency of those records:

**Graph 4**
**France 2.50% 30 May 2030 OAT issuance history**
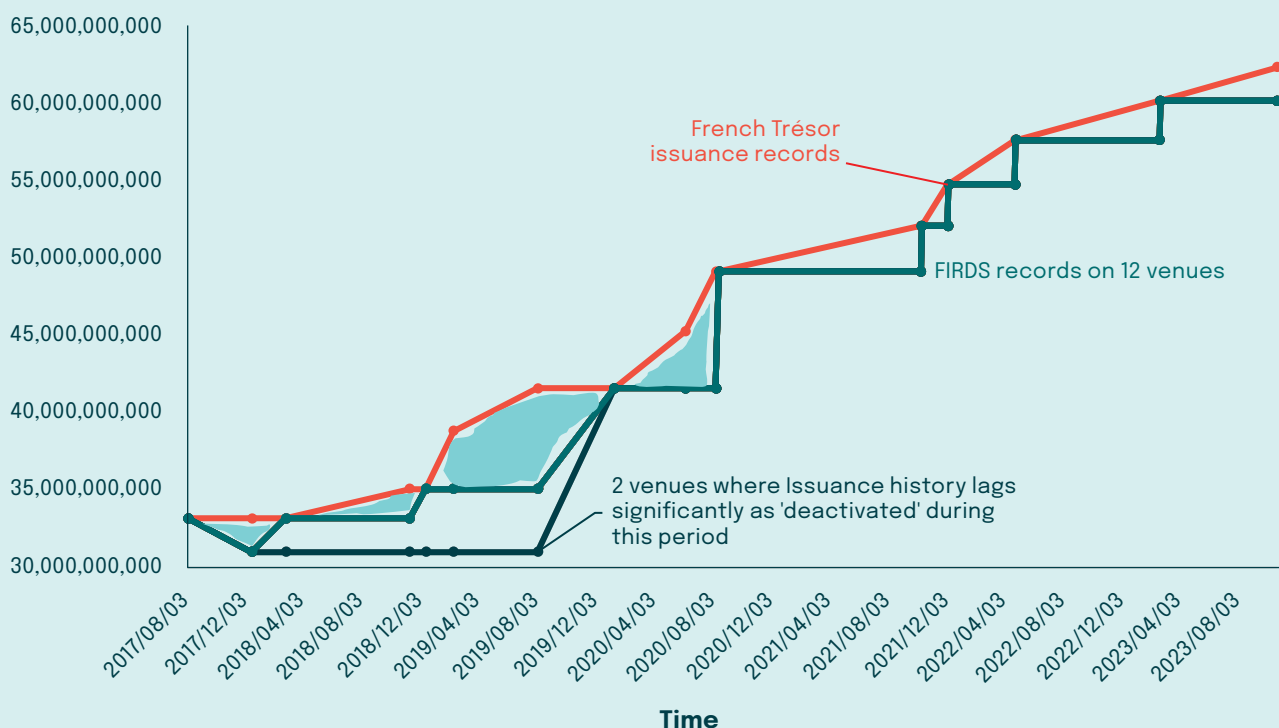


- **Red line:** indicates the official issuance records from the French Trésor. This line shows a steady increase over time, reflecting the actual issuance amounts.

- **Teal line:** represents the FIRDS records from those 12 venues.

- **Dark line:** highlights two specific venues where the issuance history significantly lags behind the others, indicated as 'deactivated' during certain periods. This means that these venues did not report updates promptly or accurately during these times.

- **Blue shaded area:** illustrates periods where there is a notable discrepancy between the FIRDS records and the French Trésor records.

This visualization emphasizes the importance of timely and accurate updates in FIRDS records to maintain data integrity and consistency with official issuance records.

FINBOURNE has examined[12] the 'official' aggregate US Treasury issuance (based on Treasury Direct sources[13]) against the aggregate contained in FIRDS ('consensus' amounts):

**Graph 5**
**Aggregate issuance volume – TD records compared to FIRDS**
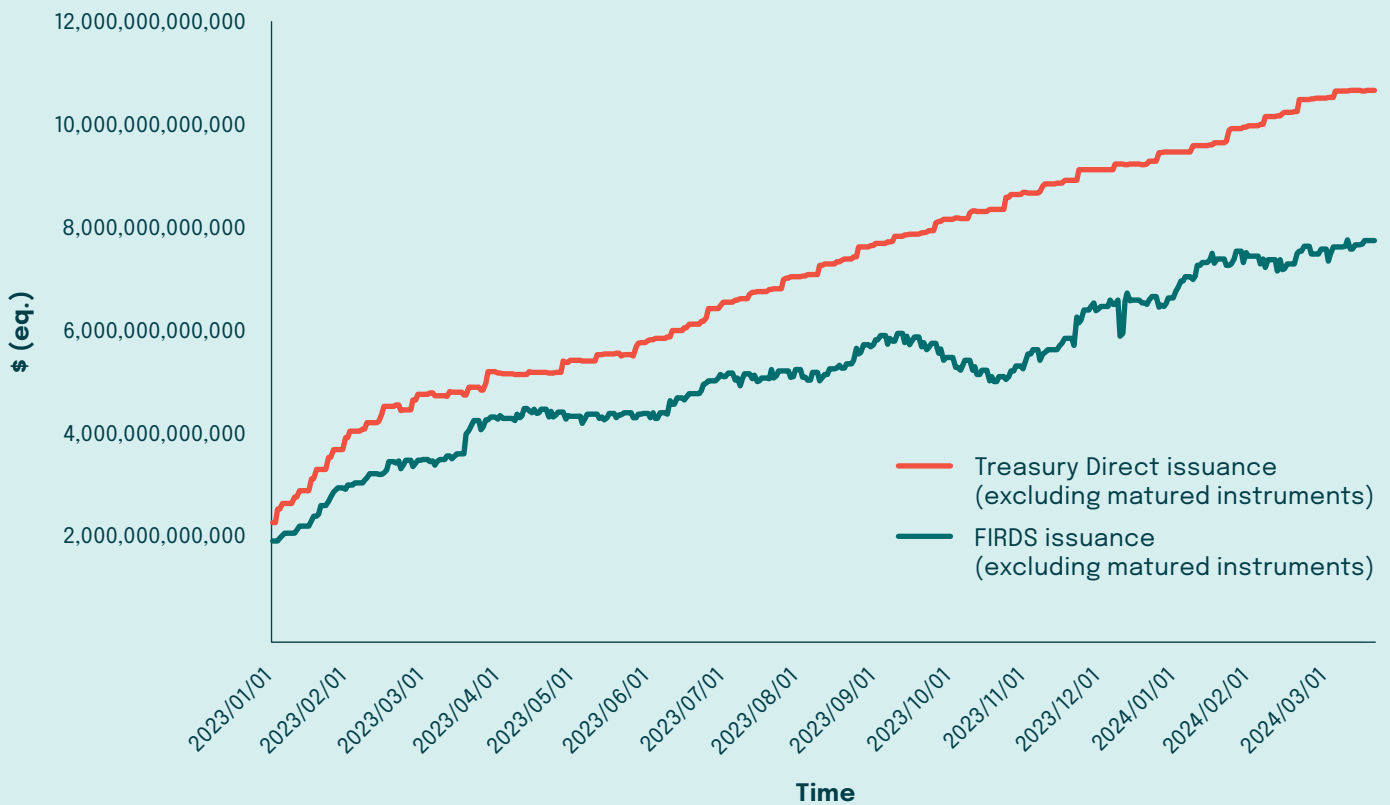


**Proposal**

Once again, the linkages can be made to the official sources available for sovereigns. For corporate and other issuances, the exercise is more distributed but applying a 'consensus' could, at least, highlight exceptions.

---

# Appendix 1
## Assessment of ESMA Data Quality Indicators[14]

ESMA outlined some of the 'data quality indicators' they monitor at EEA level to derive trends in data quality developments across data sets. They note that they calculate indicators for each jurisdiction and distribute them to the NCAs. It is worth noting that while the list below provides an insight into the data quality work for MiFIR (FITRS), it should not be assumed that these are the only data quality aspects that regularly analysed by data quality supervisors, although there is no detail given for an FIRDS DQIs.

| Name | Area | Scope | Objective | Comment |
|---|---|---|---|---|
| **BOND_1 – Bond Outliers** | Accuracy | Trading volumes | Checks the presence of daily trading values that lie far away from their reference distribution | Definition of "far away" needs clarification for trading volumes - see Appendix 2 below where ESMA present data but not clear how this definition fits the graphic |
| **BOND_2 – Notional Amount** | Accuracy | Reference data | Checks instruments with a reporting total notional amount below 10 thousand EUR | From the analysis, it is the larger trades that skew the data not the smaller ones … |
| **BOND_3 – Classification** | Consistency | Reference data | Checks whether the CFI reported for the bond is in line with its classification | Not clear how this is deployed and within FITRS there are inconsistencies and even more when cross-referenced with FIRDS CFI |
| **BOND_4 – Bond Type** | Consistency | Reference data | Checks whether the classification of the bond type is consistent between the various MICs, and whether bonds with sovereign issuers are correctly classified as such | ESMA observed that there is misclassification ("over 5% of bonds" - see Appendix 2) and our analysis - based on a consensus - shows that for some key elements (i.e. TINA) that inconsistency is even greater |
| **NQT_1** | Completeness | Quantitative data | Checks whether all expected data for non-equity instruments has been correctly submitted | Not clear what these checks are, how they operate or the outcome |
| **NQT_2** | Completeness | Reference data | Flags instruments that are missing reference data in FITRS | Not clear that cross referenced to FIRDS or how any 'consensus' is managed |
| **NQT_3** | Consistency | Reference data | Checks whether the MiFIR identifier reported by different venues for the same instruments are consistent | Not clear whether cross-referenced with FIRDS to ensure consistency |
| **NQT_4 – Maturity dates** | Consistency | Reference data | Flags instruments whose maturity date is inconsistent between FIRDS and FITRS reference data, or whose maturity data is not in line with the instrument's own characteristics | Maturity date is one of the four key elements but not the only one that needs to be compared |
| **NQT_5 – Outliers** | Accuracy | Stress test results | Check for implausible values in input factor | Once again, no detail of "implausible values" or detail of any remediation steps |

As a comparison, we looked at the Data Quality Report that GLEIF produces (Data Quality Reports – GLEIF Data Quality Management - LEI Data – GLEIF) where they apply 12 criteria linked to 85 requirements with a further 38 elements that are expected to be satisfied.  They also produce regular 'snapshots' of adherence to the standards to aid transparency and inform the market:

## Appendix 2
## ESMA Report on Quality and Use of Data (11 April 2024) §4.2 MiFiR Data

The report looked 9 different DRSPs, 7 with an APA license and 6 with an ARM license.

There are several issues that this report highlighted:

- "substantial disparities" and "disparities in transaction patterns" were identified but not explained or remediated

- 'rejection rates' at ARMs and APAs remained low (although one APA had to manage the level down during 2023) and seems to be in or around 5%

- ESMA claim that "three APAs account for more than 98% of published non-equity transactions"

The recommendations included:

1. "harmonisation of field naming conventions, the consistency in the reporting of flags, and the provision of daily aggregate CSVs would significantly improve the usability of published transparency reports" - they included some sample code to enable easier access to data as well as signaling that Level 2 RTS changes would occur in the future

2. Increased co-operation to permit transaction and transparency reporting comparison (the authorities are looking at this aspect)

3. ESMA felt confident that there is a "reasonable level of price consistency"

Chart 23
Dispersion of log median price between the same instruments reported via MiFIR transaction and transparency data
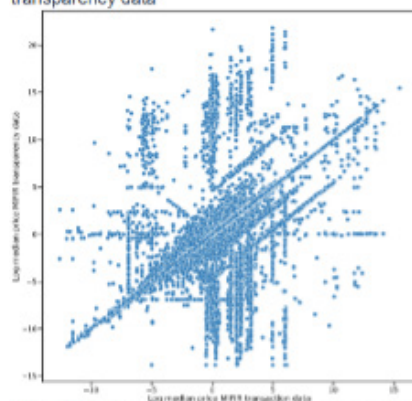


Note: Graph shows log median price reported on individual instruments and venues. Where the log median prices are equivalent between the two datasets the points scatter along 45-degree line. The further away the points are from the 45-degree line, the higher difference in the log median price between the two datasets. ESMA used clustering machine learning approach to identify features behind the clustering behaviour. Often the clusters appear to be driven by reporting by a specific executing and or submitting entity and or trading on a specific venue.
Sources: MiFIR transaction and transparency data reported in November 2022.

4. ESMA felt confident that "the completeness of quantitative and reference data submitted for both equity and non-equity instruments remains stable and of good quality " although the "consistency between the instrument classification coming from multiple venues" needs "some improvement"

5. In terms of bond type classification, ESMA states that "over 5% of bonds available for trading in 2023 on European markets are still affected by this issue"

Chart 24
Misclassification of bonds – RCA MIC vs MIC



6. ESMA also highlight (without specifying) "misreported quantitative data" where "major data corrections" were made following interventions (although there remain issues at 1 APA) as well as "an incorrect reporting practice" at "multiple" OTFs which had resulted in "duplication of transactions in the transparency system"

## Appendix 3
## References to Data Quality Issues – EU

| Issue | Source |
|---|---|
| Data quality and data use function in tandem, with one naturally fuelling the other in a virtuous cycle. The key developments concerning data quality in 2023 are reported in Section 4 for each of the dataset [sic] in scope. The data quality engagement frameworks rely on the systematic identification of data quality issues via automated dashboards, the communication of the most relevant issues to NCAs and follow-up actions, until resolution | **2023 Report on Quality and Use of Data** **11 April 2024 (ESMA12-1209242288-852)** **page 5** |
| 4.2.4.2 Data quality – to ensure accurate results in the publications mentioned above, ESMA performs a series of data quality checks on its transparency database. These checks are performed quarterly and shared with the NCAs and DRSPs with the goal of correcting any misreports, and they encompass a series of data quality indicators such as completeness, consistency, and accuracy. Moreover, during the year 2023 a new Data Quality Engagement Framework was approved, allowing ESMA to concentrate on the most impacting issues while reducing the workload for the NCAs in the correction process. | **page 29** |
| One of the main data quality issues on securities data is the misalignment in the classifications of bonds provided by different market participants. To be specific, the issue comes from trading venues providing conflicting values of BOND TYPE (RTS2 field 9) for the same instrument. ESMA has been tackling this issue with the help of the NCAs for the past years, which has helped the data quality to improve significantly. However, over 5% of bonds available for trading in 2023 on European markets are still affected by this issue. | **pages 29/30** |
| Another data quality issue in transparency data is related to misreports of quantitative data. ESMA has been working to correct these issues, both by improving the methodology for detecting anomalous values, and by collaborating with NCAs and DRSPs. Currently, the main known issues come from DRSPs and OTFs. Over the past years, ESMA contacted multiple supervised DRSPs questioning the figures reported to transparency, and helped them identify issues in their reporting process, which led to major data corrections. Some issues are however still outstanding, in particular linked to a major APAs. | **page 30** |
| Under those frameworks, the joint efforts and engagements of the industry, NCAs and ESMA has led to significant improvements of data quality across the board. Yet, there remains room for further improvement of data quality, as well as signals of data quality deterioration on certain dimensions, which ESMA will continue monitoring. | |

## References to Data Quality Issues – UK

| Issue | Source |
|---|---|
| MiFID II rules do not appear to be producing the data market outcomes we would expect from well-functioning markets in some asset classes, including on high-quality data [….] – a CT for bonds is dependent on these changes being in place for the CT to be valuable; working with market participants to achieve greater standardisation; and supervisory work with APAs to ensure they are paying adequate attention to data quality issues. | **Consultation Paper CP23/33 Consultation on Payments to data providers and forms for Data Reporting Services Providers including Policy Statement for the framework for UK consolidated tape (CP23/15) § 2.25** |
| We proposed adding two new requirements: […..] requiring the CTP to report to us every six months with observations about data quality, to help us in seeking to improve data quality | **§ 6.41** |
| In making our proposals on data quality, it was not our intention to suggest that the CTP would filter the data given to it by data providers before publication. [….] Our intention is that the CTP should feedback to data providers on any potential issues it identifies with the data they are sending so that there is an opportunity for any issues to be dealt with | **§ 6.52** |

## Appendix 4
## Trade Flags

| Acronym | Definition | Explanation |
|---------|-----------|-------------|
| AMND | Amendment Flag | When a previously published transaction is amended. |
| DUPL | Duplication Flag | When a transaction is reported to more than one APA. |
| CANC | Cancellation Flag | When a previously published transaction is cancelled. |
| COAF | Consecutive Aggregation Flag | Supplementary deferral allowing for an indefinite period without full post-trade transparency |
| PNDG | Pending Flag | When a transaction or price is currently not available but due to become available. |
| LGRS (LIS) | Post-Trade Large in Scale Transaction Flag | Transactions where deferred publication is permitted on the basis that they are large in scale compared with normal market transactions. |
| TPAC | Package Transaction Flag | A trade composed of several components/legs. |
| ILQD | Illiquid Instrument Flag | Transactions executed under the deferral for instruments for which there is not a liquid market. |
| FULJ | Full Details Flag | Individual transactions which have previously benefited from aggregated publications because of their status as non-equity instruments that are not sovereign debt. |
| FULV | Full Details Flag | Transactions for which limited details have been previously published about why an individual transaction has been given an extended time period of deferral for four weeks. |
| FULF | Full Details Flag | A transaction whereby limited details have been published about the value and average daily turnover of the transaction. |
| FULA | Full Details Flag | Individual transactions for which aggregated details have been previously published. |
| LMTF | Limited Details Flag | A supplementary deferral which requires additional information. |
| DATF | Daily Aggregated Transaction Flag | A supplementary deferral which requires additional daily aggregation information. |
| VOLO | Volume Omission Flag | A supplementary deferral allowing for an extended period without full post-trade transparency. |
| VOLW | Volume Omission Flag | Transactions for which limited details are published and for which the publication of several transactions in aggregated form for an indefinite period of time will be allowed. |

## Supporters

AFME (Association for Financial Markets in Europe) promotes fair, orderly, and efficient European wholesale capital markets and provides leadership in advancing the interests of all market participants. AFME represents a broad array of European and global participants in the wholesale financial markets. Its members comprise pan-EU and global banks as well as key regional banks,

brokers, law firms, investors and other financial market participants. AFME participates in a global alliance with the Securities Industry and Financial Markets Association (SIFMA) in the US, and the Asia Securities Industry and Financial Markets Association (ASIFMA) through the GFMA (Global Financial Markets Association). For more information, please visit the AFME website: **www.afme.eu**

BVI represents the interests of the German fund industry at national and international level. The association promotes sensible regulation of the fund business as well as fair competition vis-à-vis policy makers and regulators. Asset managers act as trustees in the sole interest of the investor and are subject to strict regulation. Funds match funding investors and the capital demands of companies and governments, thus fulfilling an important macro-

economic function. BVI's 116 members manage assets of some EUR 4 trillion for retail investors, insurance companies, pension and retirement schemes, banks, churches and foundations. With a share of 27%, Germany represents the largest fund market in the EU. BVI's ID number in the EU Transparency Register is 96816064173-47. For more information, please visit **www.bvi.de/en**

EFAMA is the voice of the European investment management industry, which manages over EUR 30 trillion of assets on behalf of its clients in Europe and around the world. We advocate for a regulatory environment that supports our industry's crucial role in steering capital towards investments for a sustainable future and providing long-term value for investors.

Besides fostering a Capital Markets Union, consumer empowerment and sustainable finance in Europe, we also support open and well-functioning global capital markets and engage with international standard setters and relevant third-country authorities. EFAMA is a primary source of industry statistical data and issues regular publications, including Market Insights and the authoritative EFAMA Fact Book.

More information is available at **www.efama.org**

# FINBOURNE ⊫

## About

FINBOURNE's solutions deliver an interconnected network of functionality and data that enables the investment community to better serve clients in a constantly evolving market.

Its investment management solutions and cloud-native data management platform ensure that investment and operations teams can increase revenue, reduce costs, and better manage risk across the investment life cycle.

## Get in touch

## Propellant.digital

### About

Propellant Digital is an award-winning technology company specialising in Fixed Income data analytics. Our founders are trusted specialists in transparency and trading data, passionate about making bond, swap and ETF markets more transparent.

Propellant Digital delivers practical, responsive technology that transforms this data into actionable insights. We help global and regional banks, asset managers, quant hedge funds, trading venues, regulators, and industry associations to analyse transparency and trading data in near real-time. Our unique analytics cloud platform leverages the most extensive Fixed Income dataset, providing users with the insights they need to improve trading strategies, streamline workflows, and make informed trading decisions.

### Get in touch

## ADAMANTIA

### About

ADAMANTIA is an independent management consulting firm specialized in the financial services and based in Paris.

Our Capital Markets team provides consultancy services to leading financial institutions across all segments of the Capital Markets activities. Our combination of deep industry knowledge, skills and experience enables us to meet our clients' challenges and drive their business changes successfully, by providing unique and adaptive solutions to their positioning.

### Get in touch