

# FINsights: CTP Whitepaper Series

#1: Tackling consolidation and aggregation

Double, Double, Toil and Trouble 1
December 2021



# **Executive Summary**

Following last week's announcement by the European Commission of its' 2021 Capital Markets Union package, we believe the developments leading to the creation of a Consolidated Tape (CT) for European capital markets are moving in the right direction.

But while the starting gun has been fired, there are still several vital steps needed to support this market mechanism: the design of an operational framework, the technological infrastructure, and a commercial model to be created - in collaboration with the market - across the four defined asset classes.

While the above concerns will be addressed in the on-going political debate, as a technology provider, fusing together domain expertise and deep technology, our focus is to resolve the data quality issues - a basic yet significant challenge.

To this end, we engaged extensively with various market stakeholders over the last 18 months to gain an insight into the key issues, barriers and use cases for a CT. These included authorities, regulators, industry and representative bodies, cloud technology providers, sell-side and buy-side participants.

All stakeholders identified data quality as a need to be solved, given that the quality of the data input will determine the effectiveness of any CT.

However, we don't believe this analysis needed to wait until a decision has been made on the form of any CT. If anything, starting this process will improve the quality and efficiency of the CT, while it is being developed, and contribute to the resiliency of the end-product.

In early Spring 2021, we commenced our own analysis, using publicly available, post-trade transaction data from a number of the largest trading venues and APAs. To-date, we've examined some **59 million** transactions across equities, ETFs, fixed income and derivatives.

We found several reasons why the available data - in its current form - could not immediately support a CT for market use. These issues can be broken down into three overarching component parts:

- consolidation and aggregation
- consistency
- coherence.

To effectively support visibility of liquidity in the market, we believe that the best way forward is to make these transaction records visible and transparent and, following on from our <u>recently shared insight into the data quality issues</u>, we will be launching a series of whitepapers, deep-diving into these three components.

In this **first whitepaper**, we tackle the issue of **Consolidation and Aggregation**, which forms a significant barrier for market participants today.

We plan to work with the members of our newly created Design Council to recognise, manage and problem-solve for these underlying barriers. This will enable a CT that is fit for purpose, meets the needs of all users and delivers FINBOURNE's core mission of lowering the cost of investing and promoting greater transparency in the market.



# Transaction records' data

# Overview of the public 'portfolio'

Our analysis comprises post-trade transparency records covering all asset classes, from the largest trading venues and APAs. This data, collected from 1 March 2021, formed part of the preliminary workings that FINBOURNE have conducted, in order to prepare for the creation of a **post-trade** CT. The data is publicly available and is provided in a variety of formats, on a delayed basis.

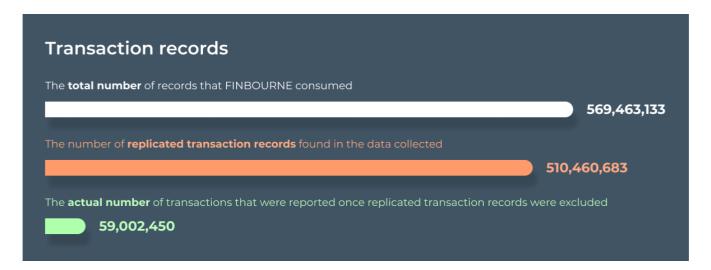
## Transaction data source details

The data sources that FINBOURNE accessed were, as identified in the ESMA Annual Statistical Report 2020i, the largest venues for the **equities**, **ETF** and **bond** markets. We also included other sources, both to deliver a substantial and objective sample and for comparison purposes:

Typo	Jurisdiction	Data Group #1	Data Group #2	Data Group #3	Data Group #4	Data Group #5	Data Group #6
Туре	Jurisalction						
APA	EU	✓	✓	✓		<b>√</b>	✓
	UK	✓	✓	✓			✓
MTF	EU			✓			✓
	UK						✓
OTF	EU			<b>✓</b>			✓
	UK			✓	✓		<b>✓</b>

For analysis purposes, we also linked the transaction records to the EU's **Financial Instruments Reference Data System ('FIRDS') database**, which covers the publication, collection and processing of additional data, to support the MiFIR transparency regime.

The high-level **overview** is as follows:







# The Three C's of CT data

The data quality issues that have arisen in relation to transaction reports, can be broken down into the following component parts, to aid the design of a series of structured solutions:



In general, we observed issues across the following areas:

- Consolidation and aggregation challenges create barriers for market participants unless they have technical SMEs to handle the myriad of codes, formats and transaction reporting conventions.
- Consistency of data is another area of concern, where we saw the same fields for transactions completed in
  different ways, leading to inconsistent treatment that required remediation. For example, we see transactions
  with the same ISIN and of similar sizes reported with inconsistent deferral flags.
- Coherence of data is the main area of concern where we saw incorrect data reported in a manner that either
  distorted any aggregates or averages of data, or led to incorrect/incoherent output as the transactions records
  are not, at present, normalised.



# Consolidation and aggregation

# What are the issues?

In this section, we break down the main technical hurdles that we found in our examination of the transaction records and explore the barriers that exist to 'simply' aggregating those transaction records and the challenges any CT Provider would face in building a platform.

# Technical issues

#### 1. Double, double ... Replication needs de-duplication

- In terms of capturing the data, some trading venues and APAs required constant monitoring to record the data (i.e. data 'grabs' every 5 minutes). However, these 'grabs' often continually aggregated existing transaction records, along with new transactions data, requiring a significant 'de-duplication' effort, as transaction records are replicated through the day.
- One venue in particular, accounted for the majority of replicated information with each 15 minute release of transaction records including all the previously reported transactions - rather than simply adding the incremental transactions, a so-called 'delta' file.
- To recap, FINBOURNE consumed 569,463,133 records for 59,002,450 transactions. This replication created a significant barrier to simply aggregating data and being able to create usable insights.

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.

#### 2. Toil ... Publication format

#### FINBOURNE found that:

A number of the larger venues published the information in JSON:

JSON is an open standard file format and data interchange format that uses human- readable text to store and transmit data objects consisting of attribute–value pairs and arrays.



#### **Example:**

```
[{"id": 2103091334410106283, "publishEntity": "TRADES", "active": true,
"counrtyOfIssue": "GB", "currency": "GBP", "instrumentCodeType": "ISIN",
"instrumentDesc": "UNILEVER ORD", "emsaAssetClass": "EQUT", "esmaAssetSubClass":
"SHAR", "executionVenue": "XOFF", "priceType": "MONE", "pubType": {"desc": "FULL
DETAILS", "messageCode": {"code": "L0006", "title": "Full Pub", "primaryMessage":
"Full trade details have been published.", "secondaryMessage": "Publication ID: %s",
"name": "FULL_PUB"}, "category": "PUBL", "eventSource": "FPUB", "code": "PUB_FULL"},
"publishSource": ____, "publishTs": "2021-03-09T13:34:41.758179Z", "quantity":
250, "instrumentCode": "GB00B10RZP78", "transactDateTime": "2021-03-
09T13:34:26.532Z", "price": 39.24, "transactionType": "NEW", "tic":
2103091334410106283, "stringPublishTs": "09-Mar-2021 13:34:41.758179",
"stringTransactDateTime": "09-Mar-2021 13:34:26.532000", "stringTic":
"2103091334410106283", "stringId": "2103091334410106283"}
```

However, the majority of trading venues and APAs published in CSV format:
 a CSV file is a Comma Separated Values file. All CSV files are plain text files, can contain num

a CSV file is a Comma Separated Values file. All CSV files are plain text files, can contain numbers and letters only, and structure the data contained within them in a tabular, or table, form i.e. it appears like an Excel file.

#### **Example:**



- Formats can change without notification or warning: we observed, during the time that we have been tracking the records, that the formats from some trading venues and APAs' CSV files changed. This required technical adjustment to earlier records, to ensure that they could be 'persisted' i.e. the columns changed which required adjustments to reconciled with the previous records.
- What is important to recognise here, is that the sheer number of formats and templates a CT Provider would need to identify and account for, is not a one-off exercise. It requires significant monitoring and maintenance, given the entities publishing post- trade data can change their structure or delivery mechanisms at any given time.
  - 3. Trouble... formatting

#### FINBOURNE found that:

the formatting of the trade fields themselves, across the venues, differed in that their order was varied:

#### **Examples:**



We can observe from these (CSV) files, that simply combining or aggregating these files will 'mix' data which prevents meaningful analysis because:

- The order of the fields differs across the venues.
- In the case of Venue D, the cells have to be dis-aggregated (and the order differs).



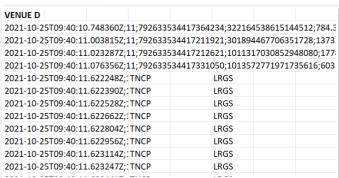
- In order to ensure its use, these files require some technical adjustment to ensure that the fields are reconfigured, reordered and remapped.
- Trade flags were represented differently in some cases, the fields were contained in one data field while in other cases, there was an individual field for each flag.

#### **Examples:**

<b>VENUE A</b>	
FLAGS	SUPPLEMENTARY DEFERRAL FLAGS
	FULJ, ILQD, SIZE
TPAC	FULV, ILQD
TPAC	FULV, ILQD, SIZE
TPAC	FULV, ILQD, SIZE
TPAC	FULV, ILQD, SIZE

VENUE B					
LMTF_FLAG	FULF_FLAG	DATF_FLAG	FULA_FLAG	VOLO_FLAG	FULV_FLAG
FALSE	FALSE	FALSE	FALSE	FALSE	FALSE





Once again, there was **variance in application of the RTS2 methodology** by the trading venues and APAS to the transaction records:

- VENUE A applied a principle of 'primary' and 'supplemental' deferral fields.
- VENUE B issued a separate field for each flag.
- VENUE C applied FIX's MMT flag convention.



- VENUE D overrode the 'single string' data field with 'random' flags i.e. the relevant flag appears but in no particular order.
- We found, in general, that a small number of feeds were over FIX, while most were not- introducing further complexity over formats and connectivity.
- One trading venue/APA had some date configurations that were inconsistent with the ISO standard.

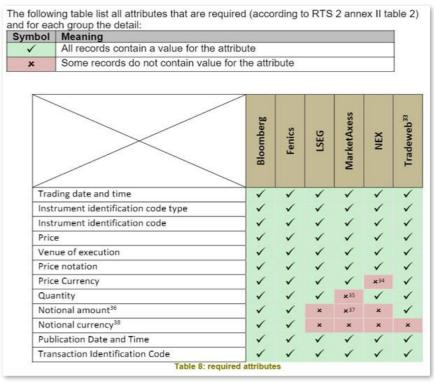
#### Example:

TRADE DATE EXECUTION TIME	PUBLICATION DATE
2020-10-13T10:34:59.000Z	14/10/2020 05:00
2020-10-12T15:16:43.000Z	14/10/2020 05:00
2019-06-19T16:06:19.000Z	14/10/2020 05:00
2019-06-19T08:08:35.000Z	14/10/2020 05:00
2020-09-14T07:51:52.531Z	14/10/2020 05:00

In the first column, the date is shown using the 'YYYY-MM-DD' standard, whereas the second column does not

 While another trading venue/APA had trade fields populated in a manner that caused difficulty with aggregation - for example, a PRICE field with 20 decimal points: "99.29596743210102030405"

### 4. Coverage of trade fields



source: etrading software, "MIFID 2 post trade data" (October 2020)

#### FINBOURNE found that:

- We could, also, access all the required data fields from the venues.
- However, the **formats** created to represent those fields inhibit the ability to easily absorb, aggregate and consolidate transaction records.
- In most cases, the fields were completed, although there were significant issues with the quality of the data provided.



While there were some records that did not contain values, there was no obvious patterns to the missing values, and it is, thus unlikely for there to exist simple solutions to remedy the problems.

#### 5. Parsing

#### FINBOURNE found that:

 A number of numeric fields contained non-numeric data e.g. "N/A" which can have the effect of slowing run-time.

#### **Examples:**



VENUE B - 'N/A' is used where there should be numbers.

#### **VENUE D**

49628127296;1435;3;2021-10-25T09:40:00.797000Z;ISIN;GB0009252882;MONE;GBX;;;SINT;2021-10-25T09:40:00.833462Z;0;;;;XXXX;
507391451248;370.75;2436;2021-10-25T09:40:00.689000Z;ISIN;JE00B4T3BW64;MONE;GBX;;;SINT;2021-10-25T09:40:00.694292Z;0;;;XXXX;
L07018821744;281.1;15;2021-10-25T09:40:00.795000Z;ISIN;GB00BY9D0Y18;MONE;GBX;;;SINT;2021-10-25T09:40:00.827291Z;0;;;XXXX;
349255497792;281.7;15;2021-10-25T09:40:00.797000Z;ISIN;GB0005603997;MONE;GBX;;;SINT;2021-10-25T09:40:00.838543Z;0;;;XXXX;
706646192240;255;17;2021-10-25T09:40:00.796000Z;ISIN;GB00BF8Q6K64;MONE;GBX;;;SINT;2021-10-25T09:40:00.842659Z;0;;;XXXX;

The use of non-numeric data presents unique problems:

 VENUE D - uses semi-colons between the single stream data which requires intricate re-parsing and configuration of the records.

#### 6. Publication method

A number of the trading venues and APAs provided access via an API:

an Application Programming Interface, enabling companies to open up their applications' data and functionality to external third-party developers.

This allows services and products to communicate with each other and leverage each other's data and functionality through a documented interface. However, we noted that, at least, one venue using API connectivity changed the API without notification, adjusted their documentation without subsequent communication or reference, which led to the 'dropping' of data and required manual adjustment to re-connect to the API.

#### 7. Lineage

 During aggregation it was not possible for FINBOURNE to determine whether any data quality or completeness were driven by issues at source or had been introduced during processing. For example, FINBOURNE could not determine if any data issues are related to the initial transactions being reported by Systematic Internalisers (SIs), OTFs or MTF or whether it is the result of data transformation by an APA.

# Practical issues

8. Self-aggregation



For any market participants that might consider the opportunity to acquire and maintain the information internally, practical considerations include:

- Handling the CSV/JSON issue.
- Being able to identify and eliminate replicated records.
- Normalising the formatting points.
- Correcting and eliminating the parsing issues.
- Although it may be well known, Excel cannot handle more than 1 million rows and its' ability to filter data, at that level of volume, is limited.

#### 9. Making the data usable

While the RTS 2 data provides 24 fields of detail, it does not include relevant basic data, for example, the issuer name. To make sense of the transaction data, FINBOURNE needed to connect with the FIRDS database to ensure a basic level of utility.

# What's next

As presented, there are numerous elements within the realm of consolidation and aggregation that will impact the efficiency and accuracy of an eventual CTP. Understanding and solving for data quality issues, such as duplication, lineage, formatting and API connectivity will be key to ensuring a robust CTP that is fit for purpose.

Our analysis is the first step in the journey ahead, and while we have identified some pressing concerns, we know that together with the collaboration of our Design Council members, and the use of our cloud-native SaaS technology, we can respond to the current data challenges, ahead of the creation of a CT.

Alongside this market engagement, we will be launching our next whitepaper in the series, providing more detail on the second of the three component issues identified: consistency of CT data. In this whitepaper, we will explore the complexities around the way in which fields are completed across trading venues and APAs and the remediation needed to overcome this issue.

# FINBOURNE's Design Council

FINBOURNE is inviting market participants with an interest in the mechanics of developing a Consolidated Tape to join its Design Council.

The first meeting will take place in December 2021 and the Council will meet through June 2022.



#### **FINBOURNE's Design Council - benefits to Members**

In return for providing views and expertise, the Design Council offers Members the following:

- an opportunity to shape the implementation in a way that could make a CT more relevant for their institution
- a forum where the elements of the operation of operation and governance of any CT entity can be discussed,
   explored and evolve an open environment where issues of data quality can be raised and examined in sufficient
   detail
- where the data discussed can be used by the Members either internally or at other fora or bodies where they participate
- where agreed by Members, analysis of data can be presented to other bodies in the form of whitepapers to help to develop the concept of market data standards or principles
- access to beta releases of the FINBOURNE CT Platform ("CT Platform") and other relevant material and services, including training
- exposure to thought leaders and the latest cloud technology in this space.

# Get in touch

Tell us what you think. If you'd like to learn more about the CT journey, or have your say in the CT

Design Council, get in touch with us at <a href="mailto:ctp@finbourne.com">ctp@finbourne.com</a>

# **ANNEX 1 - CSV FILE COMPARISONS**

	Venue A	Venue B	Venue C	Venue D
1	TRADING_DATE_AND_TIME	EXECUTION TIME	distributionTime;sourceVenue;instrum entld;transactionIdentificationCode;mi fidPrice;mifidQuantity;tradingDateAnd Time;instrumentIdentificationCodeTyp e;instrumentIdentificationCode;priceN otation;priceCurrency;notionalAmount ;notionalCurrency;venueOfExecution;p ublicationDateAndTime;transactionTo BeCleared;measurementUnit;quantityI nMeasurementUnit;type;venueOfPubli cation;mifidFlags	Execution Time
2	INSTRUMENT_ID_TYPE	PUBLICATION DATE		Execution Time UTC

# FINBOURNE |=

3	INSTRUMENT_ID	VENUE OF EXECUTION
4	PRICE	INSTRUMENT ID TYPE
5	VENUE_OF_EXECUTION	ISIN
6	PRICE_NOTATION	PRICE
7	PRICE_CURRENCY	MONE
8	NOTATION_OF_QUANTITY_ IN_ MEASUREMENT_UNIT	PERC
9	QUANTITY_IN_ MEASUREMENT_UNIT	YIEL
10	QUANTITY	ВАРО
11	NOTIONAL_AMOUNT	PRICE NOTATION
12	NOTIONAL_CURRENCY	PRICE CURRENCY
13	PUBLICATION_DATE_AND_ TIME	QUANTITY
14	VENUE_OF_PUBLICATION	QUANTITY TYPE
15	TRANSACTION_ID	QUANTITY IN MEASUREMENT UNIT
16	TO_BE_CLEARED	QUANTITY NOTATION MEASUREMENT UNIT
17	LMTF_FLAG	NOTIONAL AMOUNT
18	FULF_FLAG	NOTIONAL CURRENCY
19	DATF_FLAG	EMISSION ALLOWANCE TYPE
20	FULA_FLAG	PUBLICATION ID

21	VOLO_FLAG	ORGINAL ID
22	FULV_FLAG	TRANSACTION TO BE CLEARED
23	FWAF_FLAG	TRANSCATION COUNT

Publication Time
Publication Time UTC
CCY
Venue of Execution (MIC Code)
Venue of publication (MIC Code)
Price notation
Transaction to be cleared
MMT Flags
Transaction ID Code
Trade type
Agreement Time and Date
Agreement Time and Date UTC
Price
Volume
Buyer
Seller



24	FULJ_FLAG	FLAGS
25	IDAF_FLAG	SUPPLEMENTAR Y DEFERRAL FLAGS
26	VOLW_FLAG	NO SUBMITTED TRANS (TPCC)
27	COAF_FLAG	SUBMITTED NOTIONAL (TPCC)
28	BENC_FLAG	SUBMITTED NOTIONAL 2 (TPCC)
29	ACTX_FLAG	NOTIONAL CURRENCY 2 (TPCC)
30	LRGS_FLAG	LNO TERMINATED/RE DUCED TRANS (TPCC)
31	ILQD_FLAG	TERMINATED/RE DUCED NOTIONAL (TPCC)
32	SIZE_FLAG	TERMINATED/RE DUCED NOTIONAL 2 (TPCC)
33	TPAC_FLAG	REPLACEMENT NOTIONAL (TPCC)
34	XFPH_FLAG	REPLACEMENT NOTIONAL 2 (TPCC)
35	CANC_FLAG	SUB ASSET CLASSES (TPCC)
36	AMND_FLAG	REPLACEMENT SUB ASSET CLASSES (TPCC)
37		REPLACEMENT CURRENCIES (TPCC)
38		NO REPLACEMENT TRANS (TPCC)



# **ANNEX 2 - JSON FILE COMPARISONS**

#### Example 1

```
[{"id": 2103091334410106283, "publishEntity": "TRADES", "active": true,
"counrtyOfIssue": "GB", "currency": "GBP", "instrumentCodeType": "ISIN",
"instrumentDesc": "UNILEVER ORD", "emsaAssetClass": "EQUT", "esmaAssetSubClass":
"SHAR", "executionVenue": "XOFF", "priceType": "MONE", "pubType": {"desc": "FULL
DETAILS", "messageCode": {"code": "L0006", "title": "Full Pub", "primaryMessage":
"Full trade details have been published.", "secondaryMessage": "Publication ID: %s",
"name": "FULL_PUB"}, "category": "PUBL", "eventSource": "FPUB", "code": "PUB_FULL"},
"publishSource": "publishTs": "2021-03-09T13:34:41.758179Z", "quantity":
250, "instrumentCode": "GB00B10RZP78", "transactDateTime": "2021-03-
09T13:34:26.532Z", "price": 39.24, "transactionType": "NEW", "tic":
2103091334410106283, "stringPublishTs": "09-Mar-2021 13:34:26.532000", "stringTic":
"2103091334410106283", "stringId": "2103091334410106283"}
```

#### Example 2

```
ł
          "id": 2103020747280102964,
          "publishEntity": "TRADES"
           "active": true,
          "currency": "USD",
          "instrumentCodeType": "OTHR",
"instrumentDesc": "Deliverable forward (DF)",
"emsakssetClass": "CURR",
           esmaAssetSubClass": "DLFR",
          "executionVenue": "XOFF"
           "priceType": "MONE",
           "pubType": {
    "desc": "FULL DETAILS",
                "messageCode": {
                     "code": "L0006",
"title": "Full Pub",
                     "primaryMessage": "Full trade details have been published.", 
"secondaryMessage": "Publication ID: %s",
                     "name": "FULL_PUB"
                "category": "PUBL",
                "eventSource": "FPUB",
"code": "PUB_FULL"
           "publishSource":
           "publishTs": "2021-03-02T07:47:28.069670Z",
           'quantity": 3650000,
          "instrumentCode": "DERVCURRFORW",
"transactDateTime": "2021-01-29T10:01:23Z",
          "flags": "FULJ, ILQD", "notional": 424991.7158
           "notionalCurrency": "US_
           "price": 0.11644,
           transactionType": "NEW",
          "tic": 2103020747280102964,
          "cleared": "N",
          "stringTransactDateTime": "29-Jan-2021 10:01:23.000000",
"stringId": "2103020747280102964",
          "stringPublishTs": "02-Mar-2021 07:47:28.069670",
          "stringTic": "2103020747280102964"
    }
```



#### Example 3

{"messageId":"posttrade","sourceName":"FRA","symbol":"DE000A0N6260","tickActionIndicator":"I","mmtMarketMechanism":"7","mmtTradingMode":"U","mmtModificationInd":"-","tickId":0,"lastTradeTime":"2021-10-25T15:19:01.4972344702","lastTrade":14704e-2,"lastCty":0,"currency":"EUR","instrumentIdCode":"I","quotationType":1,"distributionDateTime":"2021-10-25T15:19:01.5071780002","instrumentId":"DE0000A0N6260","transIdCode":"300000000000003361761016351751414972344700000000000","executionVenueId":"-","notionalCurrency":"EUR","mmtAlgoInd":"-"}



#### **DISCLAIMER**

This document has been prepared by FINBOURNE Technology Limited ("FINBOURNE") on an 'as is' basis. It provides general background information regarding FINBOURNE's activities and is supplied for information purposes only. Nothing in this document should be regarded as an invitation, inducement or recommendation to engage in investment activity (a financial promotion) as defined in section 21 of the UK Financial Services and Markets Act 2000 ("FSMA") and the information contained in this document is not intended to be an offer to buy or sell any interest in any investment.

Information set forth herein is only a summary of certain information as at the time this document is provided. FINBOURNE does not make any representation or warranty (express or implied) as to the accuracy or completeness of the information provided and it shall have no liability in relation to the content or its use. You should not place any reliance on any statements contained herein and such statements are subject to change by FINBOURNE and uncertainty and contingencies outside FINBOURNE's control. The information, materials and opinions contained herein are not intended to constitute legal or other professional advice and should not be relied on, nor treated as a substitute for specific advice relevant to particular circumstances.

This document is the property of FINBOURNE and any reproduction, dissemination or re-distribution of this document or the information herein without FINBOURNE's prior written consent is forbidden.

© 2021	FINBOURNE Technology Limited. All rights reserved.
-	
i	EU Securities Market, ESMA Annual Statistical Report (18 November 2020) ESMA-50-165-1355

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

To discover more about FINBOURNE and learn how our

solutions can drive growth, increase control and improve data

access, contact us below

finbourne.com/info@finbourne.com

+44 (0)20 3880 1307

**FINBOURNE Technology** 

North America: 666 3rd Avenue, New York, 10017 UK: 1

Carter Lane, London, England, EC4V 5ER

Singapore: 790, Level 7 Capital Square,

23 Church Street, 049481