Trade Repositories

Response to a Consultation Paper

Issued by the Committee of European Securities Regulators

Version Number:	Final
Issue Date:	05 November 2009

1. Summary

This document contains the response by LCH.Clearnet Ltd to CESR's Consultation Paper CESR/09-837 "Trade Repositories in the European Union".

The stated purpose of the Consultation Paper is to collect stakeholders' views on trade repositories, including their functions, data and transparency requirements, their location and the legal framework. This is prefaced within the context of the number of European entities active in financial markets, the European origin of many underlying instruments, the number of contracts denominated in European currencies, the volumes involved and the need to satisfy the information needs of EU regulators. It is also stated that trade repositories should aim to foster transparency, thus supporting the efficiency, stability of and orderly functioning (i.e. avoidance of abusive behaviour) of financial markets.

Within these statements, and in points 3. "Functions and Characteristics of a Trade Repository", and 4. "Availability of Data by Trade Repositories", there is a clear requirement for the provision of information flows but there is also a second requirement to use the data for various operational downstream processing to support risk reduction, operational efficiency and cost saving benefits to individual participants and the market as a whole.

This operational requirement should be separated from the need for information as the two have fundamentally different business drivers and design paradigms:

	Operational	**Information**
Requirements are	market specific and driven by participants	cross-market and driven by regulators
Data design focus is	transaction throughput	analytical flexibility
Process flows are	rigid and streamlined	loose and matrix
Online data retention	transactions should only be retained until complete to avoid impacting processing times	transactions should be retained for at least 7 years to provide historical analysis
User Access and Change management	tightly controlled and managed with monthly / quarterly release cycles	managed but changes can be effected on a weekly cycle and users can be given access to design/build their own reports
SLA's / DR	real-time / instantaneous	daily / 48 hour

If a single solution is used for both operational and information requirements one or more of the following could result:

- operational processing of transactions is significantly impaired (volume and speed)
- reporting and/or analysis is very restricted (breadth and depth of queries is limited)
- the delivery of new reporting requirements may take months rather than weeks due to competing for priority against operational change requirements
- reporting analytics such as data slicing or drilling will not be possible

Historically, where one system is used for both operational processing and information analysis, the operational needs take precedent and the analytical capability is invariably so compromised that it is unable to provide the information required in a timely manner.

2. Answers to Questions

2.1 Functions and Characteristics

Do you agree with the functional definition of what constitutes a trade repository?

If the primary function of the Trade Repository ("TR") is to provide information and analysis to supervisors, the public, the markets and potentially the submitting participants themselves, then no we do not agree with the functional definition as it should not also be used for trade life cycle event management, reconciliation or downstream trade processing. Operational processing and information reporting systems are fundamentally different in their operation and design. Operational processing systems need to focus on the processing of single strings of data whereas information reporting systems focus more on large sets of data

Both operational processing and information reporting systems will contain a database to store information, however the design of these data stores are diametrically opposed. It is important to understand the makeup of a database to fully appreciate the fundamental differences between operational processing systems and information reporting systems:

- Tables – these are used to store data. Operational databases perform better when these tables are small (i.e. contain fewer columns and rows of data) so amendments / additions / deletions can be written to the table quickly. As a result of the need for smaller tables operational databases tend to have lots of tables, potentially hundreds.

- Joins – these link tables together. Joins serve a number of purposes but the two main reasons for consideration here are data integrity and query impact.
    - Data Integrity - by joining two tables together it is possible to restrict data entry in one based on the data specified in another. For example, if a table of client information includes a country field, data in the field could include "UK", "uk", "u.k.", "U.K.", "gb", "g.b.", "G.B." or any other mixture of characters that the data source determined was appropriate to represent the United Kingdom. By creating a separate table of country codes and joining this to the country code field in the client table, data entered in the client table can be restricted to a single code. Without this join, a query run against the client table using UK would not provide the correct information from a business perspective as clients with GB in the field would be ignored.
    - Query impact – by linking tables, data in different tables can be bought together and more complex cross-table queries can be run.

  Whilst joins bring data integrity, which is needed for data quality, analysis and/or reporting, they can dramatically slow the processing of data – if a table includes joins then whenever data is added, deleted or updated the data contained in the field against which the join exists has to be validated through the join.

- Indexes – these are constructed within or across tables on multiple key data facts. The primary purpose of an index is to speed up the searching and retrieval of data from tables – if an index consists of two columns of data then when searching for entries that meet two criteria, one in each column, the data only has to be "read" once; if there was no index it would have to be read twice. Operational systems invariably don't have indexes as the indexes dramatically slow the processing of

changes / additions / deletions – to add, update or remove data, any affected indexes have to be dropped, the data changes applied and then the index is re-built.

When considering the appropriate design of a database there are a few paradigms to consider:

1. The primary focus for an operational system is processing speed and accuracy, to capture and process transactions as quickly as possible. Therefore, databases supporting operational systems have large numbers of tables, few joins linking the data across these tables, and no indexes. From a reporting perspective, the absence of joins between tables means the data is unlikely to be consistent and reports will likely be flawed and it will be impossible to run queries against the data as it is not possible to link the data together in a meaningful way.

2. To address these shortcomings, the next step would be to "normalise" the data to introduce data integrity / consistency and enable meaningful queries to be run. Without going into the complexities of the different forms of normalisation and data redundancy, normalising a database effectively introduces joins between the tables to enforce data integrity. If an operational processing system included joins between tables then the data being entered into the system would have to be validated across every join as part of its processing and this would slow throughput considerably.

3. As databases get larger, normalisation will impair query performance as queries will be run across larger numbers of tables with potentially many joins. To address this, larger databases, or data warehouses, will denormalise data or consolidate data from multiple tables into as few tables as possible to reduce the number of joins and improve query performance. This design approach is known as a star schema which invariably consists of a few, maybe even just one, central fact table containing pertinent transaction data joined to a number of dimensional tables. One downside of this is a possible threat to data integrity so invariably this is addressed through the introduction of a layer approached within a data warehouse – the staging layer, an ETL layer (extract, transform and load) and a published data layer. The staging layer holds data in its original source format whilst the ETL layer extracts data from this, imposes data integrity through coded rules and then populates the published layer where reports and analysis are conducted.

Of course in practice the above design paradigms are just that, and invariably operational systems do have some level of data normalisation and do produce reports, albeit that these are usually run when the central processing operations are offline or from a replicated system. It is important to note here that running reporting queries across a large number of tables in a live operational database can cause "contention" (where more than one process is trying to access the same data) which could in turn lock the database and prevent both the operational process and reporting process from completing successfully. Although locks can often be avoided the contention can cause a significant delay to the completion of the affected processes.

Designing a database is not as black and white as the above paradigms would suggest but as data volumes grow there is usually a tendency to increase levels of normalisation will slow the performance of operational system, and eventually the volumes will be such that queries

to support reporting will also slow as the system approaches the point where denormalisation is required. Whilst this lessens the importance of the design paradigms it actually increases the need to understand the differences between processing and reporting systems as it is possible to not only compromise one for the other but to end up with a platform that cannot support either.

In addition to the fundamental system design differences there are also a number of other notable differences:

- Data retention – operational systems are focussed on supporting the life cycle of a transaction and therefore to minimise the amount of data contained within tables, transactions are usually archived once complete. Information reporting systems are designed to store huge volumes of data with many years of history to facilitate wider forms of analyses including cyclical trends.

- Data timeliness – operational systems in financial services need to be real-time and need to have real-time failover. Information reporting systems can invariably be slightly more relaxed and are usually updated daily with failover usually based on a recovery from back-up (maybe 24/48 hours).

- Data processing – operational systems have rigid processing workflows to ensure high throughput, accuracy and resilience. Analytical systems have to be flexible to facilitate the myriad of possible data paths associated with the likely possible reporting requirements. It's is also worth noting here that running analytical capability from an operational database can significantly slow the operational processes as queries will inevitably interfere with data processing.

In addition to the above, and in some ways a demonstration of the different focus of the two types of system, the need to establish a sole "official legal record" (so called golden copy) is an operational requirement to enable the two parties to the transaction to agree a single representation of that transaction that then allows it to be processed and completed. For reporting purposes it would be useful to understand the status of a trade, as that could imply certain things about a parties STP capability and operational risk profile, but it would be preferable for the TR to include all trades regardless of their status to ensure the most complete set of trade data possible. For example, there are likely to be some parties in any given industry who do not submit their trades to the TR however parties that executed trades with that non-submitting party may themselves submit trades to the TR and these "alleged" trades provide the opportunity to not only analyse the fuller trade set of the submitting party but also the partial set of trades for the non-submitting party.

What other characteristics of a TR do you consider essential?

There are a number of other characteristics which should be included but essential amongst these are:

Counterparty / Regulatory structure

This is key as it sets out the legal structure of the counterparties within the system, and should at the very least contain a two tier structure covering legal entity and parent group, although it may be necessary to include branches of legal entities. This needs to be established as a matrix against the relevant regulatory body as it may be that any given institution may determine that only the appropriate regulatory entity may view a given group,

legal entity and/or branches trading activity. Without this it will not be possible to disclose the correct level of information to any given user, whether regulatory, the public or users within the said institution.

Data Security

In addition to system/application security such as firewalls to prevent unauthorised access, the TR should also include encryption of the data at the atomic level to ensure that authorised users can only access reports and data that has been authorised. For example, if those institutions submitting data require access then they would only be able to see their own data, and regulators would only be able to see data for institutions for which they are responsible.

Data Transformation Mappings

To ensure that a full audit trail is maintained and data can be validated back to its source, it is imperative that a Source To Target Mapping is maintained. This should show the source of the data, how it has been transformed in the ETL layer and how it is stored in the published layer. This is vital where data is likely to be sourced from multiple systems and the data is to be used for regulatory analysis – if this is not maintained then it may prove impossible to validate the data produced in a report back to its source.

Data Dictionary and Naming Convention

Without a dictionary explaining the data contained within the TR and a consistent business-based naming convention, the ability for end users to define and build reports will be non-existent and a large highly specialised support function with significant key–resource risk will be needed within the operator. Key-resource risk arises because the understanding of the technical solution, the data contained within this and how to deliver reports from these invariably tends to be consolidated in those few key individuals who participated in the design and build of the TR – a full data dictionary and business-based naming convention captures this knowledge within the TR design.

Additionally the standardisation of counterparty naming conventions would provide significant benefit not just to the identification of the correct parties to a transaction within the TR but potentially throughout the entire life cycle of a trade.

Information Accessibility

Access to the TR should be via a secure web portal that allows users to run reports, slice and dice, and /or drill the results for greater understanding, and, ultimately, develop their own reports. This avoids a significant resource build up at the operator and therefore minimises the commercial costs and dependency / resource prioritisation issues that would otherwise arise.

It should be noted here that none of the above requirements for a reporting system would be requirements for an operational processing system, and all would significantly slow the operational system's capability to process transactions.

2.2 Availability of Data

In your opinion, what kind of information should be available to: regulators, market participants and the general public, respectively? Please differentiate by asset class where appropriate.

Whilst it is for the various user groupings to define their requirements – the example on the Consultation (as adapted for other asset classes) seems reasonable – it would be anticipated:

- the general public would be able to reports containing aggregated data at market level without any analytical capability
- regulators would only be able to see reports aggregated at market and counterparty level, albeit at this level only covering those entities for which they regulatory responsibility. The would also be able to analyse these reports through slicing/dicing/drilling down to individual trade level
- parties submitting data would have full access to reports containing their own data

It would also be anticipated that these reports would initially be aggregated at the submitting party's group level with the capability to drill to a sub-report based on either the legal entity and its activities by asset class, or by asset class across the party at the group level. From asset class, the reports would be capable of drilling to individual assets and then the individual underlying trades. These reports would likely be simple aggregations of directional positions, however the introduction of additional pricing feeds into the TR could also allow for a variety of pricing analyses, such as the identification of trades that are struck significantly in or out of the money.

It would also be anticipated that reports based on asset classes, or individual assets / asset pairings, would be required to identify the market participants (at the legal entity level) with the largest "position". Whilst of obvious use to the relevant regulator these may also be of use to market participants although the positioning of competitor activity would need to be desensitised. For example, a participant might see that its current position on a given equity, or to a particular currency pair ranked it 5th with a given value and percentage of the total and, depending on what level of disclosure market participants are prepared to allow, it might be able to see the values and percentages of other participants but without knowing the names of those participants.

Also, the inclusion of trades that have not been matched, confirmed or otherwise corroborated allows for some analyses of the operational performance of a given market and/or party. Apart from analysis of parties not submitting their activity to the TR through the inclusion of trades alleged against them by others parties that are submitting their activity, if trades are submitted from execution through to completion and the status of the trade is tracked as it changes, then timelines can be determined (for that market, the asset class, the individual asset, the counterparty group, the counterparty legal entity and/or the counterparty branch level). Counterparties, assets etc can then be compared against each other to understand the relative efficiency of these plus individual parties or even trades that diverge from the established timeline can then be analysed to understand cause and effect.

It is not uncommon at the outset of a reporting data warehouse initiative for there to be a lack of clarity on the reports and information required as invariably access to information inevitably then raises queries about that information that require further forms of analyses. This is another reason why the design of a reporting platform is so significantly different to that of an operational processing solution – the latter requires clear defined processing paths to ensure throughput (speed and volume) whilst the former needs a more open matrix set of paths to avoid preventing future development.

In all of the above possible forms of reporting, as stated under characteristics before, controls must be established to ensure supervisors are only able to obtain data directly relating to the entities that they themselves supervise. Market participants should have no greater access to data (apart from their own trades and positions) than the public at large. Positions at aggregated levels, and fully anonymised, may be disclosable to the public. No information should be made available to the public from which any inference could be drawn about any participant's market positions. Within the boundary of the "aggregated/anonymised" concept, it should be possible to provide data along various dimensions (for example, by currency, by jurisdiction, etc). It is important to preserve the public's trust that supervisors have appropriate information and that publication of data is not seen as an alternative - some aggregated information is already made available by BIS and ISDA.

Do you agree that trade repositories should provide adequate processes to ensure the reliability of the data provided? How can reliability be ensured?

As stated under characteristics above, it is important that a TR includes a complete data transformation mapping to ensure it is possible to trace data published in reports back to the source data received. For validation of data received, beyond the simple validation that the data feed received from a market participants or market infrastructure third parties is in itself reliable in terms of what that party transmitted, there are limited options to ensure the validity of the data provided to the TR:

- at a basic level, responsibility for the completeness and accuracy of trade data relating to an institution (i.e. all trades that that institution is a part to) should fall to that institution itself and reports can be produced from within the TR to a that institution seeking confirmation that the data held is complete and correct
- at a more intermediate level, comparisons with other reports such as BIS surveys or regulatory risk reporting may provide some validation
- a more involved basis would require regulators to validate the data submitted during the course of other regulatory activity/inspections.

Do you see any other entity with legitimate information needs with regard to OTC derivative trades recorded in a trade repository? If yes, please explain.

No.

2.3 Location

Do you see a need for establishing TR facilities in Europe if a global repository already exists elsewhere? Do you believe that a European repository is needed for each OTC asset class as described above (i.e. CDS, interest rate and equity derivative markets)? Please give reasons.

Given the global nature of markets, we do not see the need for repositories to be established in Europe if global TRs exist elsewhere, provided that European supervisors have adequate access to those TRs and appropriate influence over their operation.

If yes, what form should the trade repository facilities to be established in Europe take (e.g. single point of information, back-up facility) and which trades should be registered in such facilities (e.g. trades of European market participants, trades referring to European underlying entities)? Please specify.

Not applicable.

2.4 Legal Framework

Do you think there should be harmonised EU requirements for the regulation and supervision of trade repositories?

Yes, but these should also be harmonised with the US and other major jurisdictions where there is significant OTC derivative activity and which may develop repositories.

To what extent do you expect that protocols, common market practices and the like, surrounding proposed solutions for trade repositories, could promote harmonisation and foster safety and efficiency in the post-trading process? Please provide reasons for your position.

Given the other characteristics of a TR listed above, there is an opportunity that common practices introduced within a TR initiative, such as a common legal entity naming structure, could have a significant positive effect on the wider post-trade process. Any such initiatives that encourage greater standardisation of process and automation should be welcomed even though it is recognised that there are likely to be limits to how far standardisation can be achieved.