

# Solving the Big Problem with Big Data

Data from SEFs is streaming into swap data repositories, but can the CFTC effectively breakdown and analyse this data? What efforts are being made to harmonise trade reporting data and make it usable? Is there currently an appetite to develop an 'uber' repository that will span jurisdictions? Kirsten Hyde investigates.

"It has been often said that in order to experience change, you must first look inwards. I couldn't agree more. The first order of business for the CFTC is improving swaps data quality" – Former CFTC Commissioner, Scott O'Malia

Nearly two years has passed since swap data reporting began in the US. Yet, the country's derivatives market regulator, the Commodity Futures Trading Commission (CFTC), is unable to fully analyse the data streamed to swap data repositories (SDRs) to identify and measure risk exposures in the market – a major goal that was set in 2009 to avoid another blow up after the credit crisis.

Back then, G20 leaders agreed, following the collapse of Lehman Brothers, that all OTC derivatives contracts should be reported to trade repositories with data showing who was holding which contracts, such as interest rates or credit default swaps, in order to improve transparency, mitigate systemic risk and protect against market abuse. The aim was for regulators to have a snapshot of who is exposed to any failing lender and where risks are building up in the market.

In the US, under the Dodd-Frank Act, all swaps – whether cleared or uncleared – are to be reported to swap data

repositories (SDRs). The country's four SDRs are run by: CME Group; Intercontinental Exchange (ICE); Depository Trust and Clearing Corp (DTCC), which provides back office services for the Wall Street banks that own it; and Bloomberg, the newest SDR, which became operational this year.

However, each registered SDR has different systems architecture and reporting technology, which has resulted in entities reporting trades using different message types and in varying record formats. Each SDR receives more than 60 million messages per week, and with no uniform way of organising this data it has created challenges to the CFTC's efforts to review, analyse and automate the aggregation of data.

As an example, for the CFTC to conduct surveillance of swaps data reported to SDRs, it has had to navigate to each SDR's portal, inputting search parameters to request a report, and then manually scrolling through numerous columns of data fields for each swap.

In addition, the level of complexity and customisation inherent in swaps has further challenged the CFTC's capability to sort reported trades, despite there being fewer



trades reported in the swaps space than the futures markets.

## Challenges

According to research company Greenwich Associates, a lack of reporting standards makes using the data difficult, if not impossible.

“For example, butterfly trades – those with three legs up and down the interest rate curve – are priced and traded as a single transaction. The current clearing infrastructure requires each trade to be processed separately, however. This has left some SEFs reporting butterfly trades as three separate transactions, while others report them as one. Neither method is wrong by regulatory standards, but the lack of a standard approach to these and similar trade types creates unnecessary confusion,” says Kevin McPartland, head of research for market structure and technology at Greenwich.

“This issue is not just one for market participants, but for regulators. Data concerns are so acute, in fact, that the CFTC has teamed up with the Treasury’s Office of Financial Research (OFR) to figure out how something as seemingly

simple as trade reporting has gone so wrong,” he adds.

Speaking at the International Swaps and Derivatives Association’s (ISDA) annual London conference in September, Scott O’Malia, the new chief executive of derivatives industry body ISDA and a commissioner at the CFTC until July this year, confirmed that utilising swaps reported data has been a challenge for the Commission.

“The US has had swap data reporting for nearly two years, and while the compliance is high, the utilisation of the data is relatively low. It has been a challenge; this is not something new, this is something when I was at the Commission I was very concerned about.”

O’Malia continued: “It is a huge opportunity for regulators to utilise this data. Having complete transparency and a view in the market is key, and I think it will make market participants focus on their role knowing that regulators have this data. It really can improve regulatory oversight capacity, but the problem is you have to be very specific and accurate in terms of how you report and how you aggregate and that has been a challenge thus far.”

Differences in functionality can be traced in large part to the swap data reporting rules, which failed to provide for



Scott O'Malia,

**“The US has had swap data reporting for nearly two years, and while the compliance is high, the utilisation of the data is relatively low.”**

standardised data fields or reporting formats. It has been argued that these should have been the foundation of the CFTC's rules for the reporting of large amounts of market data and would have been if it hadn't rushed through its rule implementation and had given more thought to the practicalities of aggregating and cross-referencing specific data points.

According to one executive at an SDR, “The regulators pretty much wrote themselves a blank cheque of what SDRs have to do for them.”

Another SDR executive adds, “Dodd-Frank described data that needed to be reported. There was a lot of feedback saying ‘don't be overly prescriptive and allow the markets to sort out the ingestion of data and how it needs to be sent out’. As a result, there are different standards and proprietary codes that are being used by the data repositories.”

## Data Harmonisation Effort

The failure to use a common metric has increased the risk of misunderstanding critical market information, making the improvement of swaps data quality and the usefulness of the information one of the CFTC's top priorities.

A CFTC working group now meets regularly with the four SDRs to agree on standard data elements and content, harmonising data fields, developing common reporting guidelines and implementing a validation process to improve data quality and integrity. These efforts are focused on harmonising existing data rather than requiring new data from



Bruce Tupper

**“SDRs are a whole new market infrastructure and people realise that these things take time.”**

submitters. There is also a focus on SDRs providing direct electronic access for CFTC staff to search, filter, sort and export swaps data.

“SDRs are in an ongoing series of phases to harmonise the data across SDRs as reported to the CFTC,” says Jonathan Thursby, president, global repository services at CME Group.

“We work closely with the CFTC's Office of Data Technology and all the SDRs have been meeting, usually quarterly as a group and with calls in between, to work through building these standards which are commonly referred to as the data harmonisation effort. The second phase of this effort is well underway,” says Bruce Tupper, president, ICE Trade Vault and ICE eConfirm Services. “SDRs are a whole new market infrastructure and people realise that these things take time. We are now in the process of refining and standardising data reported to the Commission through the data harmonisation effort.”

Marisol Collazo, US chief executive officer for the DTCC Data Repository, adds, “The CFTC is really focusing on what kind of output they want to see and doing this through the data harmonisation effort with the four trade repositories.”

The CFTC last year released data harmonisation guidance on data elements standards. The first phase was for credit default swaps and included 30 data elements. The four SDRs are finalising phase 2 this year, with phases 3 and 4 to follow with approximately 30 additional data elements each.

According to the CFTC, the SDRs have prepared action plans that will implement multiple phases of delivery per asset class and field-by-field. Next, the Commission will address interest rate swaps with other commodities to follow in the future.

The first iteration, phase 1, seems to have cleaned up a lot of problems.

“The CFTC can now successfully aggregate CDS data amongst the SDRs and produce reports and we're refining this information for them. Our goal is to give the Commission the ability to efficiently aggregate data amongst the SDRs,” says Tupper.

Collazo agrees. “It's been very collaborative in terms of us working together as there has been collaboration amongst trade repositories to really understand the definition, the format and the output. That's definitely the right path forward in terms of data quality. What has not happened yet, and it's an effort that's underway at the Commission, is to attach that to report output – so, for example, focusing on a set of fields that will lead to a set of reports that will aid in the CFTC's supervision.”

She adds, “From the DTCC perspective, we are looking to globalise this effort. So we're saying, if these are the key fields that the CFTC have identified, is there an opportunity to focus on these issues for all the data that we're receiving across jurisdictions? It's important to drive forward improvement of data quality, regardless of the jurisdictions, because such global consistency is necessary to enable an effective data aggregation for systemic risk analysis which is central to regulatory oversight.”

## Cross border harmonisation

Data difference is also hampering international collaboration, although longer term, the CFTC says it will work with other regulators and industry associations to come to an agreement on universal reporting conventions. It says developing a consistent reporting convention will improve data quality and cross-market and international data harmonisation as envisioned by the G20.

However, market participants remain concerned at the lack of coordination in what data actually needs to be reported



under the various regulatory regimes and the lack of agreement on the format in which that data has to be submitted.

In addition, no framework yet exists that allows global regulators to monitor risk by looking into other countries' trade repositories. One country's data secrecy laws may prevent access to a trade repository by regulators from outside that country, even if trade data stored in that repository includes details of transactions carried out by a counterparty from another country.

At the ISDA conference, O'Malia said: "On cross-border harmonisation, the two [jurisdictions – Europe and the US] are not sharing data. There is not a bit of difference in terms of the objectives of the reporting that's in the trade repositories. There might be different fields, and the terms of conditions might be slightly different, but this is an area where both jurisdictions should have substituted compliance and be able to share that data; it's a frustration that they cannot."

The different legal frameworks present another challenge. As well as privacy laws, blocking statutes and other rules could prevent the reporting of counterparty information to a repository, and also limit the repository's ability to disclose transaction data to regulators, preventing any one regulator from gaining a clear overall picture of global derivatives activity.

## Centralised mechanism

Some market participants are hopeful, however, that supervisors recognise the challenge ahead. The Financial Stability Board, the G20's regulatory task force, last month released proposals to create a centralised mechanism to produce and share global aggregated OTC derivatives data to ensure that authorities are able to obtain a comprehensive global view of OTC derivatives market and activity.

"TR data are fragmented across many locations, stored in a variety of formats and subject to many different rules for authorities' access," the FSB said, referring to the 25 trade repositories based in 11 jurisdictions that are either operational or have announced that they will be.

Its feasibility study, which follows a consultative version of the report earlier this year, compares three basic options for aggregating OTC derivatives trade repository data; Option 1, a physically centralised model; Option 2, a logically centralised model; and Option 3, the collection and aggregation by authorities of raw data from trade repositories, the only one currently available for use.

The FSB believes that options 1 and 2 are highly preferable to Option 3, saying the latter has practical limitations that allow it to meet "only part of authorities' data needs, beyond protecting against market abuse".

Option 1, the physical model, would typically involve a central database or hub where all the data are collected from TRs, stored and subsequently aggregated within the central database for onward provision to authorities as needed.

Option 2, the logical model, would involve a physically decentralised data collection and storage system. Logical centralisation could take a number of forms, but the key feature would be a type of logical indexing mechanism that enables the use of technology to aggregate data from local TR databases rather than the use of a physically central facility. In this option, the underlying transaction data would remain in local TR databases and aggregated with the help of the central index.

One variant of logical centralisation is a model where the data is collected and stored locally but, instead of authorities using the logical indexing mechanism themselves to obtain the

**"It's important to drive forward improvement of data quality, regardless of the jurisdictions, because such global consistency is necessary to enable an effective data aggregation for systemic risk analysis which is central to regulatory oversight."**



Marisol Collazo

data from local databases, a designated agent would maintain the central index and the platform for responding to requests from authorities.

In either case, the FSB points out that a number of key steps need to be undertaken as part of the preparatory work before any formal project is launched to implement a global aggregation mechanism.

It notes that, amongst these steps, it is critical for any aggregation option that the work on standardisation and harmonisation of important data elements be completed, including the global introduction of a legal entity identifier (LEI), and the creation of a unique transaction identifier (UTI) and unique product identifier (UPI).

It also indicates in very broad terms, the types of legal and regulatory changes that would be needed to allow a central mechanism to access the necessary data from TRs and to aggregate the data for authorities.

Acting on its recommendations, the FSB has asked the Committee on Payments and Market Infrastructures (CPMI), which falls under the Bank for International Settlements (BIS), and the International Organisation of Securities Commissions (IOSCO) to develop global guidance on the harmonisation of data elements that are reported to trade repositories. A working group is due to report on its developments to the G20 at a meeting in Brisbane on November 15-16.

## Response

While the FSB says its proposals take into account public feedback it received from February, a number of senior market participants say that rules and processes within each jurisdiction, such as in the US, are likely to take precedence over the FSB proposal for an 'uber repository'.

At the ISDA conference, O'Malia called the idea "a nice vision for the future, but the reality is dealing with what we have today and making that work".

"Where regulators need to focus right now is working together to harmonise the convention of reporting, making sure we are doing an apples-with-apples comparison with the data," he added. "The sharing of data can easily be resolved, but it's going to take some time."

CME's Thursby agrees: "As a global industry we should first be focused on getting regional reporting to a quality state before focusing on the very important global picture. From a practical matter, a global TR is not feasible in a world without a global regulator. The most issue-free and readily available solution is to see traction around data sharing agreements within the regulatory community."