



ASIC
Australian Securities &
Investments Commission

Committee	Parliamentary Joint Committee on Corporations and Financial Services
Inquiry	Oversight of ASIC, the Takeovers Panel and the Corporations Legislation No.1 of the 46th Parliament
Question No.	023
Date	26 November 2021
Topic	ASIC's data lake and social media platforms
Reference	Spoken, 26 November 2021, Hansard page 48
Committee member	Senator O'Neill

Question

Mr Day: I understand; I was helping, Senator! We are very pleased with the data lake we've got. We think we're one of the first agencies to, effectively, have a secure, cloud-based data lake the way that we do. It actually provides with us a benefit and a standard format so that we can 'tip in' to that data lake a range of datasets from across all of our regulated areas, and we can then utilise them with our data analytics tools in systems and processes across our agency. In relation to the social media platform and that work, I'm not sure, and I'd have to take on notice whether or not that data forms part of that. But this is a platform that we have built with our chief data analytics office, or through that group, to effectively get us ready for the future, in terms of where the data source is coming from. And, obviously, there are new datasets coming to us because of internal dispute resolution data changes that are coming, and better data from the new corporate registry, in due course. So this is a platform for all of our data assets, going forward.

Answer

A data lake is a central data repository for storing data in its natural format, at any scale and maintained by a data governance framework including standards, policies, procedures, practices and ongoing monitoring capability. Data in its natural format can take the form of structured (tables, comma delimited, relational data); semi-structured (logs files, clickstream, social media) and unstructured (text, video, photos, audio).

In the data lake, ASIC can safely store, manage, access and analyse every type of data, document or file created or received. The data lake is an essential element to ASIC's enterprise data platform with a key feature is its ability to scale as ASIC's data needs grow.

In relation to ASIC's monitoring of social media, ASIC has a range of tools, technologies, and tactics at its disposal to view the commentary, content, and contributors across various social media platforms. Some of these tools allow us to view and understand where the conversation is being directed off-platform to other encrypted chat forums such as Telegram or Discord, for ASIC to take further action on these platforms directly.

ASIC is targeting the use of the Data Lake to leverage its power to process the large volumes of social media commentary and generate signals for ASIC to further identify conversations that may be driving share prices or sharing anomalous information related to Australia's financial markets.

In ASIC Data Strategy, we have plans to bring social media data into the data lake allowing us to achieve the above benefits. After the social media is onboarded in Data Lake it would be easy for ASIC to combine this data with other datasets within ASIC's data ecosystem.

Recently, the Markets Group introduced a new suite of analysis tools, dashboards, and capabilities, to enhance ASIC's abilities to design, develop and customise our surveillance analytics to target evolving and increasingly sophisticated financial market misconduct, such as collusive behaviour, cross market order book manipulation and insider trading rings.