

The background of the entire slide is a dark blue aerial photograph of a rural landscape, showing a patchwork of agricultural fields and a winding river. The image is semi-transparent, allowing the text to be clearly visible.

WIRED

The data challenge

Managing ever-growing commodities
data from diverse digitalised sources

REFINITIV[®]



**The town of Cushing,
Oklahoma, has a population
of around 7,000 people.
But what it lacks in numbers
of citizens, it more than
makes up for in oil supplies
– it's known as the pipeline
crossroads of the world.**

Introduction

Building insights on granular data and real-time measurements

Around two dozen pipelines converge at the town's oil hub, the largest in the United States, with around 90 million gallons of storage available at the site. As a result, the hub acts as a key node in the supply and trading of crude oil around the world. Cushing's oil operations are closely scrutinised by analysts and those making trades, and what happens there is influential and incredibly far-reaching.

The amount of oil being stored in such hubs in the United States used to only be published once or twice a week by the Energy Information Administration, says Simon Wilson, the head of oil markets at Refinitiv, making real-time analysis and decision-making all but impossible. But now, thanks to sensors and infrared cameras that allow storage levels to be monitored and the flow of oil to be tracked remotely, such data is in abundance. Sensors make it possible to determine when tanks are being used, and if they're in maintenance, construction or testing. Analysts no longer need to wait for the official volume figures to be published – instead, they can cross-check their own readings with the official data as soon as it's published.

Decisions on what trades to make have moved from being based on news reports ten to fifteen years ago, to being focused on real-time measurements captured about the movement, storage, supply and demand of commodities. Collecting these sources of data increases the transparency around what is happening within the industry and allows for a more informed picture of the world to be created. "Previously, our customers may have been happy with just one source of oil flows data and cargo tracking," Wilson says. "Now, they want to take in multiple sources. Refinitiv has to integrate two or three sources, so that they can get a blended view."

The growth in data is common across all commodities – from energy to agriculture and raw materials. But simply collecting and amassing data is not enough to get an advantage from it. Companies need to be able to integrate it into their workflows, analyse it in new ways and take action on what it may reveal. Companies that aren't

able to utilise data in this manner risk falling behind their competitors. And the amount of data available for analysis is only getting more granular. "For crude oil, you'll probably have at least 800 markers based on the different types available," Wilson says. "When it comes to gathering that data, it's those nuances around oil grade which are important." This granularity is key, Wilson explains, as many refineries are set up to handle certain grades of oil. Knowing how much of a particular grade is travelling toward a specific location can indicate the supply and demand levels for that grade.

This e-book explores how financial institutions and commodity trading organisations should be thinking about integrating data into their workflows, the first steps that need to be taken, and the complex process of standardising large datasets. It is the second in a series of four e-books produced by WIRED and Refinitiv that examine the future of commodities trading, how companies will move towards more automation, the steps they need to take to understand the new, data-heavy world – and how to get there.

Oil-storage facilities can now provide current, granular data on supply and demand



The democratised data boom

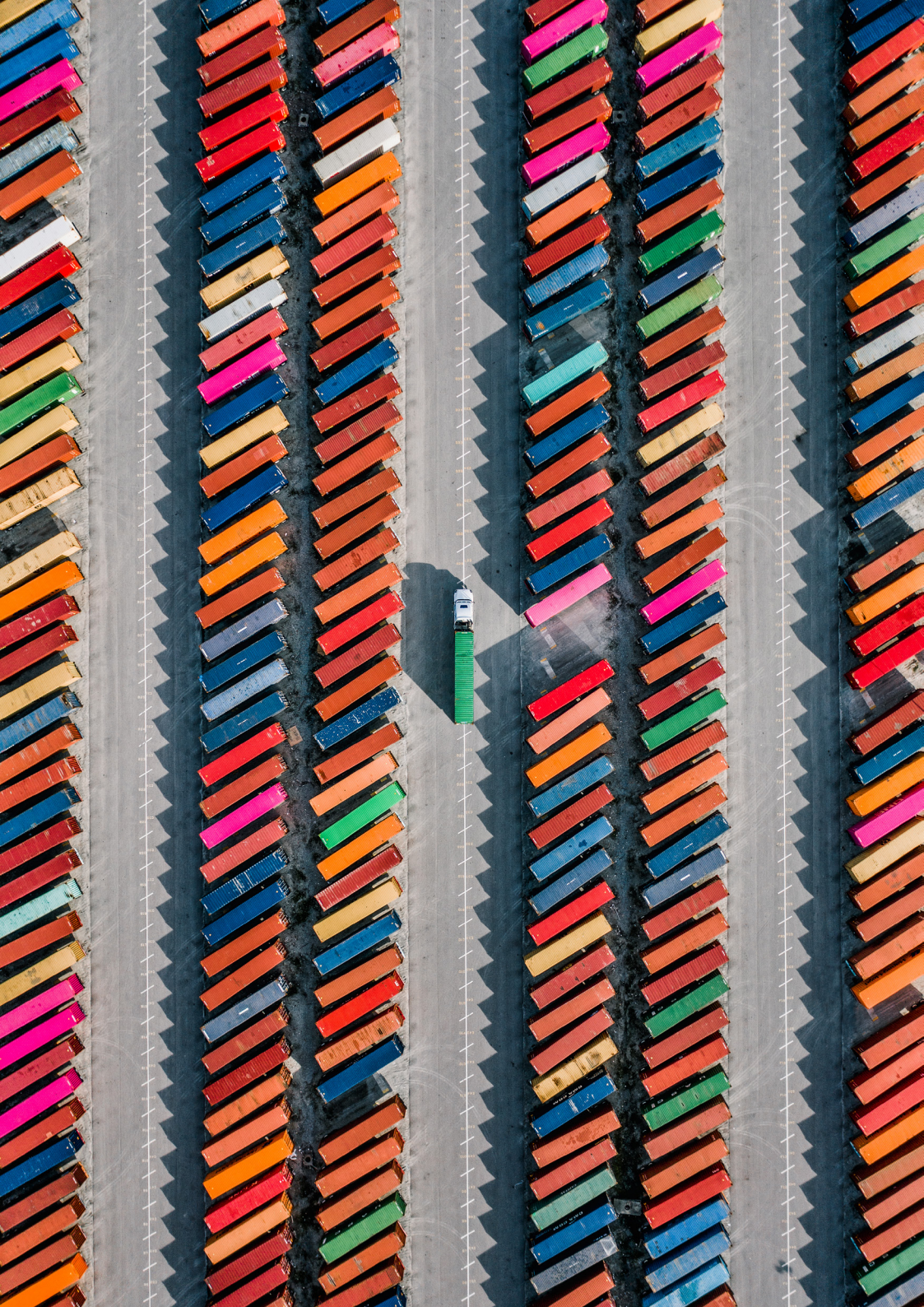
Digitalisation has spread through every sector, creating opportunities not only in analysis and trading, but in transparency, supply chains and more

"The mass-scale democratisation of data across a number of sectors has opened up new possibilities for traders and analysts," says Alessandro Sanos, global director, sales strategy and execution for commodities at Refinitiv. However, not all industries have embraced the adoption of remote sensors and digitalisation as quickly as others. "Commodity trading companies move metal – but they also move a lot of paper between one another," says Chris Evans, the director of metals at Refinitiv. Some masters of ships, he says, can still insist on having the correct paper records present, with the right authorisation stamps on them.

"Before Covid-19, I think that there was probably some reticence to engage in the digitisation of commodity supply chains," Evans explains. But the global pandemic has accelerated the process. Analysts at KPMG have said the stresses put on supply chains around the world by the pandemic saw a renewed drive to digitise parts of the process; by some counts, decades of digitisation took place in months. As well as greater transparency around supply chains, this also benefits companies' Environmental, Social and Governance (ESG) targets and risk management. It's now possible to create an "immutable data trail," according to KPMG analysis.

This digitisation, naturally, means more data is available. In recent years, commodities trading houses and analysts have been collecting as much data as they can in a bid to better understand their industries and gain insights from previously unavailable information. Once they have it, they can also combine it with the proprietary information they have about their own trades or commodity supply levels. "The very bespoke nature of metal trading through the London Metal Exchange means that there are potentially some big data gaps," Evans says. The exchange itself sees metal producers and industrial manufacturers trade alongside merchants, banks, brokers and





proprietary traders. Trades are made on everything, from zinc and tin to the copper, gold and silver that have become essential in our smartphones and other consumer technology we use daily. "With new techniques, we're now more able to fill the gaps than we ever have been," says Evans. "We're able to take certain trading data or data points in trading and interpolate between them to create a better picture of the market than we've ever been able to do before."

However, hunting for data sources can be a lengthy process. To really understand the markets, a global view is needed. This is particularly relevant for the metals industry, which has seen China emerge as a major player. China's manufacturing might make it one of the world's largest metals consumers, and as a result, a key trading market – predicting its market supply and demand accurately can provide vital insights on the overall industry.

Through its Eikon platform, Refinitiv has collated thousands of datasets and made them compatible. For commodities, Refinitiv has gathered movements on agricultural goods, databases on oil-flows, and specialist data from third parties such as S&P Global Platts and Argus Media. This can be combined with weather, shipping and news data from Reuters, plus company data and economic analysis of up to 200 countries. All of this gives companies a holistic idea of the commodities market – both regionally and internationally – so that better-informed decisions can be made.

The need for commodities traders and analysts to understand their data landscapes is growing quickly – emerging technologies are changing the way businesses operate, and for companies to take advantage of it, they need to have their data in order. Since the early 2000s, the capabilities of artificial intelligence (AI) and machine learning (ML) have increased rapidly due to the emergence of more data, better algorithms and advanced processing power.

The majority of AI and ML applications being used today are based on structured data. Algorithms are fed large datasets and trained to spot patterns in them, before making predictions about future outcomes. Throughout the 2020s, more businesses are likely to use AI within their day-to-day operations, and the technology is becoming increasingly commercialised with a number of practical use-cases. For

the world of commodities, if done correctly, AI offers the potential to predict future prices of goods and where there may be surges in demand for them. A company that can, essentially, predict the future in a chosen sector, will have significant advantages over its rivals.

But for this setup to work, an organisation needs to be able to access large volumes of data and must ensure that this data has been organised so as to be effectively processed with AI. Refinitiv says it has more than 5,000 content analysts focusing on cleaning up vast amounts of data: each year, it processes 142 million company financial data points, one million research documents each quarter, and has more than 1,200 content partners. The ability to understand data is even more crucial, as the amount of data – plus the level of detail available – is set to increase in the coming years.

Mining companies need to be much more open with their data, says Ella Cullen, the co-founder of blockchain and supply chain transparency firm Minespider. “The mining and materials industry is very intransparent, meaning the majority of companies struggle to know who they are sourcing from beyond a few tiers (levels of suppliers or clients),” she says. “Downstream brands and manufacturers face legal and brand risks if it’s uncovered that they’re sourcing from someone or somewhere with a negligible human rights or environmental agenda”.

The industry is waking up to the benefits of collecting more data about the goods and items it moves. Access to data such as where materials come from, where they are going and who has processed them means greater accountability, as all commodities can be traced through the entire chain. “There is a lot of development around physical trackers at the moment, everything from QR codes and RFID tags, to drones and satellite imagery, to materials like Stardust, which are chemical elements that can be added to minerals at the point of smelting and can be scanned at a later point to guarantee authenticity,” Cullen says. All these data points can be entered into systems and provide a clearer view of the origins, supply chain and delivery of commodities. Being able to understand, manage and interpret this granularity of data will give those in the commodities industry an advantage over their competitors.

The standardisation challenge

Diverse datasets can bring unparalleled insights, but must first be made fit for purpose – and Refinitiv has dedicated teams to achieve this

Making datasets compatible with other datasets is a crucial step, but it’s by no means a straightforward task – in fact, it’s something that only gets more complex as ever more data that is to be standardised flows in and is added to the sum. To work together, the fields in each dataset need to match, to ensure there aren’t any discrepancies. To get to this stage, a number of processes, including human verification and quality assurance checks, need to be put in place.

Refinitiv has developed an exhaustive setup that’s responsible for handling raw data and making sure it is up to scratch. Raw data can come in many forms: it can be provided in electronic files such as Excel, through APIs directly from the source, or even in locked PDF files. Whatever form it comes in, all of this data needs to be made compatible with that from other sources. To achieve this, Refinitiv has teams in Bangalore, Manila, India, Poland and beyond.

Analysts at Refinitiv can then build upon this process to develop insights from various data sources. To predict what could happen next in commodities markets, its analyst teams run more than 200 models – that’s more than 100,000 times across more than a billion data points. The expertise built up through this process – including data creation, utility and management – allows Refinitiv staff to help their external clients with similar data challenges.

Making metadata work

Tagging the data within the data adds a level of granularity that lets analysts know the who, where, what and when of larger datasets

Collecting vast quantities of data is only the first step. To really make use of this information, it needs to be standardised and tagged with metadata – that is, data about data. These processes can ensure data is ready to be analysed, visualised and, ultimately, mined for intelligence. Once this has been done, the links, relationships, and connections between data points can be spotted easily.

For instance, data can only work with data that it matches. “If we talk to different partners or even customers, they may call shipping ports something slightly different,” says Refinitiv’s Wilson. “You’ve got to make sure that you’re comparing apples to apples.” Refinitiv started by making sure vessel tracking data from different sources matched those from others, and then moved on to the names and locations of oil refineries. Wilson says the process involved “standardising each part, so making sure that we’re talking about the same refinery, and the same units in China, with the associated ports, with the crude coming into that particular port”.

This standardisation process is crucial when it comes to making use of data, but it’s not the only option available. The right metadata can provide a lot of extra information and help to create new insights. Metadata exists to provide information about larger datasets – it can be the who, when, where, or what that is contained within a dataset. It’s everything except the full content of the dataset.

Take Wilson’s example of port names. The name of a shipping port in China could be considered as a piece of metadata and labelled as such. This label can then be used to see all mentions of the port across a series of datasets. An analyst could see the port was mentioned in a report about the volume of metal being moved and also see ships that have been, or are planning on arriving at, the same port. Looking at the tag allows data from various reports, such

Once standardised, diverse data, from satellite feeds to text documents can be combined





as news sources, to be gathered in one place. By using metadata efficiently, companies can gather all the information about one place, or event, in the same location and learn from this.

Perhaps more importantly, this sort of tagging can be crucial for AI, ML and emerging data science techniques. These systems feed off organised data, and metadata can help the systems learn about the specific elements that are tagged. Refinitiv uses text analytics, natural language processing and data-mining technology to create its Intelligent Tagging system. This can generate tags for vast amounts of data and documents, which may be proprietary to a company, automatically – the process is faster than having human teams sift through all the information and manually create tags.

A permanent identifier is given to people, places, facts and events that are mentioned in data, and each element is given a score to mark how important it is to the document. This can allow for easier searches to be made, personalised content to be sent to specific people, and the creation of valuable new insights.

Making sure data is standardised and has metadata associated to it sets companies up for the future, Wilson says. “The value is in the standardisation and the aggregation of the metadata and the data layer,” he explains. “If you have that framework in place, then any technologies or any movements over time, you are able to flexibly evolve with.” Companies that ensure their data processing is in place will be better prepared for a world that’s dominated by computer science and artificial intelligence.

Industry is waking up to the benefits of collecting more data about the goods it moves. These data points can provide a clearer view of the origins, supply chain and delivery of commodities when they are all linked together.

About Refinitiv

Refinitiv, an LSEG (London Stock Exchange Group) business, is one of the world's largest providers of financial markets data and infrastructure. With \$6.25 billion in revenue, over 40,000 customers and 400,000 end users across 190 countries, Refinitiv is powering participants across the global financial marketplace. We provide information, insights, and technology that enable customers to execute critical investing, trading and risk decisions with confidence. By combining a unique open platform with best-in-class data and expertise, we connect people to choice and opportunity – driving performance, innovation and growth for our customers and partners.

WIRED

REFINITIV[®]
