OECD GOING DIGITAL TOOLKIT

POLICY NOTE

The promises and pitfalls of SupTech for corporate governance-related enforcement

OECD
BETTER POLICIES FOR BETTER LIVES

going digital

This Toolkit note was written by Emeline Denis, under the supervision of Daniel Blume. It was reviewed by the Corporate Governance Committee (CGC) and approved for publication by the CGC on 24 March 2021. The note was prepared for publication by the OECD Secretariat.

This Toolkit note is a contribution to the OECD Going Digital project, which aims to provide policy makers with the tools they need to help their economies and societies thrive in an increasingly digital and data-driven world.

For more information, visit *www.oecd.org/going-digital*.

#GoingDigital

Please cite this publication as:

Denis, E. (2021), "The promises and pitfalls of SupTech for corporate governance-related enforcement", *Going Digital Toolkit Note*, No. 10, *https://goingdigital.oecd.org/data/notes/No10_ToolkitNote_SupTechCorpGov.pdf*.

Note to Delegations:

This document is also available on O.N.E. under the reference code:

DAF/CA/CG(2021)1.

Table of Contents

Figures

The promises and pitfalls of SupTech for corporate governance-related enforcement

Digital technologies and data hold the potential to automate and improve the efficiency and effectiveness of regulatory and supervisory processes, which have become increasingly complex given the substantial increase of complex regulatory data in recent years. Securities and financial regulators have turned to supervisory technology (SupTech) tools and solutions as a means to improve their oversight, surveillance and analytical capabilities, which can in turn have important benefits for financial stability and market integrity. This Going Digital Toolkit note takes stock of the most common uses of SupTech by securities regulators to date; identifies its associated benefits, risks and challenges; and outlines considerations for devising adequate SupTech strategies, with a particular focus on corporate governance-related enforcement.

Digital technologies and data are transforming the ways in which people, firms, and governments live, interact, work and produce at an accelerating rate (OECD, 2019[1]). This Going Digital Toolkit note considers the implications of this transformation for the supervisory practices of securities market regulators, and more specifically with respect to the enforcement of corporate governance-related requirements, which can be rendered more efficient through the use of supervisory technology (SupTech).

SupTech usually refers to the use of digital tools and solutions – including hardware and software – by public sector regulators and supervisors to carry out their responsibilities (FSB, 2017[2]; BCBS, 2018[3]). Nevertheless, some variations exist[1] as to what falls under the umbrella of SupTech (World Bank, 2018[4]; di Castri et al., 2019[5]). Recognising the potential of digital technologies and data to automate and thus improve the efficiency and effectiveness of regulatory and supervisory processes, SupTech is applicable to a wide range of regulatory authorities – including securities and financial regulators – whose core functions entail protecting investors, ensuring that markets are fair, efficient and transparent, and reducing systemic risk. By improving oversight, surveillance and analytical capabilities of authorities, SupTech can have important benefits for financial stability and market integrity (FSB, 2020[6]).

Overall, the use of SupTech applications can help jurisdictions in their implementation of several key recommendations in the G20/OECD Principles of Corporate Governance ("the Principles"), which serve as the globally recognised basis for assessing countries' legal and institutional frameworks to support effective corporate governance policies and practices. Of particular relevance, Principle I.E states that "…supervisory, regulatory and enforcement authorities should have the authority, integrity and resources to fulfil their duties in a professional and objective manner. Moreover, their rulings should be timely, transparent and fully explained" (OECD, 2015[7]). Beyond enhancing the overall capacity and efficiency of supervisory oversight at a more general level, SupTech applications may be relevant to the implementation of more specific recommendations from the Principles, such as Principle III.E, which calls for the prohibition of insider trading and market manipulation and its enforcement, as well as the enforcement of other regulatory provisions, for example with respect to disclosure.

1 SupTech is defined by Dias and Staschen (2017[12]) as "technological solutions focused on improving the processes and effectiveness of financial supervision and regulation", and by the World Bank (2018[4]) as "the use of technology to facilitate and enhance supervisory processes from the perspective of supervisory authorities". Castri et al. (2019[5]) define SupTech as "the use of innovative technology by financial authorities to support their work", restricting "innovative technology" to big data and artificial intelligence (AI) tools, and "financial authorities" to supervisory and non-supervisory authorities but excluding authorities in charge of monetary and macroeconomic policies.

By extension, SupTech solutions also have the potential to alleviate the regulatory burden on regulated entities, which have themselves turned to regulatory technology (RegTech) tools to improve compliance outcomes against regulatory requirements and enhance risk management capabilities. Such solutions hold the potential to reduce costs related to regulatory reporting, data collection and risk management (ESMA, 2019[8]).

According to IBM estimates (2018[9]), poor data quality costs the United States (US) economy around USD 3.1 trillion a year, and one in three US business leaders do not trust the information they use to make decisions. Research also suggests that while financial authorities have access to a growing wealth of data to guide their decisions and actions, they tend to lack the infrastructure or skills to make use of this data, with increasing amounts of data often simply translating into more manual data processing and leading to "analysis paralysis" down the line (R²A, 2019[10]). As data continues to increase in volume, velocity, variety and complexity, it is essential that both regulators and market participants develop systems to appropriately process, monitor and analyse datasets of regulatory relevance.

This Going Digital Toolkit note takes stock of the most common uses of SupTech by securities regulators to date, identifies the main benefits, risks and challenges associated with SupTech adoption, and outlines considerations for devising adequate SupTech strategies. Drawing upon insights from surveys and studies undertaken by international bodies[2] and other research, which appear to reflect an emerging consensus around some of the benefits and challenges related to the application of SupTech for the supervision of financial institutions, this Toolkit note reviews evidence to better understand the promises and pitfalls of SupTech use for securities market oversight, with a particular focus on corporate governance-related enforcement.

Drivers and typology of SupTech developments

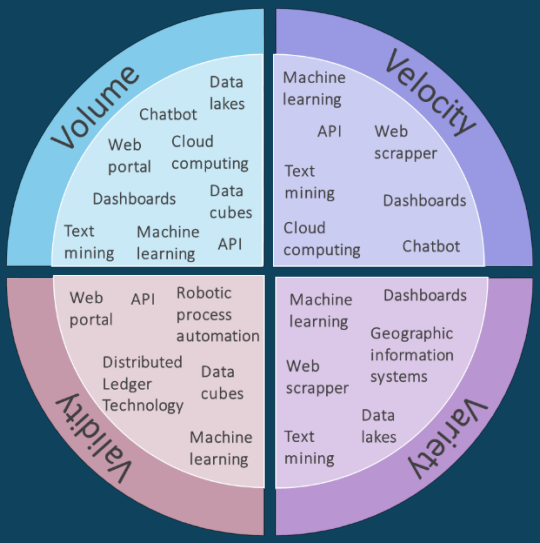
Demand and supply drivers have simultaneously spurred the development and application of SupTech tools and methods (ESMA, 2019[8]; FSB, 2020[6]). From the demand side, the increasing complexity and volume of regulations in the aftermath of the 2008 global financial crisis has in turn led to a substantial increase in regulatory data, hence pushing authorities to turn to digital tools to enhance their supervisory capability and efficiency. This trend has been met with supply drivers, including the availability of new analytical methods and tools at lower costs, which allow large datasets to be collected, stored and analysed more efficiently.

2 Including the Financial Stability Board, World Bank, International Organization of Securities Commissions, European Securities and Markets Authority, FinCoNet, etc.

In particular, as the volume and frequency of both structured and unstructured data increases substantially, so does the need for architectures or systems that are able to collect, store, analyse and visualise these new forms of data.[3] The greater availability of "big data" itself stems from the increasing volume, frequency and granularity of reporting requirements. Characterised by the "4 Vs" (volume, variety, velocity, validity), big data can pose data governance challenges for authorities, which have turned to technologies enabling sophisticated data processing techniques and generating advanced analytics (Figure 1).

3 In addition to regulatory returns from regulated entities, authorities leverage open source information (e.g. social media posts) to enhance their insights. According to a recent survey undertaken by the Financial Stability Board (FSB) (2020[6]), while regulatory, statistical, and market structured data make up the majority of data types collected from reporting institutions (45%, 22% and 12%, respectively), unstructured data amount to around one-fifth of the data collected by authorities (14% of unstructured regulatory data, 4% of unstructured statistical data, and 3% of unstructured market data). While unstructured data may offer useful insights, it is often collected in a format that makes it difficult to process and analyse.

Figure 1. SupTech tools to support data processing and analysis



Notes: The Author has defined the technologies and tools in this figure as follows. Application programming interfaces (APIs) enable real-time reporting and automated validation through direct database-to-database data transmission. Chatbots enable data collection in real time through automated capture and interpretation of qualitative data. Cloud computing enables on-demand network access to shared computing resources. Dashboards refer to interactive reporting tools enabling meaningful data visualisation. Data cubes provide for granular data storage and transmission. Data lakes enable storage of diverse structured, semi-structured and unstructured data. Distributed ledger technology (DLT) enables securely proposing, validating and recording changes to a synchronised ledger distributed across multiple nodes. Geographic information systems (GIS) enable automated analysis of spatial/geographic data. Machine learning refers to automated data analysis. Robotic process automation (RPA) automates manual, rule-based and repetitive human activities by 'bots'. Text mining automates extraction of meaning from textual data. Web portal refers to static file upload via website with built-in automated validation checks. Web scraper automates web data capture by 'bots'. While these definitions and the allocation of the technologies and tools shown in the figure to one of the 4 Vs are based on an analysis of the application of SupTech tools within securities and financial markets and are aligned with FSB's work, it should be noted that they may not align with other OECD definitions or work on digital transformation.
Source: Author, adapted from (di Castri et al., 2019[5]).

Despite the wide range of supervisory technologies available, their distinct features make their respective applications most relevant in specific areas of the data lifecycle. For instance, machine learning (ML) and natural language processing (NLP) are mostly applied by authorities for data analysis, processing and validation, while cloud computing is most often used for data storage, and blockchain is considered to offer potential for data collection (FSB, 2020[6]).

SupTech applications evolve along with technological innovations. To date, SupTech initiatives can be considered as belonging to four successive

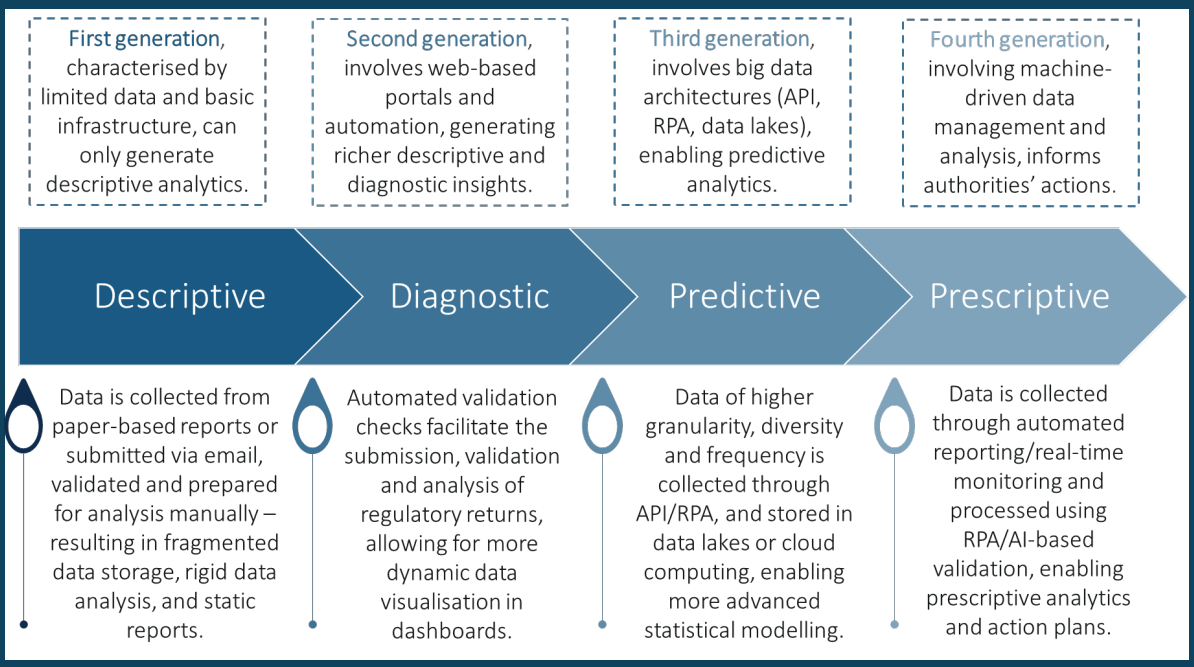
technological layers or "generations", which respectively generate descriptive, diagnostic, predictive and prescriptive analytics (Figure 2) (di Castri et al., 2019[5]). While the first generation covers primarily manual data management workflows, the second involves the digitisation[4] of certain paper-based processes in the data pipeline. These early generations of data architecture support mostly *descriptive* and *diagnostic* analytics (i.e. describing what happened and diagnosing why it happened). In a continuum, the third generation covers big data architecture, and the fourth involves artificial intelligence (AI) as its main attribute – both enabling *predictive* and *prescriptive* analytics (i.e. predicting what will happen and prescribing anticipatory action).

With supervisory authorities' use of predictive and prescriptive analytics emerging only recently, they are still at the experimental or development stages, but are gaining momentum. By fully automating data processing and optimising data storage and computation, big data architectures[5] can process larger datasets with greater computing power, in turn generating advanced insights such as predictive analytics. As AI-enabled solutions require large volumes of data and significant computing power in order to generate valid and actionable results, they are usually built upon pre-existing big data architectures. This fourth generation is characterised by machine-driven data management and analysis – which may involve natural language processing and machine learning to collect unstructured and disparate data, as well as recommendation engines suggesting courses of action. Chatbots may also be leveraged to perform tasks such as responding to and resolving complaints (di Castri et al., 2019[5]).

4 Digitisation refers to the conversion of analogue data and processes into a machine-readable format (OECD, 2019[1]).

5 Big data architectures require two key design features: 1) internal coherence of each of its layers so they can all process the speed, size and complexity of big data, and 2) built-in quality assurance and security procedures to ensure the validity and integrity of the data from the point of collection to the point of consumption by end users, thus enabling seamless end-to-end data flow without lags of size constraints (di Castri et al., 2019[5]).

Figure 2. The four generations of SupTech



Source: Author, adapted from (di Castri et al., 2019[5]).

According to a recent survey from the Financial Stability Board (FSB) undertaken among FSB members[6] (2020[6]), the first and second generations of SupTech initiatives encompass the majority of technologies used by supervisory authorities, with 49% of surveyed authorities using data analysis functions for descriptive outputs and 32% for diagnostic outputs. Only a minority of respondents report using technologies comprised in the third predictive category (11%) and the fourth prescriptive category (8%). Echoing these findings, a recent report from FinCoNet (2020[11]) based on survey responses from 21 market conduct and financial consumer protection authorities similarly demonstrates that while some SupTech tools currently deployed in this arena are used to make predictions, the majority are designed to collect or analyse data or automate workflows.

While third-generation data collection solutions and fourth-generation data analytics potentially yield the most value for authorities by enabling forward-looking supervision and greater storage and mobility capacity, technologies comprised within earlier SupTech generations can still generate sufficient

[6] The FSB includes 24 member jurisdictions, as well as 13 international organisations and standard-setting bodies. Member jurisdictions include: Argentina, Australia, Brazil, Canada, France, Germany, Hong Kong (China), India, Indonesia, Italy, Japan, Korea, Mexico, Netherlands, the People's Republic of China, the Russian Federation, Saudi Arabia, Singapore, South Africa, Spain, Switzerland, Turkey, the United Kingdom and the United States.

information and substantial efficiency gains to be beneficial as well, depending on the task at hand (di Castri et al., 2019[5]; Dias and Staschen, 2017[12]).

The benefits of SupTech

Because regulators and supervisors all rely on data, internal procedures and working tools, as well as human and other resources, they all face common challenges – albeit to varying degrees – related to low data quality and time-consuming manual procedures (Dias and Staschen, 2017[12]). SupTech applications can help authorities address these challenges by enhancing their capability, efficiency and effectiveness in terms of data collection and analysis, in particular by enabling the automation of routine tasks, the development of new analytical techniques, and the provision of better insights. By using tools to analyse increasing volumes of both structured and unstructured data of regulatory relevance, authorities can shift their focus away from labour-intensive tasks to activities requiring human judgment and expertise, allowing them to better allocate human resources and reduce costs over time.

Overall, SupTech tools are most commonly applied in the area of regulatory reporting, followed by data management, market surveillance, and misconduct analysis, respectively. While use cases have slightly decreased in the three former areas in recent years, misconduct analysis has gained traction for SupTech applications with the largest increase in the number of reported use cases by FSB members since 2016 (FSB, 2020[6]). SupTech applications can be developed in-house, by external vendors, or a combination of both.

Improving misconduct analysis

According to a recent FSB survey (2020[6]), SupTech applications have gained the most momentum in recent years for misconduct analysis, with the largest increase in the number of reported use cases by authorities since 2016. This may be because new tools are required to tackle new digitally-enabled forms of money laundering, terrorist financing, mis-selling and fraud. Another explanation may be that SupTech applications are particularly relevant to conduct supervision, which relies on the analysis of large amounts of granular, time-sensitive and unstructured data from disparate sources, it can particularly benefit from the development of big data architecture and AI tools.

Evidence suggests that authorities use advanced analytics such as machine learning, natural language processing, text mining and network analysis to enhance their capacities – especially with regards to detecting networks of related transactions, identifying anomalies and unusual behaviours, and drawing insights from extensive amounts of structured and unstructured data (Coelho, De Simoni and Prenio, 2019[13]). For instance, Mexico's National Banking and Securities Commission (CNBV) has developed a prototype for a Natural Language Processing (NLP) application to detect what a suspicious

Anti-Money Laundering/Combatting the Financing of Terrorism (AML/CFT) network is 'talking about', thus facilitating the detection of unusual transactions, relationships, and networks events to identify potential money laundering issues that cannot be identified by people.

Likewise, the Monetary Authority of Singapore (MAS) has deployed an augmented intelligence system called "Apollo" that automates the computation of key metrics used in the analysis of suspicious trading activities, and assesses the likelihood that certain types of market manipulation have occurred. As a "robo-expert", it seeks to predict the likelihood of positive prosecution outcomes for new cases by understanding how experts analyse market misconduct cases. MAS built and trained Apollo using expert reports and the trading data from cases that they had successfully prosecuted in the past.

Several benefits have resulted from its implementation. Automated trade analysis reduces the need for manual computation, helps to identify fraudulent transactions with higher market impact and provides greater insight into market trading behaviours. In addition, it allows for the testing of various case scenarios to fine-tune investigation strategies for individual cases, thus also helping with case prioritisation and guiding decisions on the appropriate courses of enforcement actions.

The Financial Superintendency of Colombia (SFC) has also recently initiated the Observer Banca Project, which uses text mining tools to analyse unstructured data – including web scraping news articles, and sentiment analysis from social media posts. In particular, it extracts news, information shared by specialised analysts (i.e. journalists and experts) and social media information, and consolidates this data in a standardised framework before implementing machine learning models to analyse market sentiment on the macroeconomic environment and across different economic sectors.

By facilitating non-structured information analysis, the project generates information repositories to support current and historical analysis, thus generating enhanced insights to better identify misconduct behaviours that might require further investigation. In addition, through a new technological development called Smart Supervision, the SFC will be able to receive real-time information on complaint handling processes carried out by financial institutions, which will serve as primary input for data analytics for market conduct supervision purposes. This new application is planned to be finalised by the end of 2021.

Improving market surveillance

Authorities can leverage big data architectures to perform real-time market monitoring. Evidence suggests that securities regulators have leveraged these technologies to transform large datasets into usable patterns for market

surveillance purposes, including to detect potential insider trading and market manipulation. However, designing and implementing tools focused on certain aspects of market surveillance can be complex due to the large volume and variety of data required (i.e. regulatory and market data and intelligence). As new technologies become available, they may facilitate their development and deployment (FSB, 2020[6]). Nevertheless, some authorities have already successfully deployed these solutions. For instance, the Australian Securities and Investments Commission (ASIC) developed a Market Analysis and Intelligence (MAI) platform, which collects real-time data feeds from all Australian primary (ASX) and secondary (Chi-X) capital markets for equity and equity derivatives products and transactions.

In particular, the MAI platform has a real-time alert monitor that detects and identifies abnormalities in securities trade. It also contains standard reports to allow analysts to drill down and analyse market data to identify trading accounts of interest that may be undertaking market misconduct such as insider trading and market manipulation. Overall, the standard dashboards within MAI include Real-Time Alert Monitor, Market Summary, Market Manipulation and Insider Trading Reports and the Market Replay, which allow for real-time or historical review of the market for a particular security.

In the European Union (EU), the German Federal Financial Supervisory Authority (BaFin) is setting up an integrated automated alarm and market monitoring system (ALMA) for analysing potential market abuse cases, including insider trading and market manipulation. Using BaFin's analytical data warehouse, visualisation software and initial AI techniques, ALMA performs evaluation of transaction data against the reporting data requirements provided by Article 26 of MiFIR[7], by extracting price-relevant information from ad hoc announcements and performing various pattern recognition functions. ALMA is also designed to be able to process information graphically via a dashboard, thus providing visual support for trading data analysts and providing them with a wide range of information in an easily accessible manner, including raw data, reporting data, directors' dealings, etc.

The Canadian Securities Administrators (CSA) also reported that it is developing its Market Analysis Platform (MAP), which collects post-trade data from exchanges, alternative trading systems (ATSs) and dealers/brokers. While the first phase delivered in October 2020 focused on equity securities, future phases will include a broader array of products traded on the exchanges and ATSs, including listed derivatives. In addition to facilitating enforcement

7 Article 26 of Regulation (EU) No 600/2014 on markets in financial instruments ("MiFIR") contains the obligation for credit institutions and investment firms which execute transactions in financial instruments to report complete and accurate details of such transactions to the competent authority.

investigation of potential insider trading and market abuse cases, the MAP system is also available for research and policy development.

Improving regulatory reporting

While some reports suggest that regulatory reporting has become increasingly complex, time-consuming and expensive for regulated entities[8], authorities face challenges related to collecting delayed and poor quality reporting data, which can in turn impact their ability to supervise (FCA, 2020[14]; European Commission, 2020[15]; European Commission, 2018[16]). In particular, as many authorities continue to rely on heavily manual processes, challenges remain as to how to make effective use of unstructured or qualitative data, such as information comprised within disclosure materials or annual reports. SupTech tools can be leveraged by authorities that must undertake complex, qualitative analyses to determine compliance with legislation or regulation that is often principle-based or comprises judgment-based rules (World Bank, 2018[4]). AI tools – including machine learning and natural language processing – are particularly relevant in that respect.

For instance, the Malaysian Securities Commission (SC Malaysia) uses AI to monitor the adoption of corporate governance best practices and quality of disclosures by listed companies on the Malaysia Stock Exchange (Bursa Malaysia). Since 2017, listed companies are required to report on their adoption of the Malaysian Code on Corporate Governance (MCCG) using a prescribed template for corporate governance reports. This template is designed to facilitate data extraction, evaluation and analysis by the AI system which considers *inter alia* the type of information disclosed, depth of explanation, and in relation to departures, the strength of alternative practices. The use of AI has enabled the SC to annually report data and observations in relation to the adoption of the MCCG and the quality of disclosures, including year-on-year progress, in the SC's Corporate Governance Monitor report. The data also supports evidence-based regulatory measures to improve corporate governance practices or address areas of concern, including practices with low score for disclosure.

With the aim of improving data collection, many authorities have also piloted the adoption of both "push" and "pull" technologies in recent years. While the former refers to pre-defined data being delivered from the regulated entity to

8 In a 2018 report, the European Commission estimated most firms' regulatory reporting costs at around 1% of total operating costs (European Commission, 2018[16]). Industry feedback suggests that the total burden on regulated entities is likely even higher, as the cost of building or amending reports tends to be higher than ongoing running costs. Several reasons can explain the increasing costs for supplying regulatory reports, including the challenge for firms to populate reports with the correct data; the spread of instructions across different pieces of interlinking regulation; unclear wording of rules; and firms subjected to multiple regulatory regimes having to submit differing reports containing similar underlying data (European Commission, 2018[16]; FCA, 2020[14]).

the regulator, the latter enables the authority to draw data from the regulated entity as required. Some authorities have also developed application programming interfaces (APIs) to allow regulated entities to submit data more easily, thus lowering reporting costs and enabling better communication between both parties (FSB, 2020[6]). For instance, the CSA is working to introduce APIs for its National Registration Database (NRD), which will give registered firms the opportunity to securely file information through their systems directly with NRD rather than making manual submissions. This will reduce administrative costs, regulatory burden and will allow them to update NRD information more efficiently and with better data quality.

Taking these efforts one step further, some authorities have begun exploring how to translate rules into a machine-readable format, in order to automate regulatory reporting and further facilitate compliance (World Bank, 2018[4]; Dias and Staschen, 2017[12]; European Commission, 2020[15]). This entails digitising reporting instructions and converting them into code to make them machine executable[9] (FCA, 2020[14]; Mohun and Roberts, 2020[17]; European Commission, 2020[15]). However, it is worth noting that while digitising regulatory reporting rules might entail additional benefits such as regulatory simplification, it is currently being hindered by the absence of common standards[10] (FSB, 2020[6]; European Commission, 2020[15]).

To address this challenge, the European Commission will develop a strategy on supervisory data in 2021 to help ensure that "(i) supervisory reporting requirements (including definitions, formats, and processes) are unambiguous, aligned, harmonised and suitable for automated reporting, (ii) full use is made of available international standards and identifiers including the Legal Entity Identifier, and (iii) supervisory data is reported in machine-readable electronic formats and is easy to combine and process" (European Commission, 2020[15]).

9 The European Commission is aiming to ensure that key parts of EU regulation are accessible to natural language processing, are machine readable and executable, and more broadly facilitate the design and implementation of reporting requirements. It will also encourage the use of modern IT tools for information sharing among national and EU authorities. As a first step in the domain of machine readable and executable reporting, the Commission has launched a pilot project for a limited set of reporting requirements (European Commission, 2020[15]). The digitisation of reporting instructions was also explored by the UK Financial Conduct Authority (FCA) and the Bank of England (BoE) during a TechSprint in late 2016, during which it was found that a small set of reporting instructions could be converted into machine-executable code, in turn enabling machines to use this code to automatically find and return regulatory reporting directly from a simulated version of a company's systems. Since then, work has progressed into a first and second phase involving the FCA, BoE and regulated banks (FSB, 2020[6]; FCA, 2020[14]; FCA, 2020[22]).

10 In Europe, some industry attempts to improve and standardise the reporting process have already been made through initiatives likes the Banks Integrated Reporting Dictionary (BIRD), Integrated Reporting Framework (IReF) and the European Banking Authority's Data (DPM) (ECB, 2021[39]).

Improving data management

The three main tasks within data management include validation, consolidation and visualisation – each referring to specific target points in the data management cycle. Validation refers to the quality control checks of completeness, correctness and consistency of data against reporting rules, whereas consolidation relates to the aggregation of data from multiple sources and in varying formats, and visualisation involves the presentation of information in a legible manner (di Castri et al., 2019[5]). A wide range of SupTech tools can be leveraged to improve data management – and in particular cloud computing, which allows for greater and more flexible storage, mobility capacity and computing power (Broeders and Prenio, 2018[18]).

For instance, Mexico's CNBV is currently implementing the second phase of a project involving cloud computing to process large amounts of anti-money laundering (AML) compliance data, thus allowing for a greater and more flexible storage, mobility capacity and computing power to support AML supervision of all supervised financial institutions. The platform will also enable the development of both basic and advanced, prospective analytics to strengthen monitoring activities and better identify atypical patterns. Likewise, Australia's ASIC is currently upgrading MAI from a non-cloud, Flex system to a cloud-based, HTML5 system, in order to ensure enhanced IT security standards within ASIC, and further improve its data ingestion, analysis and visualisation capabilities.

Challenges and risks of SupTech

Adopting SupTech solutions also comes with challenges and risks, including those that commonly arise in relation to large technology platform or software transitions, as well as risks that are transversal in nature due to the digital environment itself. The main issues and constraints principally revolve around data quality, resourcing, and skills. Challenges can also arise upon the integration of SupTech tools into legacy systems. Case studies reviewed for this Toolkit Note also identified insufficient communication across all stakeholders involved as a potential hindrance to the effective implementation of SupTech solutions. Technical issues and risks stemming from the digital nature of SupTech solutions also need to be accounted for, including risks related to: cyber and data security; third party dependencies; data localisation (potentially resulting in cross-border issues), as well as poor-quality algorithms or data, and opacity in the design and outputs of SupTech tools (i.e. a "black box issue" potentially entailing reputational risks).

Data quality, standardisation and completeness

SupTech applications rely on machine-readable data – i.e. in a format that can be processed by computer programmes. As such, quality, standardisation and completeness of data are key requirements and can pose major challenges, especially upon leveraging unstructured data collected from non-traditional sources of information (e.g. open source or social media). For instance, SC Malaysia mentions that getting the buy-in from listed companies to disclose the information and data in a structured manner was a key enabler to using AI, which required effort by listed companies to change their reporting format.

Providing sufficient amounts of quality data to build machine learning applications can also be an issue. For instance, in relation to its Project Apollo, Singapore's MAS reported the scarcity of training data – particularly expert reports associated with prosecution outcomes – as a main challenge. Having a sufficient volume of such data is a key requirement to continually improve the accuracy and robustness of the algorithms, and to validate Apollo's models and methodologies in order for its results to be admissible for use in a court of law. This is echoed by Colombia's SFC in relation to its project involving social media sentiment analysis, as significant amounts of data are required in order to train the algorithm to attribute positive or negative connotations to a sentiment – i.e. whether a sentiment is 'good' or 'bad'.

Algorithmic models and human oversight

In relation to its NLP application, the CNBV reported that having in place good communication channels between data scientists, NLP algorithms analysts and business units to combine their expertise and obtain better recommendations and continuous improvement of the NLP algorithms was a major challenge. This is linked to wider risks with regards to algorithms and their use by authorities. While algorithms can fail by detecting false positives/negatives rather than meaningful signals, there is also a risk of incorporating human biases in algorithmic models, as well as the risk of not being able to explain the outcomes of machine learning (i.e. a black-box issue that may impede accountability), all of which are exacerbated when authorities lack adequate skills and expertise. On the other hand, supervisors must also deal with the countervailing concern that if they are too transparent about the models used, regulated entities may be able to more easily game the system to avoid detection (Dias and Staschen, 2017[12]; Broeders and Prenio, 2018[18]; di Castri et al., 2019[5]).

In considering such challenges, SC Malaysia has highlighted the importance of ensuring that data scientists have a general understanding of corporate governance principles, practices and disclosures given that a basic understanding of corporate governance concepts is critical to ensure that the data scientists are able to formulate the logic that will be applied by the AI in analysing the adoption of corporate governance practices and the quality of

disclosures. As such, building AI capability requires not just more data but also better data. In this case, insightful and reliable corporate governance disclosures. In the developmental stage, a set of good disclosures by listed companies in Malaysia and other markets were selected and used to build the base of the AI.

Importantly, human intervention is required to identify and validate these disclosures in order to feed the development of the AI tool. Therefore, in order to yield benefits, SupTech tools require skilled human oversight – as technology should not be leveraged to substitute, but rather to complement and support human judgment. This has crucial financial stability implications, as tools built upon historical data associated with past instances of instability may not remain valid for predicting future crises (FSB, 2020[6]).

In addition, from a corporate governance enforcement perspective, as final decisions on whether to pursue enforcement actions are still necessarily taken by humans and based on human judgements, appeals mechanisms also provide a potential lever for considering and addressing potential biases that may be introduced through algorithmic or AI-based supervisory mechanisms. For instance, BaFin reports that defining the patterns and types of anomalies ALMA should look for represents a challenge, as the assessment of which incidents ALMA should identify as abuse is based on experience and should therefore be verifiable by analysts.

Third-party dependencies and digital security

Increased dependencies on third parties can also constitute a risk, especially with regard to cloud service providers. Although cloud-based services hold the potential to foster information sharing between authorities, and in turn improving regulatory co-operation, "public cloud" solutions raise operational, governance and oversight considerations, with particular relevance in a cross-border context, whereby authorities are unable to assess whether legal and regulatory obligations around the delivery of a service are being met. Further, interoperability limitations could create lock-in effects and over-reliance on specific platforms and providers (FSB, 2019[19]; FSB, 2017[2]).

As such, vetting and auditing processes may be required as a means to ensure adequate safeguards. In addition, greater reliance on outsourced data storage may also increase cyber vulnerabilities for authorities, which may in turn magnify financial stability risks. At present, most authorities store most of their data in-house for security reasons, and their use of cloud storage is reportedly limited to non-core activities (FSB, 2020[6]; FSB, 2019[20]).

Legacy systems

Legacy systems, along with data formats that are not compatible with SupTech, can also impede SupTech adoption. In particular, Mexico's CNBV reports that

the main obstacle to the implementation of its cloud computing project is the variety of technological infrastructure among the Mexican financial institutions. In the same vein, challenges can also arise upon the integration of SupTech tools into existing processes and procedures. For instance, Australia's ASIC reports that the rewrite of frameworks and dashboards may slightly alter legacy procedures.

BaFin also reports the setup of the technical infrastructure behind ALMA as a major challenge, as it requires integrating different databases, AI methods, a visualisation for the supervisors, a feedback mechanism and a consistent data flow through all the stages. Additionally, to work with large quantities of data, hardware needs to be updated permanently in order to guarantee a high performance. These obstacles entail that valuable product increments might be difficult to deliver even in several sprints, which might result in stakeholders being potentially dissatisfied over a longer period. This challenge also includes the need for a cultural change in the organisation to enable the whole team to work in an agile framework.

Likewise, in relation to its Data Collection Gateway, MAS reports that change management is a key challenge – as the new platform differs significantly from the current platform in terms of the collection mechanism, business process (such as exception remediation process) and technology. To address this concern, MAS has established multiple communication channels and regularly keeps tabs on financial institutions' implementation progress in order to identify potential issues early to facilitate timely resolution. Similarly, internal stakeholders are systematically engaged to reshape and fine-tune data-related processes so that MAS' business needs continue to be met.

Financial and human resources, procurement rules, and regulatory frameworks

Other challenges may be encountered when developing, deploying and maintaining SupTech solutions – including authorities' lack of adequate skills such as with respect to technology, software and hardware expertise, along with budget constraints, rigid procurement rules and obsolete regulatory frameworks. Resistance to change and organisational silos may also hinder the development of SupTech projects.

Regarding skills, for instance, BaFin's ALMA system design requires the combined expertise of database architects, data engineers, supervisors, data analysts and data scientists as well as of software engineers, thus requiring existing staff to undergo training as well as recruiting staff with expertise in data analytics. BaFin reports that bringing all these techniques and experts together is an ongoing challenge. Likewise, in relation to its social media sentiment analysis project, SFC reports that recruiting people who have advanced data science training to implement data analysis and autonomous

learning models stands as a major challenge. ASIC also reports resources and skills issues around the upgrading of MAI.

While design and testing resources are being drawn from current surveillance and enforcement teams, COVID-related remote working challenges also reduce local resources, and the communication of ideas can be more difficult at times (particularly in relation to training). In addition, the adoption of the cloud infrastructure by teams relatively new to the technology can also be a challenge. The CSA reports that securities regulators have started staffing themselves with data governance experts and employees with data analytics skills, and that some provincial regulators have also starting hiring Chief Data Officers in 2020 to address challenges raised by the complexity of data management and communication thereof. Chief Data Officers (CDOs) are usually responsible for managing data assets of an entity, and ensuring that its data needs are met.

As an illustration of issues encountered with regards to existing legal and regulatory frameworks, the CSA reported a challenge in relation to its project developing an integrated and comprehensive records filing and disclosure system (SEDAR+) in adapting the securities regulation to the new system. Indeed, while securities law in Canada is substantially harmonised, there remains some differences and unique local requirements across the provinces that must be accounted for. As such, the CSA must also undertake significant communication and change management activities to ensure that all system users, such as filers, investors, researchers and regulators are prepared for the change.

In addition, in relation to MAP, while the main challenge as reported by the CSA relates to the volume and quality associated with the development of a central depository of capital markets data across asset classes with a complex functionality toolset, another challenge is that the data collected into MAP is highly confidential. As such, legal and technical precautions are being taken to protect the trading data supplied by IIROC and broker client data provided by investment dealers.

Authorities' procurement rules may also render the design and implementation of technology solutions difficult, as evidence suggests that supervisors' procurement offices are often unfamiliar with these new technologies, and conversely, service providers are often unfamiliar with these procurement processes (di Castri et al., 2019[5]).

Considerations for devising adequate SupTech strategies

Recognising the potential of SupTech to transform data processes – in turn improving the timeliness and quality of decisions and actions – the use of SupTech tools by authorities has been gaining momentum in recent years.

According to a recent FSB survey (2020[6]), the use of SupTech strategies has grown significantly since 2016, with a vast majority of surveyed authorities having a SupTech or innovation or data strategy in place.

SupTech strategies are hereby defined as seeking to develop tools to support authorities' functions, whereas innovation/data strategies refer to institution-wide digital transformation/data-driven innovation (DT&DI) programmes that encompass the development of SupTech tools. They are not necessarily pursued in isolation (FSB, 2020[6]). SupTech applications can either be initiated by management, or originate as research questions. Evidence also suggests that SupTech applications can be explored through the use of accelerators, tech sprints, and innovation labs, regardless of whether an authority has an explicit SupTech strategy (Broeders and Prenio, 2018[18]; di Castri et al., 2019[5]).

Leadership, budget and skills

Overall, it is important that SupTech strategies be devised in consideration of authorities' needs, regulatory frameworks and technological capacities. Although there is no "one-size-fits-all" approach, authorities have identified several important considerations underpinning successful SupTech strategies, ranging from the design to the implementation stage, and covering leadership, budget and skills concerns.

A well-defined SupTech strategy requires effective leadership – such as through established CDOs – and management buy-in, as well as early engagement with end-users (i.e. 'front-line' supervisors) – which allows to overcome resistance to change. Evidence also suggests that adopting 'fast fails' approaches can enable authorities to quickly evaluate which applications merit further progress, and which ones are not fit for purpose (FSB, 2017[2]). Securing sufficient budget is also paramount for developing SupTech projects, along with adequate procurement systems.

Resources are also crucial, as having technologically skilled professionals in place with the right data expertise better enables the implementation of a flexible SupTech platform, and the adoption of a data-driven culture by organisations as a whole (Bank of England, 2019[21]; FCA, 2020[22]). Several authorities have implemented a strategy for attracting and retaining adequate skills and talent – such as through employee engagement frameworks, or by offering online training to existing staff to enhance their skills. Knowledge-based transfers between departments are also observed. In order to attain a skilled SupTech workforce, some financial services authorities have started tailoring their recruitment strategies to focus on candidates' data analysis skills (FSB, 2020[6]).

It should be noted that a "late mover" advantage applies to authorities that have recently initiated – or are considering to initiate – the development of their data infrastructure. Indeed, integrating advanced analytics tools to a data

architecture designed from scratch might prove an easier task than building new tools upon legacy systems (Coelho, De Simoni and Prenio, 2019[13]).

Collaboration between authorities, regulated entities and technology service providers within and across jurisdictions

While data analysis applications are developed to facilitate internal workflows (misconduct analysis and market surveillance solutions), data collection tools require some involvement from market participants (regulatory reporting and data management solutions). For the latter category, it is important to consult with regulated entities going forward to ensure that solutions adopted on both ends are aligned and compatible. As some supervisors have piloted and adopted SupTech frameworks on an ad-hoc and unco-ordinated basis, this can in turn create negative externalities for regulated entities. In particular, a lack of common standards – along with differing levels of technological progress within authorities – can lead to inconsistencies in SupTech approaches across jurisdictions (European Commission, 2020[15]).

In terms of automated reporting, for instance, firms with subsidiaries in more than one jurisdiction are currently unable to implement the same reporting solution for all subsidiary companies, due to cross-country variations in supervisory expectations and technological capacities (European Commission, 2020[15]). Co-ordination between authorities and regulated entities in their respective efforts to adopt innovative technologies is thus essential to ensure that their systems are aligned and compatible, in order to mitigate potential challenges and adverse effects down the line, as well as to allow both parties to reap maximum benefits from their use (Bank of England, 2020[23]).

An important caveat is that SupTech might induce market participants to adjust their behaviour accordingly. A recent study finds that authorities' adoption of SupTech solutions has a feedback effect on companies' corporate disclosure decisions, implying that companies adjust their filings when they anticipate that such disclosure will be processed by machines (Cao et al., 2020[24]). Other evidence suggests that market participants may seek to gain sufficient knowledge of SupTech applications to game the technology to their benefit (di Castri et al., 2019[5]).

Going forward, co-ordination and collaboration between authorities, regulated entities and technology service providers within and across jurisdictions is crucial to: 1) ensure the compatibility of innovative systems adopted by regulators and regulated entities; 2) foster peer learning with regards to the successes and failures of SupTech uses; and 3) consider the possibility of devising common standards and taxonomies for relevant regulatory areas in order to ensure the scalability and interoperability of SupTech tools, especially with regards to reporting solutions. By convening and fostering exchanges among a wide range of stakeholders, international organisations and standard-setting bodies can play an important role in that respect.

Annex. A selection of SupTech Initiatives

Improving misconduct analysis

Project Apollo

Responsible entity: Monetary Authority of Singapore (MAS)

Description: Project Apollo is an AI-based tool that is used, alongside other analytical frameworks, in the triaging of cases for investigation. The system automates the computation of key metrics used for trade analysis and predicts the likelihood that an expert will suspect that market manipulation has occurred. An interactive dashboard is provided for the visualisation of Apollo's results and predictions. The project was initiated in early 2018 and launched as a fully operational system in April 2020. It was intended to address pre-existing challenges, including: 1) the scarcity of people with the necessary skills, 2) the increasing demand for expert reports by criminal prosecutors, and 3) the increasing need for preliminary expert opinion during investigation. While the core functions of Project Apollo were developed in-house, the web application was developed in collaboration with external service providers.

Read more: *https://www.mas.gov.sg/-/media/MAS/News-and-Publications/Monographs-and-Information-Papers/MAS-Enforcement-Report.pdf*.

Natural Language Processing (NLP) to detect AML/CFT infringements

Responsible entity: Mexico's National Banking and Securities Commission (CNBV)

Description: As the rise of digital financial products and services poses new challenges for Mexico's financial authorities, traditional methods and models of capturing and analysing regulatory data are ill-suited to cope with the surfeit of data being generated by new platforms, products, and customers. Against this backdrop, the CNBV has developed a prototype for a Natural Language Processing (NLP) application to detect what a suspicious AML/CFT network is 'talking about', thus enabling the detection of unusual transactions, relationships, and networks events to identify potential money laundering issues that cannot be identified by humans. As the prototype was concluded, it is now at the experimental stage. It was developed by external service providers.

Read more:
https://static1.squarespace.com/static/583ddaade4fcb5082fec58f4/t/5bbf9a520d929792c05866e1/1539283556587/R2A+CNBV+Case+Study.pdf.

Machine Learning tools to analyse unstructured data

Responsible entity: Colombia's Financial Superintendency (SFC)

Description: A team within the Digital Supervisory Center of the Financial Superintendency of Colombia (SFC) has recently started compiling information from different sources and carrying out daily, weekly and monthly analyses in order to understand what has happened and anticipate market behaviour. In order to improve the recollection and analysis of non-structured data, the project allows the authority to: 1) extract news, specialised analysis and social network information; 2) consolidate data in a standardised framework; and 3) implement machine learning models to analyse market sentiment on macroeconomic environment and among different economic sectors. In particular, the project seeks to: 1) facilitate non-structured information analysis, 2) generate information repositories to support current and historical analysis, and 3) develop perception measurement tools to have real-time information on economic performance. Benefits have resulted from its implementation, as it has generated internal knowledge related to web scrapping, text mining and machine learning implementation that can be replicated by other departments in the Financial Superintendence. It has also increased information analysis capacity, and has enabled the automation of the information gathering process.

Read more: *https://www.superfinanciera.gov.co/*.

Improving market surveillance

Market analysis and intelligence (MAI) platform

Responsible entity: Australian Securities and Investments Commission (ASIC)

Description: The Australian Securities and Investments Commission (ASIC) has developed a Market Analysis and Intelligence (MAI) platform, which collects real-time data feeds from all Australian primary (ASX) and secondary (Chi-X) capital markets for equity and equity derivatives products and transactions. The MAI platform has a real-time alert monitor that detects and identifies abnormalities in order and trade messages in traded securities. It also contains standard reports to allow analysts to drill down and analyse market data to identify trading accounts of interest that may be undertaking market misconduct such as insider trading and market manipulation. Overall, the standard dashboards within MAI include Real-Time Alert Monitor, Market Summary, Market Manipulation and Insider Trading Reports and the Market Replay, which allow for real-time or historical review of the market for a particular security. The MAI platform was preceded by the SMARTS market intelligence system (MSS).

ASIC is currently in the process of upgrading MAI from a non-cloud, Flex system to a cloud-based, HTML5 system. A key driver for this upgrade is to ensure IT security standards within ASIC. As an additional benefit of upgrading MAI, ASIC will move to the latest version of its current vendor's platform which includes enhanced functionality to process analyse and visualise data. In addition, ASIC intends to leverage the enhanced functionality of the upgraded MSS to increase its surveillance capabilities of the FICC markets and further utilise information received from the Australian Tax Office. This work is being undertaken in-house and is experimental/in-development. This capability is being developed on the upgraded MAI system's sandbox environment called Kx Analyst. Datasets that will be ingested include OTC Trade Repository Data, Bond Clearing information from Austraclear and Global Legal Entity Identifier data.

The Kx Analyst environment uses proprietary KDB+ technology and interfaces with various open source languages such as Python and R, providing ASIC analysts with a single data science environment. ASIC currently receives trading account information and their related relationship information, including spousal and residential and business address information from the Australian Tax Office. From this information, ASIC has created a data set of an anonymised map of linked trading accounts. This data set will be ingested into KX Analyst and will be linked to MAI trading data to create different analytics to improve ASIC's market surveillance capability of identifying market misconduct.

Read more: *https://asic.gov.au/about-asic/corporate-publications/newsletters/market-integrity-update/market-integrity-update-issue-69-march-2016/*.

Automated Alarm and Market Monitoring System (ALMA) Project

Responsible entity: German Federal Financial Supervisory Authority (BaFin)

Description: BaFin started developing the ALMA project in 2017, using its analytical data warehouse (ADW) coupled with a visualisation software and initial AI techniques. For the automated identification of cases of insider trading in securities, ALMA extracts price-relevant information from ad hoc announcements and performs a variety of pattern recognition functions. The ALMA project does not integrate new requirements in the context of a traditional concept phase. All of the participants engaged in this project within BaFin form part of a learning process in handling the data and are working out requirements, proto- and pseudocodes and indicators on an interdisciplinary basis.

The information required is described formally in a structured query language (SQL) so that it is located by the database system of the ADW in the gigantic volume of data in an optimised reaction time. For the data analysts in the relevant divisions, these indicators and SQLs form the basis for learning about

ordinary trading activities from the system. This enables irregularities to be identified automatically and to be used to support the decision-making process.

ALMA integrates these algorithms and techniques in a single, intuitive visual interface for users at BaFin. The data transmitted to BaFin are stored in the ADW for this purpose via a number of process stages. The volume of data will be multiplied further in future as a result of new regulations coming into force.

Read more: *https://www.bafin.de/EN/PublikationenDaten/Jahresbericht/Jahresbericht2017/jahresbericht_node_en.html*.

Market Analytics Platform (MAP)

Responsible entity: Canadian Securities Administrators (CSA)

Description: The CSA is completing the replacement of its markets analysis platform used by enforcement staff for analysing potential insider trading and market manipulation violations within Canadian exchanges and alternative trading systems (ATS), as the old system is now obsolete and was mainly used to assist with assessment of market manipulation and insider trading cases. The CSA is aiming to build its own data repository, expand the use of market data and broker data analysis across multiple marketplaces and asset classes, as well as to the research and policy development areas. As the capital markets have evolved significantly over the last decade, the CSA currently requires extensive historic records of complex structural information, which necessitate upgraded technological foundations, tools and applications.

As a data repository and analysis platform, the Market Analytics Platform (MAP) brings together data from the exchanges and ATSs, through the market oversight function of the Investment Industry Regulatory Organization of Canada (IIROC), with broker client data from the investment dealers. The first phase of the project was delivered in October 2020. The MAP data will also include a broader array of products being traded on the exchanges and ATSs: for example, derivatives trading will be included in addition to equity securities. In addition to enforcement purposes such as identifying, assessing and investigating potential market abuse cases, non-personal data in the MAP system will be available to CSA staff for research and policy development, which was not the case with the previous system. The key challenges encountered while developing MAP relate to the volume and quality of data associated with the development of a central depository of capital markets data across asset classes with a complex functionality toolset.

Read more: *https://www.securities-administrators.ca/aboutcsa.aspx?id=1735&terms=Market%20Analytics%20Platform%20(MAP)*.

Improving regulatory reporting

Use of AI to evaluate the quality of corporate governance disclosures by listed companies

Responsible entity: Malaysian Securities Commission (SC Malaysia)

Description: SC Malaysia uses AI – including machine learning and natural language processing – to monitor the adoption of corporate governance best practices and quality of disclosures by listed companies on the Malaysia Stock Exchange (Bursa Malaysia). Since the revision of the Malaysian Code on Corporate Governance (MCCG) in 2017, listed companies are required to report on their adoption of the MCCG using a prescribed template for corporate governance reports, introduced to improve readability and comparability of information, and designed to be system-friendly (e.g. use of drop-down options to ensure standardisation of entry) to enable data extraction, evaluation and analysis by AI, which considers among others the type of information disclosed, depth of explanation, and in relation to departures, the strength of alternative practices. There is an evaluation parameter for all 36 best practices in the MCCG.

Prior to these enhancements, including the use of AI, monitoring the adoption of the Malaysian Code on Corporate Governance and the quality of disclosures by either regulators or investors required extensive resources to undertake heavily manual activities, posing some challenges, including to obtain current data and observations on the adoption of best practices. Further, the observations or data would be relatively dated, given the time required for manual extraction of the information. The use of AI has enabled the SC to annually report data and observations in relation to the adoption of the MCCG and the quality of disclosures, including year-on-year progress, in the SC's Corporate Governance Monitor report. The data also supports evidence-based regulatory measures to improve corporate governance practices or address areas of concern – including practices with low score of for disclosure.

Read more:
https://www.sc.com.my/api/documentms/download.ashx?id=ff69ce0d-a35e-44d4-996a-c591529c56c7.

Data Collection Gateway (DCG)

Responsible entity: Monetary Authority of Singapore (MAS)

Description: MAS' new regulatory reporting platform – Data Collection Gateway (DCG) – aims to uplift the MAS' data collection capabilities by addressing challenges faced by both MAS and financial institutions using the current platform. These challenges include limitations on the complexity of data-validation rules, scalability of the collection mechanism to support large

datasets as well as constraints on agility due to the lengthy turnaround times. With DCG, more complex business validations can be embedded, which will help improve the quality of data submitted. Validations are performed on data submitted and relevant exceptions are raised to submitters for remediation before the submission is transmitted for downstream processing, thus reducing the extent of manual validation required by MAS' data administrator.

The project was initiated in early 2019, with the pilot based on one of MAS' largest regulatory submissions. The pilot was launched in April 2020 when the gateway was made available for use by financial institutions for testing. The systems were developed by external service providers. Prior to project initiation, MAS gathered feedback from financial institutions on data collection. With the feedback gathered, the project team scanned the market to find out i) how other regulators were collecting data and ii) what platforms were available before detailing a set of desired functional requirements that was used to procure a solution from the market.

As part of the change management effort, the project team engaged the financial institutions regularly to understand the challenges faced submitting data via the DCG during the pilot phase and address the challenges, where possible. On the operations tracking front, regular usage statistics were also compiled to understand how financial institutions were interacting with the system during the pilot phase. A key part of MAS' digital transformation strategy is to involve stakeholders (i.e. financial institutions) as early as possible. This ensures that expectations are aligned up front, and that the necessary investments/expenditures are accounted for as early as possible.

Read more:
https://masnet.mas.gov.sg/portal/masnet/vmedia/get/?path=1034/dcg-user-guide.pdf.

SEDAR+ National Filing System

Responsible entity: Canadian Securities Administrators (CSA)

Description: The CSA currently has several national systems where issuers, insiders and registrants (such as brokers, dealers and financial advisers) file various reports, documents and information mandated by CSA members' securities laws and regulations. Along with a number of separate databases, such as the Cease-Trade Order Database or Disciplined List, these systems also provide disclosure information for use by the investing public. Historically, the CSA national filing systems were developed as separate systems and databases without a view to having comprehensive, high quality data for use in advanced analytics for investing or regulation. Collecting data in separate systems does not allow for data integration and cross-reference. In addition, outdated technologies also may represent challenges from a digital security standpoint.

As such, the CSA is developing an integrated and comprehensive records filing and disclosure system (named SEDAR+) to unify and modernize its existing national systems and databases, which include: System for Electronic Document Analysis and Retrieval (SEDAR), System for Electronic Disclosure by Insiders (SEDI), Cease-Trade Order Database (CTO), National Registration Database (NRD), National Registration Search (NRS), Disciplined List (DL), Various filings currently made in paper format or in local electronic filing systems.

To ensure a smooth transition, SEDAR+ will be rolled out in phases. The first phase, which will replace SEDAR, the CTO Database, the Disciplined List and certain filings made in paper format or in local electronic filing systems, is expected to launch in late 2021. When completed, SEDAR+ will be a web-based portal that will function as a single point of contact among all national systems and provide easier access for filers and investors alike, thus enabling better data collection and search capabilities, improved usability, increased speed and enhanced cybersecurity.

Read more: *https://www.securities-administrators.ca/systems_replacement.aspx?id=1995&terms=sedar%20*.

Improving data management

Cloud computing to process large data volumes

Responsible entity: Mexico's National Banking and Securities Commission (CNBV)

Description: The CNBV is currently implementing the second phase of a project involving cloud computing to process large amounts of anti-money laundering (AML) compliance data, thus allowing for a greater and more flexible storage, mobility capacity and computing power to support AML supervision of all supervised financial institutions. This is also aimed at strengthening monitoring of financial institutions, through prospective analysis and the identification of atypical patterns. The rationale for developing this project is the existence of diverse data formats and the lack of a data warehouse system, data validation processes and consolidation mechanisms of AML compliance data of more financial institutions. Overall, the platform will enable the development of basic and advanced analytics to strengthen monitoring activities.

Read more: *https://www.cnbv.gob.mx/*.

Digital forensics

Responsible entity: Japan's Securities and Exchange Surveillance Commission (SESC)

Description: SESC investigations are getting complex and bigger because of the diversification of electric devices, improvement on security and the prevalence of new IT services such as cloud services. To cope with these changes, SESC is developing an advanced environment for the preservation, restoration, analysis, and storage of electronic data and enhancing digital forensic technology. In 2015, the SESC launched the Digital Forensics Strengthening Project Team, which consolidates the digital forensics personnel and digital forensics tools assigned to each section of the SESC and examines various issues such as environmental improvement. The Digital Forensic Solutions Office was established in the same year as a section that efficiently and effectively utilises appropriate personnel and tools for inspections and surveys and systematically develops human resources through centralised management. The Digital Forensic Solutions Office has been maintaining a server to handle the increase in the capacity of data acquired in inspections and surveys, collecting, selecting and procuring information such as the performance and trends of digital forensic tools, and improving and maintaining an environment for data analysis and browsing.

Read more: *https://www.fsa.go.jp/sesc/*.

References

Arner, D., J. Barberis and R. Buckley (2016), *FinTech, RegTech and the Reconceptualization of Financial Regulation*, https://ssrn.com/abstract=2847806. [35]

BaFin (2017), *BaFin's 2017 Annual Report*, https://www.bafin.de/EN/PublikationenDaten/Jahresbericht/Jahresbericht2017/jahresbericht_node_en.html. [40]

Bank of England (2020), *Transforming data collection from the UK financial sector*, https://www.bankofengland.co.uk/paper/2020/transforming-data-collection-from-the-uk-financial-sector. [23]

Bank of England (2019), *The Future of Finance Report*, https://www.bankofengland.co.uk/-/media/boe/files/report/2019/future-of-finance-report.pdf?la=en&hash=59CEFAEF01C71AA551E7182262E933A699E952FC. [21]

BCBS (2018), *Sound Practices: implications of fintech developments for banks and bank supervisors*, https://www.bis.org/bcbs/publ/d431.htm. [3]

Broeders, D. and J. Prenio (2018), *Innovative Technology in Financial Supervision (Suptech)-the Experience of Early Users*, https://www.bis.org/fsi/publ/insights9.htm. [18]

Cao, S. et al. (2020), *How to Talk When a Machine is Listening: Corporate Disclosure in the Age of AI*, https://dx.doi.org/10.2139/ssrn.3683802. [24]

CCAF (2019), *The Global RegTech Industry Benchmark Report*, https://www.jbs.cam.ac.uk/wp-content/uploads/2020/08/2019-12-ccaf-global-regtech-benchmarking-report.pdf. [31]

Coelho, R., M. De Simoni and J. Prenio (2019), "Suptech applications for anti-money laundering", Vol. FSI Insights on policy implementation, no 18, August, https://www.bis.org/fsi/publ/insights18.htm. [13]

Danielsson, J., R. Macrae and A. Uthemann (2017), *Artificial intelligence, financial risk management and systemic risk*, http://158.143.149.9/sites/default/files/downloads/publications/SP-13.pdf. [36]

di Castri, S., M. Grasser and A. Kulenkampff (2020), *The 'DataStack': A Data and Tech Blueprint for Financial Supervision, Innovation, and the Data Commons*, https://ssrn.com/abstract=3595344. [27]

di Castri, S., M. Grasser and A. Kulenkampff (2018), *An AML SupTech Solution for the Mexican National Banking and Securities Commission (CNBV): R2 A Project Retrospective and Lessons Learned*, https://static1.squarespace.com/static/583ddaade4fcb5082fec58f4/t/5bbf9a520d929792c05866e1/1539283556587/R2A+CNBV+Case+Study.pdf. [38]

di Castri, S. et al. (2019), *The Suptech Generations*, https://www.bis.org/fsi/publ/insights19.htm. [5]

Dias, D. and S. Staschen (2017), *Data Collection by Supervisors of Digital Financial Services*, https://www.cgap.org/sites/default/files/researches/documents/Working-Paper-Data-Collection-by-Supervisors-of-DFS-Dec-2017.pdf. [12]

ECB (2021), *The ESCB's long-term approach to banks' data reporting*, https://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/reporting/html/index.en.html. [39]

ESMA (2019), *Report on Trends, Risks and Vulnerabilities*, https://www.esma.europa.eu/sites/default/files/library/esma50-report_on_trends_risks_and_vulnerabilities_no1_2019.pdf. [8]

European Commission (2020), *Digital Finance Strategy for the EU*, *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0591*. [15]

European Commission (2019), *Expert Group on Regulatory Obstacles to Financial Innovation (ROFIEG):30 Recommendations on Regulation, Innovation and Finance: Final Report to the European Commission*, *https://ec.europa.eu/info/sites/info/files/business_economy_euro/banking_and_finance/documents/191113-report-expert-group-regulatory-obstacles-financial-innovation_en.pdf*. [25]

European Commission (2018), *Summary Report of the Public Consultation on the Fitness Check on Supervisory Reporting*, *https://ec.europa.eu/info/sites/info/files/2017-supervisory-reporting-requirements-summary-report_en.pdf*. [16]

FCA (2020), *Data Strategy*, *https://www.fca.org.uk/publications/corporate-documents/data-strategy*. [22]

FCA (2020), *Digital Regulatory Reporting: Phase 2 Viability Assessment*, *https://www.fca.org.uk/publication/discussion/digital-regulatory-reporting-pilot-phase-2-viability-assessment.pdf*. [14]

FinCoNet (2020), *SupTech Tools for Market Conduct Supervisors*, http://www.finconet.org/FinCoNet-Report-SupTech-Tools_Final.pdf. [11]

FSB (2020), *The Use of Supervisory and Regulatory Technology by Authorities and Regulated Institutions*, https://www.fsb.org/wp-content/uploads/P091020.pdf. [6]

FSB (2019), *FinTech and market structure in financial services*, https://www.fsb.org/wp-content/uploads/P140219.pdf. [20]

FSB (2019), *Third-party dependencies in cloud services: Considerations on financial stability implications*, https://www.fsb.org/2019/12/third-party-dependencies-in-cloud-services-considerations-on-financial-stability-implications/. [19]

FSB (2017), *Artificial Intelligence and Machine Learning in Financial Services: Market Developments and Financial Stability Implications*, https://www.fsb.org/wp-content/uploads/P011117.pdf. [2]

Globalscape (2017), *The True Cost of Compliance with Data Protection Regulations: Benchmark Study of Multinational Organisations*, https://www.globalscape.com/resources/whitepapers/data-protection-regulations-study. [29]

IBM (2018), *The Four V's of Big Data*, https://www.ibmbigdatahub.com/infographic/four-vs-big-data. [9]

IOSCO (2020), *The use of artificial intelligence and machine learning by market intermediaries and asset managers: Consultation Report*, https://www.iosco.org/library/pubdocs/pdf/IOSCOPD658.pdf. [34]

Mohun, J. and A. Roberts (2020), "Cracking the code: Rulemaking for humans and machines", *OECD Working Papers on Public Governance*, No. 42, OECD Publishing, Paris, https://dx.doi.org/10.1787/3afe6ba5-en. [17]

OCC (2020), *OCC Assesses $400 Million Civil Money Penalty Against Citibank*, https://occ.treas.gov/news-issuances/news-releases/2020/nr-occ-2020-132.html. [33]

OECD (2019), *Going Digital: Shaping Policies, Improving Lives*, OECD Publishing, Paris, https://doi.org/10.1787/9789264312012-en. [1]

OECD (2019), *Recommendation of the Council on Artifical Intelligence*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. [37]

OECD (2019), *The Path to Becoming a Data-Driven Public Sector*, OECD Digital Government Studies, OECD Publishing, https://doi.org/10.1787/059814a7-en. [41]

OECD (2018), *G20/OECD Policy Guidance on Financial Consumer Protection Approaches in the Digital Age*, https://www.oecd.org/finance/G20-OECD-Policy-Guidance-Financial-Consumer-Protection-Digital-Age-2018.pdf. [30]

OECD (2015), *G20/OECD Principles of Corporate Governance*, OECD Publishing, Paris, https://doi.org/10.1787/9789264236882-en. [7]

R²A (2019), *The State of RegTech: The Rising Demand for "Superpowers"*, https://static1.squarespace.com/static/583ddaade4fcb5082fec58f4/t/5b926ac0575d1f6f95502741/1536322258686/The+State+of+RegTech+-+The+Rising+Demand+for+%E2%80%9CSuperpowers%E2%80%9D_standard.pdf. [10]

SC Malaysia (2020), *Corporate Governance Monitor 2020*, https://www.sc.com.my/api/documentms/download.ashx?id=ff69ce0d-a35e-44d4-996a-c591529c56c7. [28]

The New York Times (2020), *Citigroup is fined $400 million over 'longstanding' internal problems*, https://www.nytimes.com/2020/10/07/business/citigroup-fine-risk-management.html. [26]

World Bank (2018), *From Spreadsheets to Suptech : Technology Solutions for Market Conduct Supervision*, https://openknowledge.worldbank.org/handle/10986/29952. [4]

World Bank and CCAF (2019), *Regulating Alternative Finance: Results from a Global Regulator Survey*, https://www.jbs.cam.ac.uk/wp-content/uploads/2020/08/2019-11-ccaf-regulating-alternative-finance-report.pdf. [32]