

SupTech: Leveraging Technology for Better Supervision

JULY 2018

Contents

- Introduction..... 2**
- Common Challenges faced by Supervisors..... 2**
 - Data Collection, Management and Governance 3
 - Data Analyses and Other Procedures..... 3
 - Data Analysis 3
 - Working Tools and Processes 4
- Current and Emerging SupTech and its Benefits 4**
 - The Potential Benefits of SupTech and Examples of Uses..... 4
 - Supervisory Data Collection and Management..... 4
 - Data Analytics..... 6
 - Supervisory Processes 8
- Risks and Challenges in Adopting SupTech..... 8**
- Conclusion 10**
- Annex: Glossary of Terms..... 12**
- Key References 14**
- Additional Readings..... 14**

This document was prepared exclusively for use in association with programs offered by Toronto Centre. The information in this note has been summarized and is made available for learning purposes only. It should not be regarded as complete or accurate in every detail. No part of this document may be reproduced, disseminated, stored in a retrieval system, used in a spreadsheet, or transmitted in any form without the prior written permission of Toronto Centre.

Toronto Centre and Toronto Centre logo are trade-marks of Toronto Leadership Centre.
© Copyright Toronto Leadership Centre 2018. All rights reserved.

Introduction¹

Technological innovation is revolutionizing the financial services industry and is poised to transform financial supervision as well. There are at least two reasons for supervisors to adopt supervisory technology (SupTech). First, the technology available today could help supervisors achieve greater efficiency and effectiveness in pursuing their goals. Second, without investing in technology, supervisors may be unable to deal with developments in the financial sector itself (such as the rise of FinTech) and any possibly related expansion of their statutory mandates.²

The FinTech industry has been investing heavily in RegTech and SupTech. The former aims at improving risk management and compliance at financial services providers (FSPs), while the latter supports financial supervision. The Basel Committee for Banking Supervision (BCBS) defines SupTech as the use of technologically enabled innovation by supervisory authorities.³ This Note focuses on SupTech solutions that rely on three major technological developments: a) exponentially higher and more affordable data storage capability and computer processing power that, in the past, required prohibitive investments; b) the availability and diversity of digital data, due to the digitization of the economy and of financial services in particular; and c) the recent advances in artificial intelligence (AI), such as machine learning (ML). These developments support the use of technologies such as ML, cloud computing, natural language processing (NLP), big data analytics, and distributed ledger technologies (DLT) for supervisory purposes in a cost-effective manner.⁴ SupTech can be built by external vendors (large or small), in-house by supervisors, or a combination of the two.

SupTech can involve improving regulatory data, automating and streamlining procedures and working tools, or significantly enhancing data analytics (e.g., faster data crunching and newer, richer and more complex analyses that are impractical using conventional analytics and data). In particular, the integration of multiple types of data in supervisory analysis can increase accuracy and efficiency of decision-making, shifting resources away from process-oriented tasks. In doing so, SupTech could help better allocate supervisory resources and become a key ally in supervisors' quest for more effective risk-based supervision.⁵ However, without quality data, SupTech will have limited impact. Increasing quality of regulatory reporting should be the starting point to create a solid basis for advanced analytics enabled by SupTech.

The focus of this Note is on how SupTech can help supervisors meet common challenges in discharging their responsibilities, with focus on data collection, data analytics and digitization of processes and working tools. The Note provides examples of current uses of SupTech and uses that are still in trial stage. Finally, it discusses the challenges and risks of embracing SupTech. A Glossary of Terms is found in the annex.

Common Challenges faced by Supervisors

Supervisory activities, from licensing to enforcement, rely on data, internal procedures and working tools, and human and other resources. While supervisors in emerging and developing economies (EMDEs) are likely to face greater obstacles in discharging their responsibilities due to lack of adequate regulatory powers, regulations, expertise or resources, virtually all supervisors face, to varying degrees, challenges

¹ This Note was prepared by Denise Dias on behalf of Toronto Centre.

² This TC Note builds off and expands upon the TC Note [FinTech, RegTech and SupTech: What They Mean for Financial Supervision](#), 2017. Detailed discussion of the impact of FinTech for supervisors is found in FSB (2017 and 2017a) and BCBS (2017).

³ BCBS (2018) p 42.

⁴ See Glossary of Terms in the annex.

⁵ See the TC Note *Risk-Based Supervision*, 2018.

such as low data quality and time-consuming manual procedures. This section summarizes these challenges by organizing them into two areas: i) data collection, management and governance, and ii) data analysis and other procedures.

Data Collection, Management and Governance

Supervisors use data from a variety of sources, the core of which is reported by FSPs, using templates dictated by the supervisory agencies. The template-based approach to regulatory reporting presents weaknesses that impact data quality, such as: long time-to-report; inaccuracies; burdensome validation; duplicate and inconsistent data across templates; high compliance costs for FSPs; and relative inflexibility for quick changes to reporting requirements. Many of these issues arise due to manual intervention at FSPs and supervisory agencies, inadequate information system infrastructure, and incompatibility between reporting standards and the taxonomies used internally by FSPs.⁶ Supervisors may lack coordination or adequate information system infrastructure to integrate reporting requirements across departments, increase speed of data transfer and processing, store large volumes of data, automate submission management and validation, and share data internally and with other authorities.

In addition, supervisors use a range of other data from external sources such as publicly available information, media reports, and data from other authorities. This data is most often unstructured, comes in a wide variety of formats (e.g., Word, HTML, PDF), and may be collected, stored and duplicated in a multitude of personal or shared folders. These characteristics impair the efficient use of such data. With the adoption of more risk-based and intrusive supervisory approaches leading to an increase in the volume, frequency and granularity of reporting requirements, these weaknesses could be an important impediment to effective supervision and pose high compliance costs on FSPs.

Data Analyses and Other Procedures

Data Analysis

A key element of supervision is the analysis of regulatory returns to feed macro- and micro-prudential monitoring, ad hoc assessments, as well as other supervisory objectives, such as AML/CFT compliance and consumer protection. While many supervisors in developed economies use sophisticated tools for analyzing reported data, those in EMDEs often rely on outdated tools (e.g., Excel) and manual procedures for extracting, consolidating, analyzing, visualizing, and reporting, due to the lack of adequate hardware, software or expertise. Even if data quality is high, poor data analytics still can lead to suboptimal decision-making.

In addition to regulatory returns, supervisors analyze less structured data, varied in format and content. This can include external structured data (e.g., complaints data from financial ombudsmen, data from national statistics agencies), and unstructured data such as management, auditor, annual and other reports by FSPs, self-assessments, documents with narrative texts downloaded from the internet (e.g., websites of FSPs), consumer surveys, and documents produced and/or stored at the supervisory agency (e.g., inspection reports, decision memos, meeting minutes, communications with FSPs, etc.). Searching, finding, sorting, analyzing and producing insights from such unstructured data – which may well constitute the majority of relevant data for supervisory purposes – remain major challenges for supervisors.

⁶ Ravi Menon, Managing Director of the Monetary Authority of Singapore (MAS) summarizes the challenges with manual procedures: “Today, financial institutions’ data submissions to MAS often involve manual processes to extract the information from their databases and fill up the MAS-provided form or template. And over at MAS, processing that data is also done manually.” <https://www.bis.org/review/r171115a.htm>

Working Tools and Processes

Both offsite and onsite work require tools to help plan, implement, and record the steps taken by supervisors. The level of standardization of tools such as inspection plans, interview guides, and inspection reports, varies widely across countries, but there is no doubt that achieving formality and standardization is easier when using digital tools. Most supervisors in developed economies use software to help organize, standardize and record supervisory activities. Many EMDE supervisors lack such tools. Also, even when software is used to organize tasks and produce documents, there might be insufficient or inefficient follow-on consumption of their content, given their unstructured nature.

Processing licensing and other applications can be time-consuming because of the use of manual procedures for checking sufficiency and formality of applications, gathering and analyzing existing supervisory or publicly available data, running background checks, analyzing business and financial feasibility plans, and gathering opinions across departments and authorities. Not all supervisors have digital and interactive tools to increase efficiency. Similarly, enforcement is often lengthy and based on manual processes from investigation and evidence-gathering to negotiation and sanction, relying mostly on unstructured data.

Current and Emerging SupTech and its Benefits

The Potential Benefits of SupTech and Examples of Uses

SupTech offers opportunities for automation, greater mobility, better data and data management and enhanced analytics. It could lead to lower costs over time and better allocation of supervisory resources by gradually shifting them from “robotic” tasks (e.g., data crunching and paper gathering) to tasks that depend on human judgment, expertise and experience. By increasing efficiency and effectiveness, SupTech could contribute in the quest for better risk-based supervision.

Supervisory Data Collection and Management

Without quality data (i.e., accurate, timely, sufficient) no SupTech tool has real value. SupTech enables supervisors, at a minimum, to shift away from the traditional template-based reporting toward the automated collection of highly granular data that can be used to generate consistent aggregates for different reports. SupTech using ML can spot quality issues such as data gaps, inconsistencies, and errors, and automate data cleaning, consolidation, validation and quality assurance. While the standardization of highly granular data requires close coordination between supervisors, IT vendors and the industry, such reform could integrate reporting requirements across departments and authorities, leading to less complexity, greater transparency and lower costs in the long run. High quality granular data can increase cross-entity and cross-time comparability, create new avenues for analysis, and reduce the burden of validating data at the aggregate level. It could also facilitate the harmonization of internal risk data (for the FSPs’ own purposes) and regulatory data.⁷

One example is the reporting system implemented by the Austrian central bank (OeNB) in 2015, a type of input-based approach. Standardized highly granular data (“basic cubes”) is inputted by banks to a database at AuRep, a company owned by the banks. AuRep works as a bridge between banks’ systems and OeNB’s, and as a single data warehouse whose costs are shared among the banks. OeNB, instead of collecting formatted templates, uses automatically transforms the “basic cubes” into desired reports. The reporting of granular data allows OeNB to change report formats with minimum or no impact on FSPs.⁸

⁷ Such harmonization could facilitate meeting the BCBS’s Principles for Effective Risk Data Aggregation and Risk Reporting (known as BCBS 239).

⁸ More detail in https://www.bis.org/ifc/events/ifc_isi_2015/010_turner_presentation.pdf.

One of the key aspects of this project was the creation of a single data model integrating the data needs of all OeNB departments.⁹

Another approach is the one being piloted by the Philippines central bank. An Application Program Interface (API) was developed for banks to automatically report highly granular and near real-time data. The tool offers many back-office functions, such as automated validation and data visualization, and report customization. Among other objectives, the central bank intends to reallocate staff currently dedicated to manual data validation, improve data analytics, and reduce non-compliance with reporting obligations.

Yet a different option is the data pull approach, in which the supervisory agency pulls the data automatically from the systems of FSPs, according to pre-determined specifications, and assumes the task of standardizing it and transforming it into desired reports, when the data is already in its own database. The Rwandan central bank implemented such a solution in 2017.

Emerging approaches to regulatory reporting would make reporting requirements digestible by the FSPs' operational systems. This could, in turn, fully automate the production of supervisory reports in a straight-through process that is not reliant on human interpretation and intervention. Referred to as "machine-readable" or "machine-executable regulation", this approach requires that supervisors issue regulation in the form of software (code) that is then run by the FSPs' systems. Both the Monetary Authority of Singapore (MAS) and the UK's Financial Conduct Authority (FCA) are exploring this approach. The MAS has set the goal of making all data requests through machine-readable code: "the data will flow seamlessly from financial institutions' databases to our forms and ultimately to the supervisory dashboards."¹⁰ The UK's FCA launched in early 2018 a Call for Input¹¹ for the use of DLT and ML for "smarter regulatory reporting".

DLT is also being explored by security regulators to reduce complexity and costs of voluminous reporting, as it could provide a high level of data security and integrity while keeping costs of data aggregation and transfer manageable. According to the ECB,¹² a DLT regulatory reporting model could provide market participants with a transparent, secure and trusted rule-set of reporting obligations.

There are not many examples of holistic SupTech solutions to integrate the collection of structured standardized reporting with the collection of unstructured data from external and internal sources. There are emerging and established tools, for instance, to automate the search, download and storage of unstructured data. These are being combined with, for instance, ML and NLP to conduct preliminary analysis to save supervisors' time. The UK's FCA is piloting a tool to collect standard consumer agreements from websites of FSPs, check compliance with basic mandatory disclosures and identify abusive contractual clauses that can harm consumers. Some securities regulators use software to do preliminary analysis of documents such as emails received from market participants and newsletters sent by market participants to investors. The US' CFPB uses an online Consumer Complaints Database (CCD)

⁹ The European Central Bank (ECB) is looking into implementing the input-approach within the European Reporting Framework (ERF). The ERF's Expert Group on Statistical and Banking Data Dictionary has developed a Banking Data Dictionary with a harmonized model for input data and rules for the transformation of the input data to reporting data. See https://www.ecb.europa.eu/stats/ecb_statistics/co-operation-and-standards/reporting/html/index.en.html

¹⁰ www.mas.gov.sg/News-and-Publications/Speeches-and-Monetary-Policy-Statements/Speeches/2017/Singapore-FinTech-Journey-2.aspx

¹¹ <https://www.fca.org.uk/publications/calls-input/call-input-smarter-regulatory-reporting>

¹² ECB (2017).

for both structured and unstructured real-time data on consumer complaints, which is automatically processed by analytics software.¹³

Data Analytics

Improving data collection is only a means to an end. The most compelling use of SupTech is to enhance data analytics for both micro-and macro-prudential purposes. SupTech solutions have two main benefits for data analytics: alleviating the burden of data crunching through automation, and opening up avenues for newer, richer and more complex analyses that would be impractical using conventional analytical tools and data. ML and big data are combined, for instance, to improve stress testing and crisis simulation exercises, accuracy of forecasts, credit risk monitoring and modeling, real-time detection of suspicious and anomalous trades and transactions.¹⁴ Moreover, SupTech is improving data visualization with interactive dashboards and charts, so that supervisors can extract better and timely insights.

Some examples include:

- A deep learning tool deployed in early 2018 by Mexico's CONSAR, the national pensions regulator. Previously, data analysts had to import reported granular data to an Excel-based risk dashboard that flagged potentially risky and anti-competitive behavior and frauds. The labor-intensive Excel system was replaced by a deep learning tool trained with past data to incorporate and enhance the existing risk model. In addition, the new tool overlays other data, such as agent data, transaction data (e.g., geolocation and time stamp), transaction device data, and customer biometric data (e.g., voice and fingerprints) to enhance the analysis. The risk dashboard and warning system are now web-based and constantly updated through automated extraction of real-time data from a central database held at Procesar, a company owned by the pension administrators.
- The MAS is testing a tool to analyze high volumes of suspicious transaction reports and substitute labor-intensive and time-consuming procedures to filter the reports that demand further investigation (by humans) from false positives.
- The Mexican banking commission (CNBV) is piloting a system to automate the collection of suspicious transaction report statistics and granular transaction data in a central database at CNBV via an API. The system uses algorithmic models to produce a risk dashboard, data visualization, alerts, reports, and outlier flags.
- The US Federal Reserve uses big data in its Comprehensive Capital Analysis and Review (CCAR) stress-testing process based on granular loan data used to project losses in each retail product.¹⁵
- The US' CFPB uses structured and unstructured complaints data to create company profiles, monitor activities, identify emerging risks, and inform its risk-based methodology to prioritize FSPs. The system provides trend analysis, early warning tools, and scorecards. An algorithm-based tool named 'Spikes and Trends' flags short-, medium-, and long-term changes in complaints behavior.

Most traditional data analytics software focuses on one or a couple of structured data types and misses potential correlations across different data types and sources. The combined use of technologies such as ML, deep learning, Optical Character Recognition (OCR), NLP, and big data analytics allows supervisors to integrate the analysis of multiple data sources and formats, which was impossible with traditional software. Based on multiple data streams, SupTech can provide interactive dashboards, powerful

¹³ World Bank (forthcoming).

¹⁴ Other examples in macro-prudential oversight are found in FSB (2017a) and BIS (2015).

¹⁵ Jagtiani et al (undated).

visualization tools, network graphs (e.g., to identify concentration risks), real-time transaction monitoring, and topic modeling, for example for misconduct risk.

The Financial Stability Board (FSB) mentions the potential of allying ML with NLP to identify patterns in the combination of trading data with behavioral data (e.g., communications among traders/employees).¹⁶ By finding patterns and relationships across a wider range of variables than what traditional analytics can handle, SupTech allows more complex scenario-building and identification of deviations that merit further investigation. SupTech is also much better than traditional analytics in conducting network analysis and misconduct monitoring in large sets of trading and transaction data. Moreover, it can improve macro-economic analysis such as by using housing prices and gauging consumer sentiment from big data.

Securities supervisors have been early adopters of such technologies:

- The MAS is implementing a ML tool to detect trade syndicates and price manipulation in the stock market based on pattern recognition in real-time data. The tool can detect, for instance, accounts acting in concert to manipulate price. The MAS is also using NLP to do preliminary analysis of prospectuses, including of initial coin offerings (ICOs).
- The US SEC has a specialized data analytics team that uses ML and NLP and other technologies to analyze unstructured data (e.g., narrative disclosures on the web, news articles, regulatory filings in text form) to assess risk probabilities. It is also exploring deep learning for topic modeling.
- The US' Financial Industry Regulatory Authority (FinRA) uses ML to improve red flags that could signify rules violations.¹⁷
- The London Stock Exchange combines IBM's ML service and a cybersecurity solution to improve its market surveillance.¹⁸
- The Australian Securities and Investments Commission (ASIC) uses SupTech to analyze real-time and historical trade data, create alerts, and assess complex thematic risks.
- Numerous securities regulators are exploring DLT for real-time monitoring and auditing of trades and agreements.¹⁹

Automated pattern recognition through the combination of data types and technologies is also the basis of SupTech for AML/CFT and fraud monitoring. Supervisors are now capable of digesting high volumes of transaction data and incorporating other relevant variables such as transaction geolocation and time stamp, mobile phone usage data, and consumer sentiment from social media. Central banks are exploring overlaying social media with traditional statistics to assess risk of bank runs.²⁰ Bank of Italy has been researching the use of Twitter and data from real estate websites to improve macro-prudential analyses.²¹

Finally, SupTech can integrate existing databases at supervisory agencies to maximize data utilization and reduce redundancies. The MAS is using an API to unify access to all databases and provide a "Discovery Platform" that centralizes all data analytics.

¹⁶ FSB (2017a).

¹⁷ Jagtiani et al (undated).

¹⁸ Idem.

¹⁹ ECB (2017).

²⁰ BIS (2015).

²¹ Banca d'Italia (2018).

Supervisory Processes

SupTech can be used to digitize, automate, streamline or transform operational and administrative procedures, increasing their level of standardization and efficiency. The benefits could range from improving general performance (e.g., reducing response time to requests and applications), reducing costs of archiving, and increasing availability of digital data that can be used by new data analytics software, as well as for public dissemination. These outcomes could liberate brainpower for complex activities that require human judgment, including negotiations, managing relationships and decision-making. They could also increase transparency of supervisory processes.

Numerous supervisors, in particular in developed economies, for years have used specialized software for managing processes such as licensing and internal approvals, annual planning and task management, as well as to create digital documents, such as inspection reports. Digitization is particularly important when dealing with a large number of supervised FSPs. For instance, to monitor money laundering and terrorism financing risks in nonbank FSPs, the Brazil central bank developed a web-based system to collect structured and unstructured data, remotely interact with FSPs, and automate the generation of documents such as inspection reports and letters.²²

Automating tasks and digitizing documents do not automatically produce data that is readily usable by analytics software. However, SupTech solutions can integrate such data into the data streams that feed new types of analytics software. The tool used by the US CFPB is an example of how the automation of a process (handling consumer complaints) generates data that is immediately consumed by analytics software. Other examples include the pilot run by the Philippines central bank for a chatbot to collect, answer and process consumer complaints, using ML. The data generated by the chatbot will feed data analytics tools. Similarly, the UK's FCA is piloting a chatbot powered by ML to automate interactions with FSPs and rationalize queries about regulations and other interactions.

Supervisors disseminate reports, datasets and other information to the public, and much of the underlying work relies on manual processes for data extraction, cleaning and layout. At least part could be automated with SupTech. For example, the complaints database used by the US CFPB automatically generates different public versions of the data for public dissemination, including through the CFPB Open Data API.²³ The US SEC uses SupTech to create and disseminate public datasets based on reported data and documentation. To facilitate the work, the SEC has standardized most of text-based filings into formats that can be easily used by analytics tools. The “Nigeria Data Stack”, being built by the Nigerian central bank, will make granular payments data accessible in real-time to the general public through an API, and overlay other data such as demographics statistics.²⁴

Efforts to create global data standards and data identifiers, such as those by the FSB, are enablers of SupTech focused on data sharing. Emerging prototypes for cross-border data sharing use DLT to provide a secure environment and improve coordination, for instance, to spot and act upon signs of stress in financial markets and facilitate crisis management. In the context of FinTech, such efforts may help deal with FSPs offering services that defy supervisory jurisdictions, including cryptocurrency trading and ICOs.

Risks and Challenges in Adopting SupTech

There are risks and challenges in adopting SupTech. For a start, algorithms can fail. False positives, false negatives, and spurious correlations from algorithms can decrease the effectiveness of risk-based

²² World Bank (forthcoming).

²³ <https://dev.socrata.com/foundry/data.consumerfinance.gov/jhzv-w97w>

²⁴ <http://bfa-global.com/projects/payments-and-transactions-data-stack-in-nigeria/>

methodologies and tarnish the supervisor's reputation. The complexity and opacity of technologies such as deep learning can lead supervisors to lose control of automated processes, impacting their accountability and transparency obligations. Excessive reliance on data analytics could hamper the supervisor's ability to spot weaknesses that might not be visible in the data or measured quantitatively, and lead to a situation where supervision is insensitive to qualitative factors not captured by SupTech.

Some risks are similar to the risks faced by FSPs when they use similar technologies. For example, heightened cyber-security and third-party risks when data travels over the Internet or is handled by outsourced parties such as cloud providers. Shifting completely or partially to digital data and algorithm-based procedures exposes supervisors to greater potential impact of technology, electricity, telecommunication or other operational disruption, compared to paper-based tools and procedures. Cyber-attacks are one of the most significant threats in a highly automated environment and could lead to data loss and sequestration of computers. Third-party risks are significant, including the exposure of supervisory agencies and FSPs to the same few vendors (e.g., cloud or ML providers).

There is the risk of incorporating and reinforcing human biases in algorithm models, the risk of not being able to explain the outcomes of ML, particularly deep learning (the black-box issue),²⁵ and the risk of using an algorithm that does not fit the supervisor's needs or is flawed from the outset. These risks are heightened if the supervisory agency lacks expertise, experience with, or understanding of SupTech solutions (including their limitations) or lacks clear objectives for the SupTech implementation.

In general, providing quality digital data to SupTech, such as to build ML applications, can be a challenge. Datasets to train algorithms can be incomplete, and many supervisors face general data quality issues, even if only standardized regulatory reports are used. Also, big data, in particular social media, is often unreliable and low in quality and can be manipulated by third parties (including unethical FSPs willing to "game" the supervisor's algorithms). The use of social media may also raise data privacy issues depending on the country's legal framework.

Another challenge for the adoption of SupTech is the persistence of manual operational processes at FSPs. The more digitized their processes and documentation, the more digital data can be used in SupTech. Mexico's CONSAR, in order to create real-time digital offsite supervision, imposed, by regulation, a shift of the whole industry away from paper processes (e.g., forms used to transfer customers from one pension administrator to another). Supervisors adopting SupTech may need to assess whether they will face hurdles because of low level of digitization across FSPs.

Arduous integration with legacy systems could be another challenge. Legacy systems and data formats may be incompatible with SupTech, making it difficult for historical data to be fully utilized along with new data. Supervisors may also lack storage capacity and computer processing power to use SupTech such as big data analytics. Moreover, legacy systems at FSPs, as well as data and organizational silos, may frustrate SupTech focused on improving regulatory reporting.

SupTech is still a developing industry and there are no off-the-shelf solutions to meet all supervisory needs. Most often, solutions need to combine internal and external software development. Many supervisors may lack software engineers, and adequate technology, data analytics and business expertise to, for instance, handle high volumes of granular data and to propose, choose, develop, implement, and maintain SupTech. Even if it exists at a point in time, expertise can be difficult to retain, particularly in areas of high demand, such as ML. Lack of expertise also limits the supervisor's ability to compare SupTech offers. In fact, there may be only a few options of SupTech providers in a country. The options can be further reduced by budget constraints and burdensome procurement rules at supervisory agencies.

²⁵ The MAS is working on the development of explainable ML for supervisory purposes to address this issue.

Finally, even if none of the above was a problem, the combination of resistance to change, internal bureaucracy, organizational silos, and politics can get in the way of SupTech projects. Resistance comes in many forms, such as fear to be substituted by machines and codes, and the uncertainties (e.g., the impact on individual performance evaluations) of changing methodologies and procedures to which supervisors are already used.

Conclusion

SupTech has the potential to enhance human judgment and decision-making and mitigate common supervisory challenges. It can improve quality, timeliness and relevance of supervisory data, digitize and automate processes and expand analytical capacity, helping supervisors better deal with the new financial services landscape, expanded regulatory perimeters, and products and business models that challenge traditional practices. Sophisticated data analytics can overlay highly granular reported data with additional data streams and formats to reveal previously uncovered patterns, connections and networks that are relevant for macro- and micro-prudential supervision, market monitoring and enforcement across all financial sectors. With SupTech, supervisors can finally become forward-looking, data-driven, real-time supervisors. By liberating brainpower from time-consuming, process-oriented tasks, SupTech could lead to more effective risk-based supervision.

However, SupTech projects could require shifts in mindset, as well as organizational reforms and, depending on the state of current resources at supervisory agencies, significant investment in software and/or hardware, and specialized expertise in areas such as data science, software engineering and ML.

Overall, there could be many challenges to implementing SupTech, such as:

- Lack of clearly articulated objectives and goals for SupTech;
- Lack of inter-departmental and inter-agency coordination for integrating needs that could be addressed by a single SupTech solution;
- Limited availability of customizable SupTech solutions to fit multiple supervisory needs;
- Difficulty in finding and retaining in-house expertise;
- Insufficient budget or administrative constraints to procure SupTech projects;
- Persistence of manual processes and documentation at FSPs;
- Difficulty in integrating SupTech solutions with legacy systems and data;
- Existence of data silos at supervisory agencies and FSPs.

There are also risks related to the technology and the data used in SupTech, which could impact the supervisor's effectiveness and reputation, including:

- Risk of false alerts or erroneous outputs from ill-designed algorithms;
- Risk of using opaque ML tools whose output cannot be explained (black box);
- Risk of relying excessively on data analytics and losing sight of qualitative aspects not captured by SupTech;
- Heightened cyber-security and data security risks;
- Enhanced third-party risks (e.g., cloud computing and algorithm providers);
- Heightened operational risk, more broadly (e.g., power or communication disruptions);
- Risk of incorporating and reinforcing human biases in SupTech algorithms;
- Data privacy risks in using alternative data such as social media;
- Unreliability and low quality of certain big data types, such as social media;
- The potential for third-party manipulation of big data used as input for SupTech, such as by FSPs trying to "game" the supervisor's algorithms.

To make the most of SupTech and address its challenges and risks, supervisors may consider creating a SupTech strategy to align investment and implementation to supervisory goals and approaches, covering, for instance:

- Clearly articulated supervisory and policy objectives that could be enhanced by SupTech;
- Priority areas for SupTech investment;
- Specific SupTech solutions to be sought;
- Limitations of SupTech solutions;
- Internal processes for translating regulatory interpretation and policy choices into algorithms;
- Algorithm management and governance;
- Infrastructure and organizational arrangements needed, including computing and storage capacity and integration of data management and governance frameworks;
- Skillsets needed, and policy to acquire and retain expertise;
- Policy on vendor contracting, particularly to retain control of source codes;
- Robust cyber-resilience program;
- Effective operational risk management framework;
- Investments or reforms needed at FSPs, and strategy to acquire their buy-in.

SupTech attends to the desire to use data to its full potential, but no advanced analytics software can increase data quality. Addressing current data quality issues should be a priority for SupTech. Integrating reporting needs across departments, moving away from template-based approaches and increasing data granularity can require upheaval of data infrastructure and governance at supervisory agencies and FSPs, and elimination of existing data silos. Regulatory reporting is the core of supervisory analytics and improving it should be the starting point of SupTech to create a solid basis for any advanced analytics.

Implementing change and gearing mindset shifts to embrace SupTech for more effective supervision require strong leadership. Top-level commitment at supervisory agencies is essential, including to assign adequate budget and implement organizational reforms. It involves effective communication with all internal and external stakeholders to articulate the objectives, benefits and risks of SupTech and mitigate resistance to change. Piecemeal, unstructured SupTech projects lacking top leadership and a holistic strategy may, on the contrary, create apathy toward technology and delay the achievement of supervisory goals.²⁶

Finally, as experience with SupTech is still limited, supervisors would benefit from the formation of global and/or regional forums to facilitate their sharing of experience and knowledge, collaborate with academia, and to share or co-create algorithms and SupTech provider reference lists.

²⁶ See the TC Note series on Leading Organization Change at:
<http://res.torontocentre.org/guidedocs/TC%20Notes%20Compendium%20April%202018.pdf>

Annex: Glossary of Terms

Application Program Interface (API)	An API is a set of rules and specifications followed by software programs to communicate with each other, and an interface between different software programs that facilitates their interaction. ²⁷
Artificial intelligence (AI)	Artificial intelligence is the science of making computer programs perform tasks such as problem-solving, speech recognition, visual perception, decision-making and language translation. AI can ask questions, discover and test hypotheses, and make decisions automatically based on advanced analytics operating on extensive data sets. Machine learning (see below) is one subcategory of AI. ²⁸
Big data	Big data refers to the large volume of data that can be generated, analyzed and increasingly used by digital tools and information systems. This capability is driven by the increased availability of structured data, the ability to process unstructured data, increased data storage capabilities and advances in computing power. ²⁹
Big Data analytics	Big Data analytics focuses on, for instance, discovering patterns, correlations, and trends in the data, or customer preferences. It can be based on machine learning or other technologies.
Chatbot	A chatbot is a computer program designed to simulate conversation with human users and is widely used for online customer services at FSPs and beyond. More recent chatbots use ML for improved performance.
Cloud computing	Cloud computing refers to the use of an online network (“cloud”) of hosting processors to increase the scale and flexibility of computing capacity. This model enables convenient on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage facilities, applications and services) that can be rapidly released with minimal management effort or service provider interaction. ³⁰
Deep learning	Deep learning is a type of ML that mimics neuron networks of the human brain, having its roots in the 1950s. Recent breakthroughs led machines to effectively learn to recognize patterns in large datasets, including numbers, sounds, speech, text, images and other. The word “deep” relates to the numerous layers of virtual neurons used to process data. ³¹
Distributed ledger technology (DLT)	DLT such as blockchain are a means of recording information through a distributed ledger, i.e., a repeated digital copy of data at multiple locations. These technologies enable nodes in a network to securely propose, validate and record state changes (or updates) to a synchronized ledger that is distributed across the network’s nodes. ³²
FinTech	Technologically enabled financial innovation that could result in new business models, applications, processes or products with an associated material effect on financial markets and institutions and the provision of financial services. ³³

²⁷ BCBS (2018) p 42.

²⁸ BCBS (2018) p 42.

²⁹ BCBS (2018) p 42.

³⁰ BCBS (2028) p 42.

³¹ <https://www.technologyreview.com/s/513696/deep-learning>

³² BCBS (2018) p 42

³³ FSB (2017).

Information system infrastructure or IT infrastructure	The term “infrastructure” in an information technology (IT) context refers to a company’s collection of hardware, software, networks, data centers, facilities and related equipment used to develop, test, operate, monitor, manage and/or support information technology services. A related term is information system architecture, which refers to the conventions, rules, and standards used as technical framework to design or integrate various components of the information system infrastructure.
Internet of Things (IoT)	The internet of things (IoT) is the networking of physical devices, vehicles, buildings, and other items embedded with electronics, software, sensors, actuators, and network connectivity that enable these objects to (a) collect and exchange data and (b) send, receive, and execute commands. ³⁴
Machine learning	Machine learning is a sub-field of AI that focuses on giving computers the ability to learn without being specifically programmed for such through hand-inputted codes. It is focused on parsing out and learning from large amounts of data, in order to make a determination or prediction. Machine learning uses a variety of techniques, including neural networks and deep learning. In the past, AI tried to mimic human behavior through rules-based methods, i.e., logic-based algorithms. Today, machine learning is data-based, that is, computers analyze a large volume and variety of data to recognize patterns, which do not need to be intuitive or rational, or translated into programming codes. This type of machine learning is already having impact on financial services and financial supervision.
Natural language processing (NLP)	Regulations, like books and human speech, use natural language, that is, formats that humans can read and understand. NLP technology can transform natural languages into computer codes that can be understood by computers. This is the technology behind iPhone’s Siri and Amazon’s Alexa, for instance.
Network analysis	Network analysis is the mathematical analysis and mapping of complex relationships and flows between related activities, institutions, people, etc. Network graphs are an output of such analysis.
Optical Character Recognition (OCR)	OCR is the conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo (for example the text on signs and billboards in a landscape photo) or from subtitle text superimposed on an image (for example from a television broadcast).
RegTech	Regtech (regulatory technology) is defined as any range of fintech applications for regulatory reporting and compliance purposes by regulated financial institutions. This can also refer to firms that offer such applications. ³⁵
Taxonomy and data dictionary	Taxonomy is the classification of data (e.g., into categories and sub-categories) according to a pre-determined conceptual framework. Data dictionary encompasses the concepts, attributes and allowed formats of individual data points, and the relationship between data points.
Smart contracts	Smart contracts are programmable applications that, in financial transactions, can trigger financial flows or changes of ownership if specific events occur. Some

³⁴ BCBS (2018), p 43.

³⁵ BCBS (2018) p 43.

	smart contracts are able to self-verify their own conditions and self-execute by releasing payments and/or carrying out others' instructions. ³⁶
SupTech	Suptech (supervisory technology) is the use of technologically enabled innovation by supervisory authorities. ³⁷
Topic modeling	Topic modeling is a statistical model that uses ML and NLP to identify recurring topics or themes across a collection of documents. This technique is commonly used in text mining to discover semantic structures (e.g., sentences or combination of terms) within texts. Text mining is the process of deriving information from text, such as by identifying patterns.
Unstructured (vs. structured) data	Unstructured data is data in non-standardized formats that cannot be organized in traditional databases with searchable fields for easy sorting, extraction and analysis.

Key References

Basel Committee on Banking Supervision (BCBS). 2018. *Sound Practices: implications of fintech developments for banks and bank supervisors*. February. <https://www.bis.org/bcbs/publ/d431.htm>

Denise Dias. 2017. *FinTech, RegTech and SupTech: What They Mean for Financial Supervision*. Toronto Centre Note. <http://res.torontocentre.org/guidedocs/FinTech%20RegTech%20and%20SupTech%20-%20What%20They%20Mean%20for%20Financial%20Supervision.pdf>

Financial Stability Board. 2017. *Financial Stability Implications from FinTech: Supervisory and Regulatory Issues that Merit Authorities' Attention*, June. <http://www.fsb.org/wp-content/uploads/R270617.pdf>

Financial Stability Board. 2017a. *Artificial intelligence and machine learning in financial services: Market developments and financial stability implications*. November. <http://www.fsb.org/wp-content/uploads/P011117.pdf>

Additional Readings

Banca d'Italia. 2018. *Listening to the buzz: social media sentiment and retail depositors' trust*. By Matteo Accornero and Mirko Moscatelli. Working Papers 1165. February. https://ideas.repec.org/p/bdi/wptemi/td_1165_18.html

Bank of International Settlements (BIS). 2015. *Central Bank's use of and interest in 'big data'*, Irving Fisher Committee, 2015. <http://www.bis.org/ifc/publ/ifc-report-bigdata.pdf>

Bank of International Settlements (BIS). 2017. *The nature of evolving risks to financial stability*. Keynote address by Agustín Carstens, General Manager, at the 53rd SEACEN Governor's Conference/High-Level Seminar and 37th Meeting of the SEACEN Board of Governors. Bangkok, 15 December 2017. <https://www.bis.org/speeches/sp180214.pdf>

Basel Committee on Banking Supervision (BCBS). 2013a. *Principles for effective risk data aggregation and risk reporting*, January. <http://www.bis.org/publ/bcbs239.pdf>

³⁶ BCBS (2018) p 43.

³⁷ BCBS (2018) p 43.

European Banking Authority (EBA). 2018. *Recommendations on outsourcing to cloud service providers*. EBA/REC/2017/03 https://www.eba.europa.eu/documents/10180/2170125/Recommendations+on+Cloud+Outsourcing+%28EBA-Rec-2017-03%29_EN.pdf/e02bef01-3e00-4d81-b549-4981a8fb2f1e

European Central Bank (ECB). 2017. *The potential impact of DLTs on securities post-trading harmonization and on the wider EU financial market integration*. Advisory Group on Market Infrastructures for Securities and Collateral, September 2017. https://www.ecb.europa.eu/paym/intro/governance/shared/pdf/201709_dlt_impact_on_harmonisation_and_integration.pdf

Financial Conduct Authority (FCA). 2018. *Call for input: Using technology to achieve smarter regulatory reporting*. February. <https://www.fca.org.uk/news/press-releases/fca-launches-call-input-use-technology-achieve-smarter-regulatory-reporting>

Financial Conduct Authority (FCA). 2018. *Algorithm Trading Compliance in Wholesale Markets*. February. <https://www.fca.org.uk/publication/multi-firm-reviews/algorithmic-trading-compliance-wholesale-markets.pdf>

Institute of International Finance (IIF). 2016. *RegTech in financial services: technology solutions for compliance and reporting*, Institute of International Finance, March. https://www.iif.com/system/files/regtech_in_financial_services_solutions_for_compliance_and_reporting.pdf

Jagtiani, Julapa, Larry Wall and Todd Vermilyea. Undated. *The Roles of Big Data and Machine Learning in Banking Supervision*. Banking Perspectives. The Clearing House.org. <https://www.theclearinghouse.org/banking-perspectives/2018/2018-q1-banking-perspectives/articles/big-data-ml-bank-supervision>

Micheler, Eva and Whaley, Anna. 2018. *Regulatory Technology*. April. <https://ssrn.com/abstract=3164258>

Packin, Nizan Geslevich. 2018. *Regtech, compliance and technology judgment rule*. Chicago Kent Law Review, Volume 93 Issue 1 FinTech's Promises and Perils, Article 7. December. <https://scholarship.kentlaw.iit.edu/cgi/viewcontent.cgi?article=4198&context=cklawreview>

Piechocki, M., Dabringhausen, T. 2015. *Reforming Regulatory Reporting: From Templates to Cubes*. BearingPoint. Paper prepared for the Irving Fischer Committee on Central Bank Statistics, workshop on "Combining micro and macro financial statistical data for financial stability analysis: Experiences, opportunities and challenges", Warsaw, Poland, 14-15, December. <http://www.bis.org/ifc/publ/ifcb41o.pdf>

Securities and Exchange Commission (SEC). 2017. *The Role of Big Data, Machine Learning, and AI in Assessing Risks: a Regulatory Perspective*, speech by Scott W. Bauguess, Acting Director and Acting Chief Economist, DERA, US Securities and Exchange Commission, Champagne Keynote Address at OpRisk North America 2017, New York, New York, June 21 2017. <https://www.sec.gov/news/speech/bauguess-big-data-ai>

Wall, Larry D. Undated. *Some Regulatory Implications of Machine Learning*. Federal Reserve Bank of Atlanta. <https://www.philadelphiafed.org/-/media/bank-resources/supervision-and-regulation/events/2017/fintech/resources/some-financial-regulatory-implications-of-artificial-intelligence.pdf?la=en>

World Bank (upcoming). *From Spreadsheets to Suptech: Technology Solutions for Market Conduct Supervision*. Discussion Note. 2018.

Wright, Paul. 2018. *Risk-Based Supervision*. Toronto Centre Note. March.
<https://res.torontocentre.org/guidedocs/Risk-Based%20Supervision.pdf>