Financial Stability Institute

FSI Insights
on policy implementation
No 19

The suptech generations

by Simone di Castri, Stefan Hohl, Arend Kulenkampff and Jermy Prenio

October 2019

JEL classification: C45, C88, C89, G20, G38, O31, O32

Keywords: Suptech, data collection, data analytics, innovation, big data, AI, artificial intelligence

BANK FOR INTERNATIONAL SETTLEMENTS

FSI Insights are written by members of the Financial Stability Institute (FSI) of the Bank for International Settlements (BIS), often in collaboration with staff from supervisory agencies and central banks. The papers aim to contribute to international discussions on a range of contemporary regulatory and supervisory policy issues and implementation challenges faced by financial sector authorities. The views expressed in them are solely those of the authors and do not necessarily reflect those of the BIS or the Basel-based committees.

Authorised by the Chairman of the FSI, Fernando Restoy.

This publication is available on the BIS website (www.bis.org). To contact the BIS Media and Public Relations team, please email press@bis.org. You can sign up for email alerts at www.bis.org/emailalerts.htm.

ISSN 2522-2481 (print)
ISBN 978-92-9259-303-2 (print)

ISSN 2522-249X (online)
ISBN 978-92-9259-302-5 (online)

Contents

The suptech generations ... 1

Executive summary ... 1

Section 1 – Introduction ... 3

Section 2 – Suptech redefined ... 4

Section 3 – Generations of technology used by financial authorities ... 6

Section 4 – Suptech strategies ... 7

Section 5 – Suptech use cases ... 9

- Misconduct analysis and market surveillance ... 11
- Reporting and virtual assistance ... 12
- Data management ... 12
- Micro- and macroprudential supervision ... 13

Section 6 – Suptech development and deployment ... 14

Section 7 – Concluding remarks ... 15

References ... 18

Glossary ... 19

The suptech generations[1]

Executive summary

Suptech initiatives have gained momentum but it remains unclear exactly what falls within its scope. The term is defined by Broeders and Prenio (2018) as the use of innovative technology by supervisory agencies to support supervision. Since that publication, an increasing number of supervisory authorities are beginning to explore suptech applications in different areas of supervision. In addition, other non-supervisory financial authorities (eg financial intelligence units) have also used or experimented with innovative technologies to support their work. However, the Broeders and Prenio definition only refers to "innovative technology" without defining it. Consequently, the differing stages of technological progress across financial authorities have led to differences in the way "suptech" has been interpreted.

This paper examines these developments by analysing suptech initiatives in 39 financial authorities globally. Most of these financial authorities responded to a survey on suptech strategies and use cases conducted jointly by the BIS's Financial Stability Institute (FSI) and the Regtech for Regulators Accelerator (R^2A).[2] The survey responses were supplemented by information from the two previous FSI Insights papers on suptech,[3] as well as by information from the online tracker developed by R^2A.[4]

Suptech is more broadly defined as the use of innovative technology by financial authorities to support their work. For the purposes of this paper, the term "innovative technology" refers to the application of big data or artificial intelligence (AI) to tools used by financial authorities. This new definition clarifies the scope in terms of suptech users (ie including non-supervisory financial authorities such as financial intelligence units)[5] as well as the types of technology used (big data or AI).

Not all initiatives examined for this paper meet the above definition of suptech and could be considered more appropriately as belonging to different "generations" of technology. The first generation involves data management workflows with intensive manual input, and mostly delivering descriptive analytics. The second generation digitises and automates certain manual processes in the data pipeline. The third generation covers big data architecture whereas the fourth generation involves the addition of AI as the defining characteristic. Suptech straddles the third and fourth generations. In particular, third-generation data collection solutions and fourth-generation data analytics solutions are considered suptech for the purposes of this paper.

1 This paper was authored by experts/members of the Financial Stability Institute of the Bank for International Settlements in collaboration with members/experts of the BFA's RegTech for Regulators Accelerator (R^2A). Stefan Hohl (stefan.hohl@bis.org) and Jermy Prenio (jermy.prenio@bis.org), Bank for International Settlements; Simone di Castri (sdicastri@bfaglobal.com) and Arend Kulenkampff (akulenkampff@bfaglobal.com), BFA's RegTech for Regulators Accelerator (R^2A).

The authors are grateful to contacts at the covered financial authorities; to the participants of the second FSI meeting on the use of innovative technology in financial supervision held in Basel on 5–6 June 2019 for the insightful discussions; and to Patrizia Baudino, Christopher Calabia, Jon Frost and Leonardo Gambacorta, for helpful comments. Cissy Mak provided valuable administrative support with this paper.

2 Regtech for Regulators Accelerator is a non-profit, donor-funded accelerator programme administered by BFA Global that aims to help financial authorities in emerging markets and developing economies explore specific suptech solutions by providing support in building prototypes.

3 Broeders and Prenio (2018) and Coelho et al (2019).

4 See http://vendors.r2accelerator.org/?v=tracker.

5 However, the paper does not include authorities in charge of monetary or macroeconomic policies that may also be using similar tools. See eg Tissot et al (2015).

While suptech will help authorities to become more data-driven, the technologies that authorities use should be appropriate to the size, complexity and development of the sectors they oversee. For example, investments in big data architecture and AI tools might not be appropriate for an authority in a low-income jurisdiction that supervises only a handful of financial institutions providing basic financial products and services. Moreover, authorities should also be aware of the issues and challenges associated with suptech. Broeders and Prenio (2018) outlined some of these issues and challenges. In particular, the lack of transparency in some of the suptech data analytics solutions is a critical issue. This underscores the continued need for human intervention in the form of supervisory expertise to further investigate the results of analyses and when deciding on a course of action.

Almost half of the financial authorities covered have explicit suptech strategies or are in the process of developing them. The approaches taken vary. Some specify suptech roadmaps with a deliberate path towards adopting big data and AI processes and systems. Others have developed suptech applications as part of an institution-wide digital transformation and data-driven innovation programme. This is broadly aimed at moving the whole institution to more automated and digitised processes as well as adopting advanced data collection and data analytics tools. A well defined strategy can help authorities optimise the potential benefits of suptech for their organisation. But for authorities who want to explore specific suptech tools first before committing substantial resources, there are helpful institutionalised or one-off methodologies such as innovation labs, accelerators or tech sprints. These methodologies may also be embedded into authorities' existing or future suptech strategies.

The suptech use cases observed cluster mainly around misconduct analysis, reporting and data management. Conduct supervision and the work of financial intelligence units look at huge amounts of unstructured data. As such, they can particularly benefit from the development of big data architecture and AI tools. Virtual assistance, microprudential, macroprudential, and market surveillance make up a smaller share of the sample set.

Suptech solutions have emerged only recently, are mostly experimental in nature and are being developed within financial authorities. The majority of suptech initiatives reported are still in either the experimental or development stages, with less than a third operational. Most of the suptech initiatives covered in the paper are being developed internally or jointly with external developers or other organisations such as universities. Suptech initiatives developed solely by external vendors account for only a quarter of all reported initiatives. This could be due to the experimental nature of these initiatives, among other reasons. Consequently, many initiatives may lack clearly defined functional requirements or technical specifications with which to engage external parties. This suggests the importance of strategic partnerships between financial authorities, other governmental agencies, and academia as well as research organisations to help overcome the challenges associated with the experimental nature of these initiatives.

Further international coordination and collaboration may help to accelerate suptech development. Global standard-setting bodies and international organisations provide platforms for authorities to exchange information on their suptech initiatives. These international platforms could also be used potentially to collaborate on the development of suptech solutions that may be useful to a number of authorities or to address related cross-border issues affecting the development of suptech (eg data localisation). A good example is the BIS Innovation Hub that aims to foster international collaboration on innovative financial technology within the central banking community. Such platforms will help authorities to benefit from peer learning, including from the experience of different types of authority (central banks, prudential regulators, conduct regulators etc), especially given the dearth of specialist providers. They should also reduce the need for individual authorities to independently work on similar solutions, thus increasing efficiency. In addition, given the inherently small market for suptech solutions, which limits business opportunities for private providers, accelerators set up or funded by international organisations could play an important role in helping authorities explore specific suptech tools.

Section 1 – Introduction

1. **Broeders and Prenio (2018) defined supervisory technology (suptech) as the use of innovative technology by supervisory agencies to support supervision**. The paper outlined the suptech applications used by some supervisory authorities, and mapped these applications onto the different areas of supervision. It also examined the practical experience of "early movers" who recognised the potential of suptech to turn risk and compliance monitoring from a backward-looking into a predictive and proactive process.

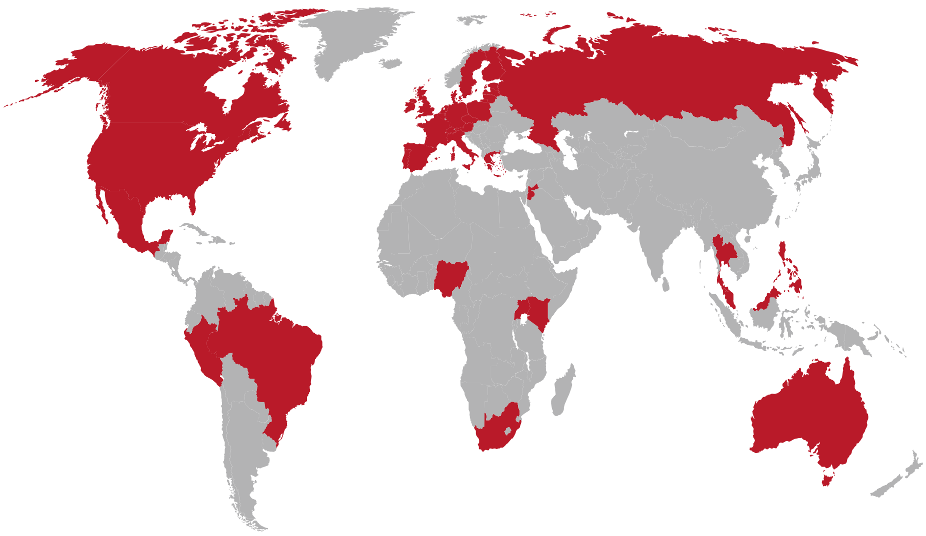
2. **Since then, suptech initiatives have gained momentum around the world.** An increasing number of supervisory authorities are beginning to explore suptech applications in different areas of supervision. In addition, other non-supervisory financial authorities (eg financial intelligence units) have also used or experimented with innovative technologies to support their work. There now seems to be wider recognition of the potential of suptech to transform data processes in financial authorities, which would in turn enable better and more timely decisions and actions.

3. **However, it remains unclear as to what falls under the umbrella of "suptech".** The above definition only refers to "innovative technology" without defining it. Consequently, the differing levels of technological progress within financial authorities have led to differences in the way "suptech" has been interpreted.

4. **To examine these developments, this paper looks at suptech initiatives in 39 financial authorities from 31 jurisdictions.** Graph 1 shows the financial authorities covered in this paper. Most responded to a survey on suptech strategies and use cases conducted jointly by the FSI and the R^2A[6] during the second quarter of 2019. The survey responses were supplemented by information from the two previous FSI Insights papers on suptech,[7] as well as by information from the online tracker developed by R^2A.[8]

Graph 1

Country	Authority
Australia	APRA, ASIC, AUSTRAC
Austria	OeNB
Brazil	BCB
Brunei	AMBD
Canada	FINTRAC
Czech Republic	CNB
Denmark	DFSA
EU	EIOPA
Germany	BBk
Greece	BoG
Guernsey	GFSC
Hong Kong	HKMA
Italy	BoI
Jordan	CBJ
Kenya	CBK
Lithuania	BoL
Malaysia	BNM
Mexico	BoM, CNBV, CONSAR
Netherlands	DNB
Nigeria	CBN
Peru	SBS
Philippines	BSP
Poland	KNF
Russia	CBR
Rwanda	BNR
Singapore	MAS
South Africa	SARB
Switzerland	FINMA
Thailand	BOT
United Kingdom	FCA
	BoE
United States	FRB, FRBNY, SEC, CFPB



6 Regtech for Regulators Accelerator is a non-profit, donor-funded accelerator programme administered by BFA Global that aims to help financial authorities in emerging markets and developing economies explore specific suptech solutions by providing support in building prototypes.

7 Broeders and Prenio (2018) and Coelho et al (2019).

8 See http://vendors.r2accelerator.org/?v=tracker.

5. **Based on this new set of information, this paper takes a fresh look at suptech and provides an update on developments in this field.** Specifically, Section 2 offers a new definition of suptech – on the one hand broadening it to include more authorities, while on the other hand narrowing it to specific technologies. Section 3 then traces the evolution or the different generations of technology used by financial authorities, which culminates in generations that can be considered as suptech. Section 4 describes the suptech strategies pursued by different authorities. Section 5 and 6 provide an overview of the different suptech use cases, and the manner of development and status of deployment, respectively. Section 7 concludes.

Section 2 – Suptech redefined

6. **This paper redefines suptech as the use of innovative technology by financial authorities to support their work**. For the purposes of this paper, the term "innovative technology" refers to the application of big data or artificial intelligence (AI) to tools used by financial authorities, while "financial authorities" refer to both supervisory and non-supervisory authorities.[9] This new definition clarifies the scope in terms of users of suptech (ie including non-supervisory financial authorities such as financial intelligence units) as well as the types of technology used (big data or AI).

7. **Big data encompasses technologies that significantly increase the volume, variety, velocity and validity of data under management — the so-called four Vs of big data.** Big data involves data sets that are orders of magnitude larger than can be accommodated by common spreadsheet applications.[10] The European Central Bank's (ECB's) AnaCredit project is one example where the boundaries of big data are already being pushed in the context of supervision.[11] This is compounded by many more varieties of data than were previously considered by authorities, including both structured tabular data as well as unstructured web content such as email, images and social media posts (eg "tweets"). The speed or velocity of big data measures not only the time between the generation of data and their collection, but also how quickly it is turned into reports and actions. Finally, validity speaks to the quality of data. To guard against the "garbage in, garbage out" problem, the data must be subjected to validation checks and other quality controls or else they cannot be trusted to deliver accurate and reliable information. Consistent metadata standards, such as the Statistical Data and Metadata eXchange (SDMX) reporting standards in the case of AnaCredit, are crucial in this regard.[12]

8. **A big data architecture comprises the processes and systems that enable and govern the collection, processing, storage, analysis and visualisation of data.** To qualify as such, the layers of the architecture must be internally coherent such that each can handle the speed, size and complexity of big data as defined above. Crucially, the architecture must have built-in quality assurance and security features to ensure the validity and integrity of the data from the point of collection to the point of consumption by end users. This end-to-end flow of data should be seamless, speedy and scalable, without bottlenecks, lags or size constraints.

9. **A number of big data tools can be leveraged to construct such an architecture.** Application programming interfaces (APIs) can ferry large volumes of data directly between databases without human

[9] However, the paper does not include authorities in charge of monetary or macroeconomic policies that may also be using similar tools. See for example Tissot et al (2015).

[10] Microsoft Excel spreadsheets, perhaps the most ubiquitous format in financial regulatory reporting, are limited to around 1 million rows and 16,000 columns. Similarly, workbooks are limited to roughly 2 gigabytes limit depending on the version.

[11] Cœuré (2017).

[12] European Central Bank (2017).

intervention, thereby overcoming the size limitations of file transfer via email or web portals as well as cutting down on time-consuming and error-prone manual submission. Similarly, robotic process automation (RPA), a form of business automation technology driven by robotics software ("bots"), can substitute for laborious validation and transformation of data, further reducing the scope for human error and speeding up data turnover times. Highly efficient and scalable storage solutions, particularly cloud-based computing, can accommodate big data at rest. Advanced document and data management systems such as "data lakes" can also handle unstructured data. Distributed ledger technology (DLT) allows for automatic validation through a consensus algorithm that replicates, shares and synchronises digital data across different locations. Finally, big data visualisation tools such as dynamic dashboards allow for seamless data interrogation with minimal latency, allowing humans (ie financial authorities) to quickly absorb and understand data.

Graph 2

10. **AI is defined by the Financial Stability Board as the theory and development of computer systems able to perform tasks that traditionally have required human intelligence.**[13] In practice, AI can help in the analysis of large and complex data that would otherwise be impossible for humans to perform. It encompasses machine learning (ML), natural language processing (NLP), visual analytics and all of their respective sub-branches. Financial authorities are now exploring or using AI applications, particularly to enable the integration and analysis of large volumes of information from disparate sources. Graph 2 sketches a non-exhaustive mapping of big data and AI technologies (including their definitions)

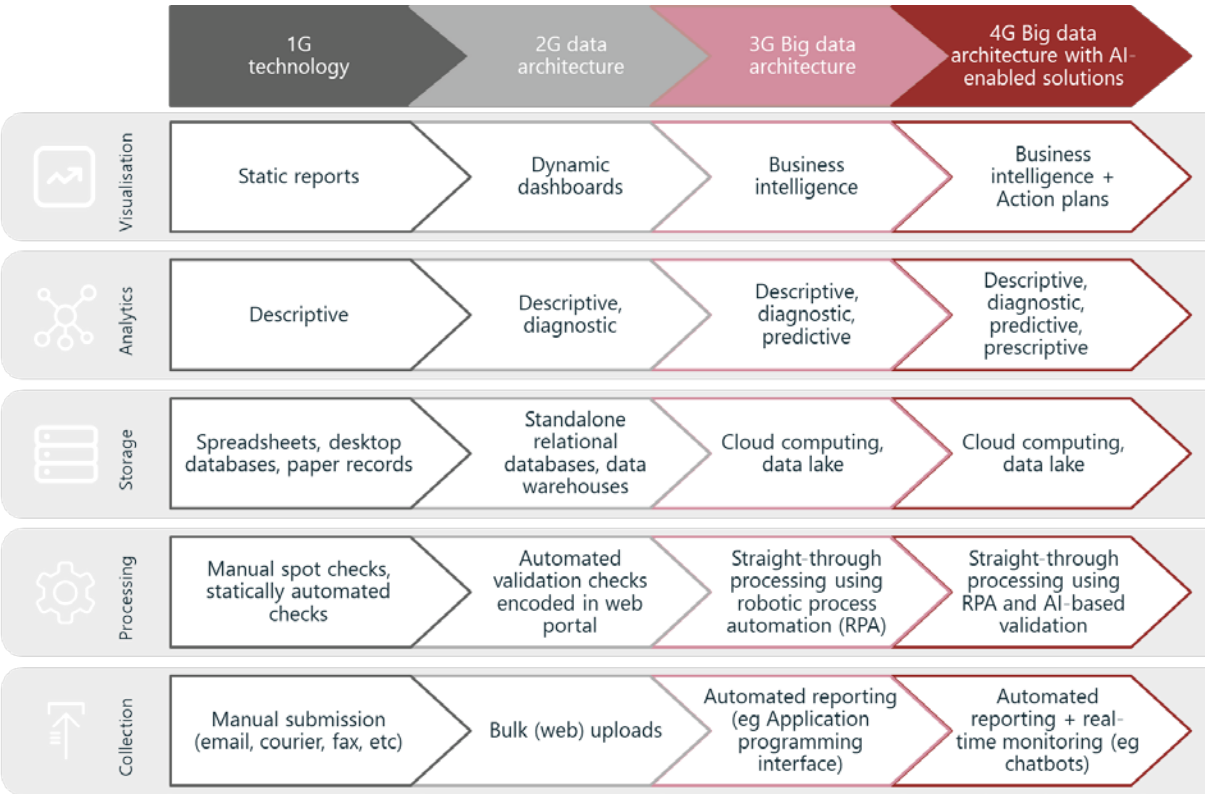
[13] Financial Stability Board (2017).

and their impact on data. As an example, a "data lake", which is a scalable storage solution for data of various formats, allows the storage of massive volumes and variety of data, thus it is mapped accordingly (ie to both "volume" and "variety").

Section 3 – Generations of technology used by financial authorities

11. **Not all initiatives examined for this paper meet the definition of suptech suggested above and some could be more appropriately considered as belonging to different "generations" of technology.** Some constitute IT infrastructure upgrades that entail no material increase in the four Vs of big data. Such early generations of data architecture support mostly descriptive and diagnostic analytics – respectively, "what happened and why did it happen?" – whereas big data and AI enable predictive and prescriptive analytics – "what will happen and what should I do about it?" Depending on the task, the former can still generate both sufficient information and significant efficiency gains while laying the groundwork for a big data architecture and increasing AI readiness. Graph 3 shows the evolution of technological progress in terms of generations of technology used by financial authorities. These are not necessarily discrete categories, but rather a continuum culminating in a big data architecture supporting advanced AI applications. Suptech straddles the third and fourth generations. In particular, third-generation data collection solutions and fourth-generation data analytics solutions are considered suptech for the purposes of this paper.

Graph 3



12. **The first generation involves data management workflows that are heavily manual and mostly descriptive analytics.** This has been the starting point for most financial authorities. It involves data collection in which reports are submitted either in paper form or via email, which imposes file size restrictions and introduces operational and security risks. Staff of financial authorities validate data manually (eg "spot checks" or statically automated checks using macros), and extraction, transformation

and loading (ETL) of data to prepare for analysis are also done manually. Storage is fragmented across disjointed spreadsheets or desktop databases, or in paper records. Data analysis is performed in relatively rigid and simplified spreadsheet models and visualisations are rendered in static reports that require manual updating. Because of data and infrastructure limitations, analytics tend to be descriptive in nature.

13. **The second generation covers the digitisation and automation of certain paper-based and manual processes in the data pipeline.** Typically, this involves web-based portals or bulk uploads (eg file transfer protocol) for the submission of regulatory returns coupled with automated validation checks built into the upload protocol. Database rationalisation and the automation of ETL processes to prepare data for analysis are other common features. Some amount of straight through processing allows for more dynamic data visualisation in business intelligence (BI) dashboards, while improved analytical processing allows for deeper diagnostic analysis (eg scorecards) as well as richer descriptive insights.

14. **The third generation covers big data architecture.** Such architectures are built with technology stacks that support data of higher granularity, diversity and frequency than could be accommodated previously. On the input end, data ingestion and consolidation are fully automated, for instance, using a combination of APIs and RPA. Data storage and computation are optimised for seamless and continuous data interrogation, which may entail the use of cloud storage and "data lakes". Larger data pools coupled with greater computing power enable more advanced statistical modelling, including predictive analytics (eg econometric forecasting).

15. **The fourth generation involves the addition of AI as the defining characteristic.** Generally, AI-enabled solutions or tools presuppose an underlying big data architecture since most AI models require large volumes of data and significant computing power for their results to be valid, meaningful and actionable. Hence, digital transformation and big datafication[14] can be considered enablers of AI. Furthermore, the fourth generation takes automation one step further by having "machines" drive parts of data management and analysis, as well as inform authorities' actions. The former might entail leveraging natural language processing to scrape data from the web or using ML to match and merge disparate data sets. The latter can take the form of recommendation engines that suggest courses of action or even chatbots that execute supervisory tasks previously performed by humans, such as responding to and resolving customer complaints.

Section 4 – Suptech strategies

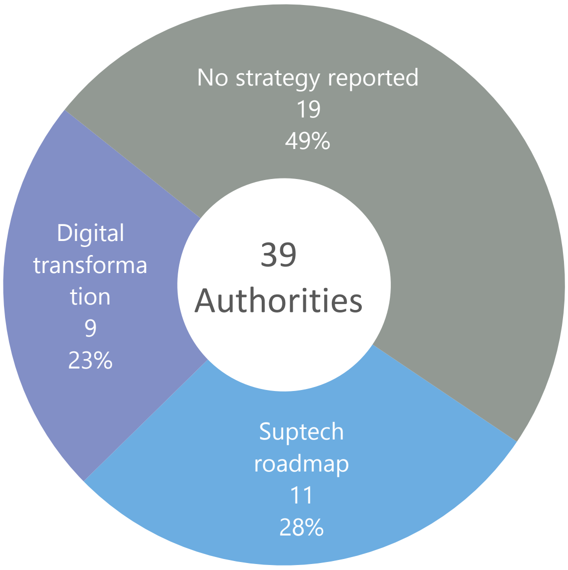
16. **About half of the financial authorities covered in the paper have explicit suptech strategies or are developing them, but approaches vary (Graph 4).** At least two types of approach can be discerned from this study and may be described as (i) specific suptech roadmaps, and (ii) institution-wide digital transformation/data-driven innovation (DT&DI) programmes. These approaches are not necessarily pursued in isolation. For instance, a DT&DI programme can subsume a suptech roadmap. A further distinction among approaches can be made between top-down and bottom-up strategies. In the former, use cases and scope of work have been mostly decided upon in advance. In the latter, solutions emerge by trial and error, diagnostic exercises, or as transplants from the private sector.

17. **Authorities without an explicit strategy tend to pursue suptech projects with an experimental focus or on an ad hoc basis.** These projects are chosen based on the particular needs of individual departments, or in an opportunistic fashion in response to a technological or market development. A menu of methodologies is available to financial authorities that may wish to explore and eventually implement suptech solutions. These include a more institutionalised approach such as

[14] The use of big data infrastructure, tools and processes to enable a data-driven organisation.

"innovation labs", or one-off programmes such as "accelerators" and "tech sprints", as explained below. These methodologies could also be used by authorities with explicit suptech strategies.

Graph 4



18. **Specific suptech roadmaps set out a deliberate path to adopting big data and AI processes and systems to support the work of financial authorities.** Generally, this involves (i) making a formal commitment to innovation and setting out a work programme; (ii) selecting and prioritising use cases; (iii) preparing the data architecture (including IT infrastructure upgrades); and (iv) building solutions by way of various development methodologies. This approach is followed by the Australian Securities and Investment Commission (ASIC), and the Monetary Authority of Singapore (MAS), among others. ASIC, for example, has mapped out a succession of initiatives, starting with the establishment of an innovation hub in 2015, followed by "trials", and finally developing and demonstrating tangible suptech tools.

19. **Institution-wide DT&DI programmes broadly aim at shifting to automated/digital processes and systems, and adopting advanced data analytics tools.** As the name suggests, institution-wide DT&DI programmes have a broader scope, of which suptech forms part. For instance, Deutsche Bundesbank (DB) is developing a bank-wide digitalisation strategy, to which the banking supervision department contributes by promoting the development of suptech. A similar approach is being pursued by the Bank of Thailand (BOT). On the other hand, the Australian Prudential Regulation Authority's (APRA's) Data Transformation Program (2017–20) seeks to change how APRA collects, stores, uses and innovates with data, in which context suptech solutions are being explored. The Hong Kong Monetary Authority (HKMA) has embarked on a digitalisation programme, with various of its units working on projects. In this context, the HKMA banking departments are considering developing a suptech roadmap. Other authorities with a similar arrangement are the Central Bank of Malaysia (BNM), and the Bangko Sentral ng Pilipinas (BSP), among others.

20. **Whether or not a financial authority has an explicit suptech strategy, innovation labs allow an innovation centre to be set up within the organisation.** Innovation labs, whether in the form of units or programmes, allow technical solutions to be tested for supervisory use cases in a secure development environment. Generally, they receive dedicated funding and technical assistance from departments across the organisation. The Central Bank of the Republic of Austria (OeNB) has a bank-wide innovation lab where project ideas related to new technologies are evaluated and initiated. The Financial Transactions and Reports Analysis Centre of Canada (FINTRAC) has a data exploitation laboratory aimed at exploring and implementing advanced data analytics solutions. The French Prudential Supervision and Resolution

Authority (ACPR), on the other hand, operates an "intrapreneurship" programme that aims at encouraging staff members to suggest or lead innovative projects that would improve the ACPR's tools and processes. Selected projects are supported in their design through an innovation methodology provided by the Bank of France's innovation centre, Le LAB.

21. **The use of "accelerators" is another way to explore suptech solutions.** Accelerators are either one-off or ongoing programmes that involve outside parties. They aim to develop explorative proofs of concept (POCs) or prototypes during a defined timeline. As an example, R^2A is a donor-funded programme that seeks to help financial authorities in capacity-constrained emerging market economies explore specific suptech solutions. It uses lean design and agile development methodologies together with open-source software to develop scalable prototypes at low cost. The designs of the prototypes are co-created by staff members of the relevant units within a financial authority, R^2A consultants and sometimes external vendors specialising in a specific use case. These prototypes then serve as blueprints for the development of a full-scale final product should the financial authority decide to go ahead. The Bank of England's accelerator had a similar objective of developing POCs, including for suptech use cases. The difference was that innovative firms play a lead role in design and development, rather than the relevant units within the bank.[15]

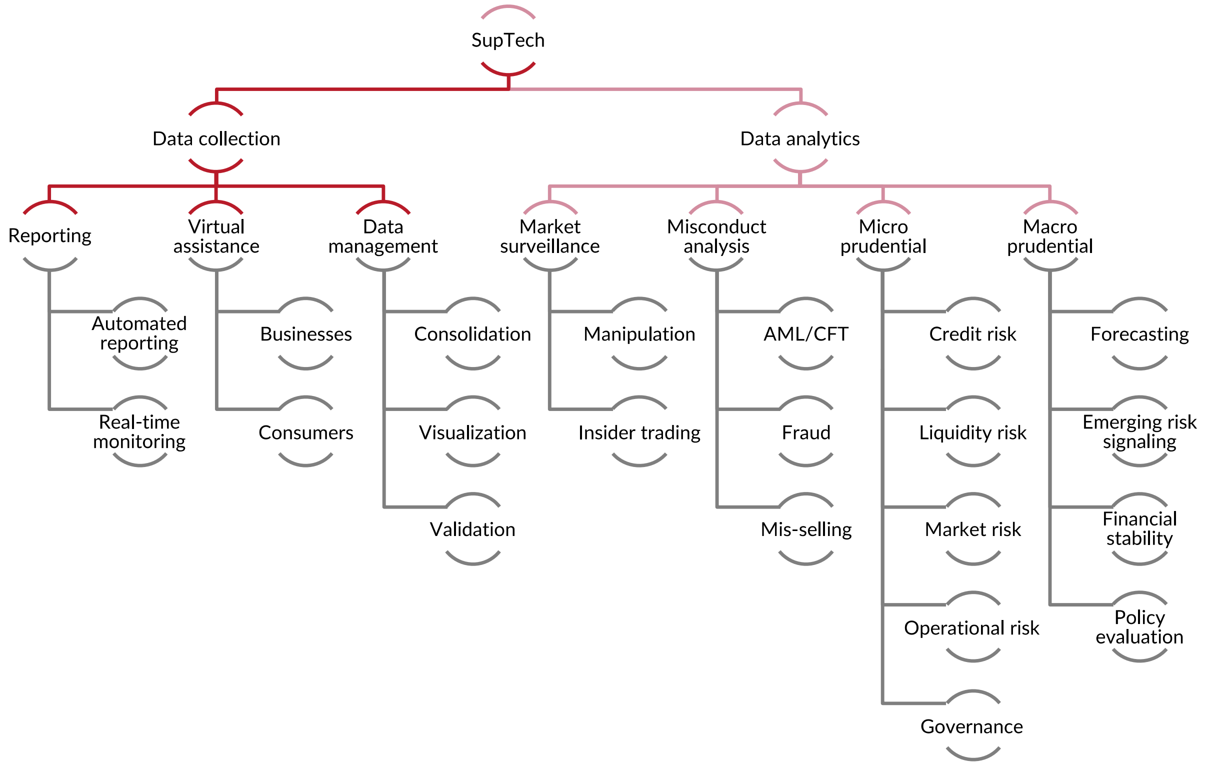
22. **Exploration of suptech solutions may also be aided by "tech sprints" (sometimes referred to as "hackathons" or "codeathons").** Like accelerators, tech sprints seek to develop POCs, albeit in a much shorter time frame. Tech sprints typically take the form of intensive workshops lasting two to five days that bring together in-house and industry experts around specific challenges or use cases. The United Kingdom's Financial Conduct Authority (FCA) has pioneered the use of tech sprints since 2016, including two on the topic of regulatory reporting. The 2017 tech sprint on digital regulatory reporting (DRR) has informed the FCA's initiative to re-engineer its data architecture and replace its existing web portal-based system, which has reached the end of its shelf life. To implement the tech sprints, the FCA receives technical assistance from consulting or technology firms that furnish facilities (eg a venue, high-speed internet), software (eg cloud computing and AI platforms) and synthetic data for the event.

Section 5 – Suptech use cases

23. **The universe of suptech use cases has remained broadly unchanged since the first comprehensive study of the field by Broeders and Prenio (2018).** The same but slightly expanded taxonomy was employed to categorise reported suptech applications according to the primary use (Graph 5). Thus, for data collection applications, end-to-end solutions fall under "reporting", whereas point solutions targeting specific sections of the data pipeline are categorised separately under the corresponding headers (eg validation solutions would fall under "data management"). Only use cases that fall under the definition of suptech provided in Section 2 are included here.

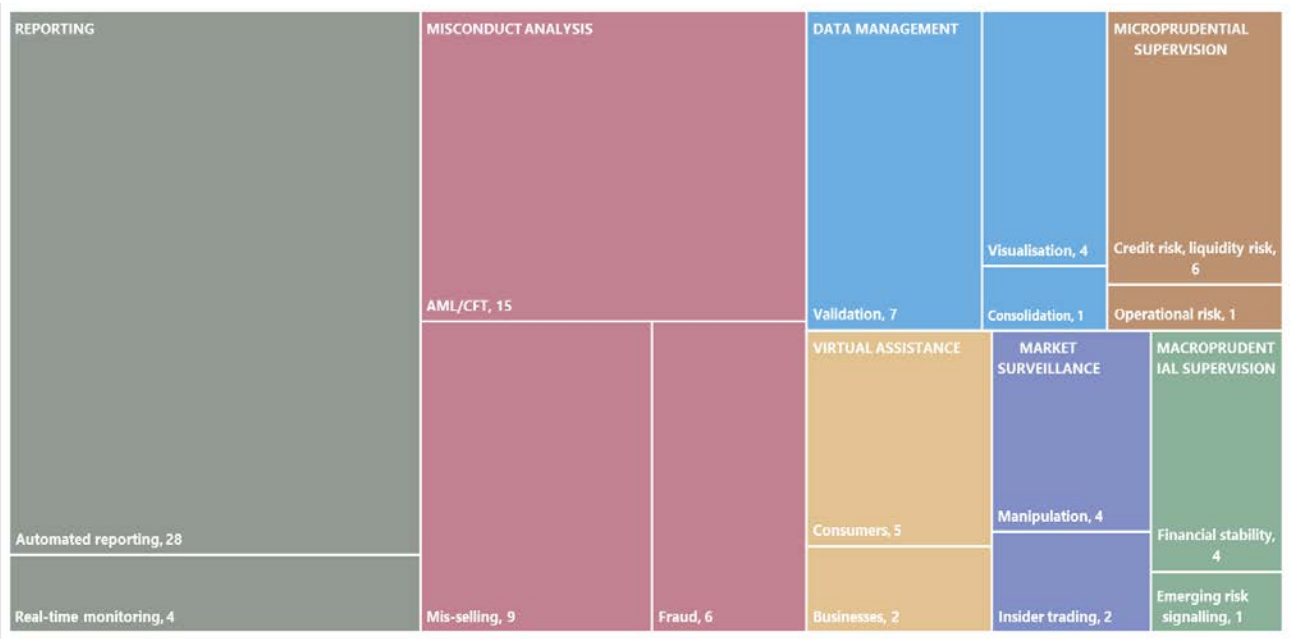
[15] See Bank of England's frequently asked questions on Bank of England accelerator.

Graph 5



24. **The suptech initiatives observed cluster mainly around misconduct analysis, reporting and data management.** Virtual assistance, microprudential, macroprudential and market surveillance make up a smaller share of the sample set (Graph 6). In total, we examined (99) use cases that fall under the new definition of suptech as described above.

Graph 6



Misconduct analysis and market surveillance

25. **New tools are required to combat the new forms of money laundering, terrorist financing, mis-selling and fraud made possible by digital technology.** This need likely explains the preponderance of misconduct use cases. It also reflects the fact that latest-generation suptech solutions lend themselves especially well to misconduct analysis. This is because these solutions generally deal with granular, time-sensitive and hard-to-parse unstructured data, for which big data and AI tools are well suited. In the field of anti-money laundering (AML) and combating the financing of terrorism (CFT), added impetus comes from the global trend of de-risking in correspondent banking relationships, which in part has been driven by weaknesses in AML/CFT regulatory and supervisory frameworks.[16] Coelho et al (2019) provide specific examples of suptech tools for AML/CFT purposes.

26. **Among the observed suptech use cases, ML is used to support anomaly detection, network analysis and risk scoring.** Mexico's National Banking and Securities Commission (CNBV) uses a combination of clustering, logistic regression, artificial neural networks and random forest models to detect suspicious patterns in granular transactional and account-level data. It is also experimenting with an NLP tool that flags selected names and companies from news related to money laundering schemes and links these to other data sources, both unstructured (social media) and structured (watchlists, accounts). The United Kingdom's FCA performs network analysis on orders and executions data to construct webs of market participants and identify collusive behaviour indicating insider trading, while the Netherlands Bank (DNB) employs a similar technique to link individuals sending funds to the same counterparties in high-risk jurisdictions along various routes. The Bank of Italy (BdI) is exploring a tool that would apply a combination of network analysis and self-organising map techniques to detect behaviours typical of fraud in the gold declaration database. Meanwhile, FINTRAC has a scoring model that uses geospatial and principal component analysis to rank supervised institutions according to their likelihood of non-compliance with AML rules.

27. **Text mining is a popular ML technique for AML and anti-fraud use cases where parsing voluminous textual data is involved.** The Systematic Integrity Risk Analysis (SIRA) reports submitted by Dutch financial institutions to the DNB are an example of documents that have been subjected to this form of automated textual analysis. ASIC, the Bank of Mexico (BoM), and the FCA also employ a combination of web-scraping and text mining to audit promotional materials, prospectuses or financial advice documents disseminated by financial institutions. MAS has developed an event impact analysis tool that scrapes web data from news sites and uses NLP to automatically detect so-called "hot events" and categorise them as "laundering" or not. MAS and the Czech National Bank (CNB) also use such a solution to check whether initial coin offerings (ICOs) are being marketed to residents in their jurisdictions. One variation explored by ASIC through a trial went one step further by testing voice analytics and voice-to-text (VA&VT) technology to monitor marketing calls for life insurance. The solution promises to analyse, flag and report calls that contain inappropriate sales tactics such as pressure selling or inaccurate information about products. Similarly, BoM developed a proprietary optical character recognition solution for converting images of trust contracts into text and a text mining tool that will help identify key words.

28. **From a market surveillance perspective, a big data architecture makes it possible for financial authorities to perform real-time market transaction monitoring.** Securities market supervisors, such as ASIC and SEC, transform enormous data sets into usable patterns for market surveillance purposes, including in the detection of potential insider trading and market manipulation. For example, ASIC's market analysis and intelligence (MAI) platform collects real-time data feeds from all Australian primary and secondary capital markets (ASX and Chi-X) for equity and equity derivatives products and transactions. The SEC uses analytics that harness the power of big data, primarily for text

[16] Financial Stability Board (2015).

arising from regulatory filings and tonal analysis, to drive its surveillance programmes. They also foster innovation in market risk assessment initiatives.[17]

Reporting and virtual assistance

29. **Reporting use cases can be distinguished between "automated reporting" solutions and "real-time monitoring".** The former draws on advances in big data technology to automate the collection of data from supervised entities, especially prudential returns. Real-time monitoring, by contrast, provides a "live" view of a given supervisory domain. Data collected from virtual assistance solutions help in this regard.

30. **Automation is increasingly achieved through APIs, but also other technology that enables machine-to-machine communication.[18]** APRA, BOT, BSP, CNBV, DB, the Central Bank of Brazil (BCB), the Central Bank of Jordan (CBJ), the European Insurance and Occupational Pensions Authority (EIOPA), the Polish Financial Supervision Authority (KNF) and the South African Reserve Bank (SARB) all have API projects under consideration or in the pipeline. Other financial authorities have expressed an intention to adopt them eventually. The OeNB has been pioneering the data cube approach, which the Bank of Greece (BoG) is also exploring.[19] The National Bank of Rwanda (BNR) continues to stand apart for its "data pull approach", which connects the entire financial sector to a centralised electronic data warehouse.

31. **Real-time monitoring leverages a mix of APIs, web-scraping, chatbots, text mining and others to fetch data on demand or as a continuous stream.** Real-time monitoring is then coupled with straight through processing and dynamic visualisation to provide an instant read-out of performance indicators. Such solutions go hand in hand with other use cases, in particular with misconduct analysis and market surveillance. For example, the chatbot prototyped by the BSP captures consumer complaints via popular messaging apps and renders summary statistics and performance metrics in a dynamic dashboard near-instantaneously, giving supervisors a live view of market conduct in the domestic banking sector. Similarly, CNBV's and ASIC's market surveillance tools enable real-time monitoring and scrutiny of transaction and fund flows.

Data management

32. **Data management use cases refer to suptech initiatives that target specific points in the data life cycle.** This is in contrast with end-to-end solutions that cover the data pipeline in its entirety (eg real-time monitoring) or focus chiefly on collection (eg automated reporting). The three key data management tasks are validation, consolidation and visualisation.

33. **Data validation refers to the quality control checks of completeness, correctness and consistency of formatting and calculation in accordance with reporting rules.** Suptech can substitute for time-consuming and error-prone manual "spot checks" or spreadsheet-based formulas, which are unsuited for working with large data sets. Most initiatives combine elements of RPA and ML algorithms that detect atypical data points in submissions. OeNB has applied ML and deep learning algorithms to predict the probability that a data set contains errors that need to be rectified by the reporting entity, as

[17] Bauguess (2017).

[18] Auer (2019) makes a case for *"embedded supervision"* for automated reporting for blockchain-based financial markets. This is a framework that allows compliance with regulatory goals to be automatically monitored by reading the market's ledger, thus reducing the need for firms to actively collect, verify and deliver data.

[19] The data cube approach is described in Piechocki and Dabringhausen (2015).

well as knowledge graphs for master data "plausibilisation" (a method for removing unrealistic or erroneous values).

34. **Consolidation involves the integration of data from multiple sources and in varying formats.** Here too ML techniques can prove helpful. Disparate data sets often contain relevant information about different dimensions of the same subject (eg a company, an individual) but integrating them can be exceedingly cumbersome. If unique identifiers[20] are not available or are inconsistent across data sets and manual merging is difficult, then ML approaches can be used to match data sets using probabilistic scores. For example, the DB uses both a centralised data platform and ML-based merging techniques to build its so-called "House of Microdata".[21]

35. **Visualisation use cases refer to interfaces that sit atop big data architectures and provide seamless and interactive user experience with minimal latency.** Often they substitute for static, spreadsheet-generated dashboards that require manual updating. To extract the most meaningful and actionable insights from data, big data dashboards allow for numerous analytical operations, such as drilling up (ie summarising data along one dimension) and drilling down (ie navigating deeper along a dimension), as well as slicing, dicing, pivoting and overlaying data across multiple dimensions. For instance, MAS is developing a supervisory dashboard that simplifies the experience of using data and provides supervisors with "at-a-glance" visibility of the health of the financial institutions in their portfolio. It also enables an easy comparison of performance between different financial institutions and peer groups, alerts users to outliers and/or significant changes of indicators, and allows drilling down to derive deeper insights on certain risks.

Micro- and macroprudential supervision

36. **Use cases for microprudential supervision seek primarily to leverage ML applications to improve prudential oversight of risks, such as liquidity and credit risks, as well as governance and culture.** The DNB has studied the use of neural networks to detect liquidity problems at banks in anticipation of potential deposit runs.[22] Sentiment analysis tools such as the BdI's twitter-based indicator may also be deployed for this purpose.[23] The Federal Reserve System is experimenting with the use of NLP to identify and extract topical information from voluminous text for microprudential purposes. On the other hand, the Central Bank of the Russian Federation (CBR) has designed software that accelerates the credit risk internal ratings-based (IRB) approach supervision process, for example, by running statistical tests to check the level of accuracy, discrimination and stability of banks' models for calculating capital requirements.[24] Similarly, the BOT has operationalised a credit risk assessment model that applies logistic regression and random forest algorithms to granular data from commercial credit contracts in order to generate a credit score for individual lenders.[25] The BOT has also developed a NLP platform that is used

[20] The legal entity identifier (LEI), initiated by the FSB, is an important example of unique identifiers. It enables a clear and unique identification of entities participating in financial transactions.

[21] Buch (2018).

[22] See Triepels et al (2017).

[23] Accornero and Moscatelli (2018).

[24] Central Bank of the Russian Federation (2018).

[25] Bank of Thailand (2018).

to analyse executive committee meeting minutes to study corporate governance ("behaviour and culture analysis").[26]

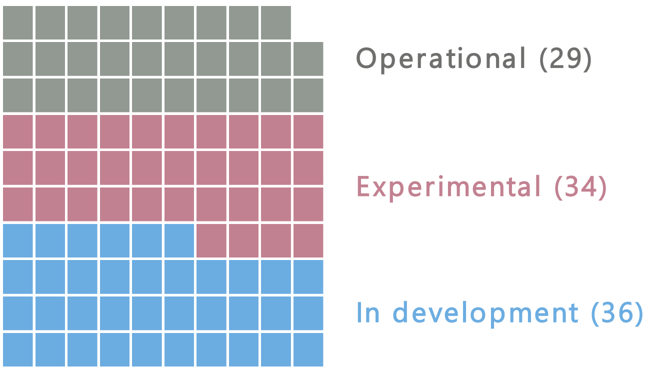
37. **Use cases for macroprudential supervision use similar ML techniques.** The BoG is experimenting with ML and deep learning to anticipate bank insolvencies, essentially creating an early warning system for bank failures that can serve both micro- and macroprudential purposes.[27] The BdI is using ML techniques to analyse real estate ads in a popular online portal to forecast housing prices and inflation. BdI is also exploring the use of deep neural networks to detect potential liquidity problems that might affect payment system participants.

Section 6 – Suptech development and deployment

38. **Suptech solutions have emerged only recently, with a marked take-off in 2019.** The emergence of big data- or AI-enabled suptech followed the take-up of regtech by financial institutions. The main impetus for the latter came from efforts by financial institutions to ease a rising regulatory compliance burden, coupled with the growing availability of big data and AI tools. Suptech has generally lagged behind developments in the fintech/regtech space. Aside from a deliberate strategy to sequence and pace IT upgrades gradually, the reasons for the relatively late embrace of suptech may be ascribed to (i) concerns among financial authorities about the uncertain value and risks of suptech; (ii) resource constraints; and (iii) a limited product offering for suptech solutions from a small pool of specialised technology vendors. The inertia inherent in legacy IT systems is another factor.

39. **Less than a third of suptech initiatives covered in the paper are operational, with the majority still either in experimental or development stages (Graph 7).** The experimental nature of many suptech initiatives suggests financial authorities are willing to explore the subject, provided that IT budgets and infrastructure are not unduly compromised. Experimentation takes the form of open-ended innovation programs or projects that result in POCs. POCs are effective in validating ideas with uncertain prospects, paving the way for prototypes and eventually full-scale development. Laboratory-like settings also allow staff to use low- or no-cost software, such as freely available machine learning models in R and Python libraries, to test ideas without expending scarce resources or introducing potential cyber security threats lurking within the open-source code. Once a POC or prototype has demonstrated a minimum viability, the models can either be rigorously inspected for bugs and vulnerabilities through penetration tests or reconstituted in a controlled environment.

Graph 7

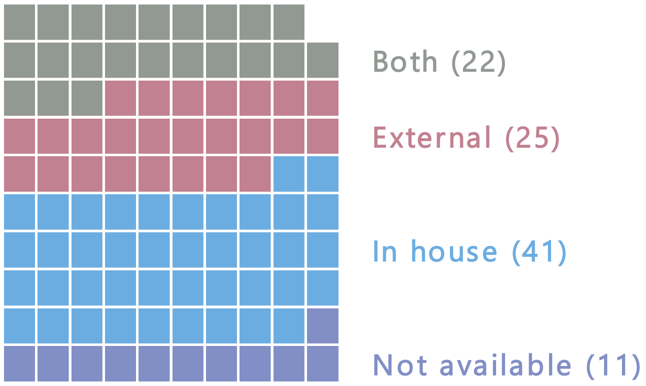


[26] Meeting minutes are collected in machine-readable form (PDF and Microsoft Word) via a web portal and stored in the BOT's document and record management system, before they are run through a Python-based topic modelling algorithm.

[27] Petropoulos et al (2017).

40. **Nearly three quarters of suptech initiatives covered in the paper were or are being developed internally or jointly with external developers and organisations such as universities, with only a quarter under development by external vendors (Graph 8).** This skew likely reflects the experimental and open-ended nature of these projects. Some of these initiatives may lack clearly defined functional requirements or technical specifications (details that are required in traditional "request for proposals") with which to engage outside developers, or sufficient research and development resources with which to entice them. On the supply side, there is a dearth of specialised products or providers in the suptech space. For example, when the United Kingdom's FCA undertook to develop digital regulatory reporting, it could not find a single vendor with the requisite technology or expertise in this space and instead had to launch two tech sprints to incubate the idea. Despite a compelling value proposition and growing demand for suptech solutions, the small addressable market for prospective suptech vendors, both in terms of the quantity of clients and their relative spending power, plus rigid procurement processes and long lead times can discourage private sector participation. Anecdotal evidence also points to a mismatch between financial authorities' needs and outside innovators' capacities.

Graph 8



41. **External vendors mentioned in the survey range from a handful of firms specialising in suptech, to a larger pool of regtech solutions providers and major enterprise IT firms providing commercial off-the-shelf (COTS) solutions.** In a few cases, financial authorities opt for smaller local vendors. What these smaller vendors lack in economies of scale and scope they frequently make up for through lower servicing costs, more customisation, and greater ease of communication and coordination between co-located clients and developers. Data localisation restrictions and public policy objectives for building domestic data science capacity may also factor into procurement decisions.

42. **Some initiatives take the form of strategic partnerships between financial authorities and other governmental agencies, industry, academia and research organisations.** Such arrangements are especially helpful where there is a common need for data pooling and intelligence sharing. Suptech can help integrate data streams from multiple sources into a common pool from which richer insights can be drawn to serve different user groups. Academic partnerships, meanwhile, can be fruitful for exploratory projects on the cutting edge of suptech research.

Section 7 – Concluding remarks

43. **Financial authorities have refined their use of technology over the years, leading to technologies that this paper would consider as suptech.** Authorities have constantly leveraged different

technologies in order to develop tools that help support their work to better carry out their mandates. As outlined in this paper, these technologies have resulted in a range of useful tools, from spreadsheets, desktop databases, dashboards, web-based portals etc all the way to the use of big data architecture and the application of AI. The application of big data or AI to tools used by authorities represents the latest generations of technologies that are considered to be suptech for the purposes of this paper.

44. **It is important for authorities to continue enhancing existing tools, including those not considered suptech.** Some authorities covered in this paper reported using or developing tools that do not fall under the definition of suptech offered in this paper. However, these tools still represent improvements on the tools used previously. Indeed, authorities can only benefit by taking advantage of advances in technology. Newer technologies enable more efficient processes for data collection and analyses, resulting in more timely and better decision-making. At the same time, the technologies that authorities use should be appropriate to the size, complexity and development of the sectors they oversee. For example, investments in big data architecture and AI tools might not be appropriate for an authority in a low-income jurisdiction that supervises only a handful of financial institutions providing basic financial products and services. Moreover, authorities should also be aware of the issues and challenges associated with suptech, as outlined in Broeders and Prenio (2018). In particular, the lack of transparency in some of the suptech data analytics solutions is a critical issue. This underscores the continued need for human intervention in the form of supervisory expertise to further investigate the results of analyses and when deciding on a course of action.

45. **Pursuing suptech will help authorities to become more data-driven.** Data are vital for financial authorities. They inform decisions and any actions taken with respect to institutions or markets. Earlier generations of technology resulted in tools that hindered the optimal use of data. This resulted in less efficient data collection and analysis. These earlier tools tended to trap data in silos, and could not exploit newer sources of data or deliver timely insights. Suptech has the potential to address these shortcomings. Big data architecture supports seamless data processes, while AI allows large volumes of information to be integrated from disparate sources and analysed. But the acquisition of such tools needs to be accompanied by a corresponding data expertise and mindset if a data-driven culture is to be successfully embedded in the organisation.

46. **There are many ways of exploring suptech tools and these are not mutually exclusive.** Most financial authorities covered in the paper either have suptech strategies in place or are in the process of developing them. Approaches to suptech strategy vary; they range from developing specific suptech roadmaps to incorporating suptech into an institution-wide DT&DI programme. These approaches are typically pursued in combination. A well defined strategy can help authorities optimally realise the potential benefits of suptech. But for authorities who want to explore specific suptech tools first, before committing substantial resources, there are helpful avenues, such as innovation labs, accelerators or tech sprints. These methodologies can also be included in authorities' existing or future suptech strategies.

47. **Suptech is still in its infancy but it is gaining traction.** The latest big data or AI-centred generations of technology that supervisors are either using or exploring have emerged only recently. In fact, most suptech solutions are still in either the experimental or the development stages. But suptech is gaining momentum. While the universe of suptech use cases has remained broadly unchanged, there seems to be a significant increase in the number of authorities as well as the number of initiatives that are exploring suptech tools.

48. **Suptech tools are applied mainly in misconduct analysis, reporting and data management.** Given the sheer amount of information, big data and ML tools show huge potential to support authorities' activities in misconduct analysis, including in the fight against financial crime. These tools will support financial authorities in better tackling non-financial risks, an area which has inflicted massive reputational damage on financial institutions.

49. **The experimental nature of most suptech initiatives may have prevented a greater number of external parties from participating in the development of suptech solutions.** Most solutions are

being developed within financial authorities, or at least partly using internal resources, with only about a quarter being developed solely by external parties. This may be due to the experimental nature of these initiatives, among other reasons. Consequently, some initiatives may lack functional requirements or technical specifications that are sufficiently defined to engage external parties. Thus, strategic partnerships with other authorities, academia and research organisations will be important in overcoming the challenges associated with the experimental nature of these initiatives.

50. **International coordination and collaboration could help to accelerate suptech development.** Global standard-setting bodies and international organisations provide platforms for authorities to exchange information on their suptech initiatives. These international platforms could also be used potentially to collaborate on the development of suptech solutions that may be useful to a number of authorities or to address cross-border issues affecting suptech development (eg data localisation). A good example is the recently announced BIS Innovation Hub, which is designed to foster international collaboration on innovative financial technology within the central banking community. Such platforms can help authorities to benefit from peer learning, including from different types of authority (central banks, prudential regulators, conduct regulators etc), helping to offset the lack of specialist providers. They also reduce the need for individual authorities to independently work on similar solutions, thus increasing efficiency. In addition, given the inherently small market for suptech solutions, which limits business opportunities for private IT providers, accelerators set up and/or funded by international organisations can play an important role in helping authorities to explore specific suptech tools.

References

Accornero, M and M Moscatelli (2018): "Listening to the buzz: social media sentiment and retail depositors' trust", Bank of Italy, February.

Auer, R (2019): "Embedded supervision: how to build regulation into blockchain finance", *BIS Working Papers*, no 811, September.

Bank of England: "Frequently asked questions on Bank of England Accelerator", Bank of England.

Bank of Thailand (2018): "Credit scoring model: Tools to assess credit quality", Bank of Thailand, August.

Bauguess, S (2017): "The role of big data, machine learning, and AI in assessing risks: A regulatory perspective", keynote address, OpRisk North America, New York, June.

Broeders, D and J Prenio (2018): "Innovative technology in financial supervision (suptech) – the experience of early users", *FSI Insights on policy implementation*, no 9, July.

Buch, C (2018): "Can technology and innovation help? New data-generating possibilities", speech at the European Central Bank conference on statistics, Frankfurt, 12 July.

Central Bank of the Russian Federation (2018): "Questions and directions regulatory development and supervisory technologies (RegTech and SupTech) in the financial market in Russia", October.

Coelho, R, M De Simoni and J Prenio (2019): "Suptech applications for anti-money laundering", *FSI Insights on policy implementation*, no 18, August.

Cœuré, B (2017): "Policy analysis with big data", speech at the conference on economic and financial regulation in the era of big data organised by the Bank of France, November.

Di Castri, S, M Grasser and A Kulenkampff (2018a): "An AML suptech solution for the Mexican National Banking and Securities Commission (CNBV): R2A project retrospective and lessons learned", RegTech for Regulators Accelerator, August.

——— (2018b): "A chatbot application and complaints management system for the Bangko Sentral ng Pilipinas (BSP): R^2A project retrospective and lessons learned", RegTech for Regulators Accelerator, August.

European Central Bank (2017): Guideline 2017/2335 of the European Central Bank of 23 November 2017 on the procedures for the collection of granular credit and credit risk data (ECB/2017/38), November.

Financial Stability Board (2015): "Report to the G20 on actions taken to assess and address the decline in correspondent banking", November.

——— (2017): "Artificial intelligence and machine learning in financial services: market developments and financial stability implications", 1 November.

Petropoulos, A, V Siakoulis, E Stavroulakis and N Vlachogiannakis (2017): "Predicting bank insolvencies using machine learning techniques", Bank of Greece, April.

Piechocki, M and T Dabringhausen (2015): "Reforming regulatory reporting – from templates to cubes", BearingPoint, December.

Tissot, B, T Hülagü, P Nymand-Andersen and L Comino Suarez (2015): "Central banks' use of and interest in 'big data'", Irving Fisher Committee on Central Bank Statistics, August.

Triepels, R, H Daniels and R Heijmans (2017): "Detection and explanation of anomalous payment behavior in real-time gross settlement systems", Netherlands Bank.

Glossary

application programming interface: a set of rules and specifications followed by software programmes to communicate with each other, and an interface between different software programmes that facilitates their interaction; APIs enable direct database-to-database data transmission enabling granular, real-time reporting and automated validation.

artificial intelligence: the theory and development of computer systems able to perform tasks that traditionally have required human intelligence.

big data: a generic term that designates the massive volume of data that is generated by the increasing use of digital tools and information systems. Big data encompasses technologies that significantly increase the volume, variety, velocity and validity of data under management.

chatbot: virtual assistance programmes that interact with users in natural language; chatbots enable automated capture and interpretation of qualitative data, enabling data collection in real time.

cloud computing: use of an online network ("cloud") of hosting processors so as to increase computing capacity and its flexibility.

dashboards: customisable, dynamic interactive reporting tools that automatically fetch and render data in meaningful and actionable visualisations.

data cubes: granular data storage and transmission solution enabling real-time data collection.

data lake: scalable storage solution for diverse structured, semi-structured, and unstructured data.

distributed ledger technology: a network to securely propose, validate and record changes to a synchronised ledger distributed across multiple nodes.

geographic information systems: automated analysis of spatial or geographic data.

machine learning: a method of designing a sequence of actions, with the aim of solving a problem, that optimise automatically through experience and with limited or no human intervention. ML enables automated anomaly detection, merge-sort, scoring and other use cases.

natural language processing: an interdisciplinary field of computer science, artificial intelligence and computation linguistics that focuses on programming computers and algorithms to parse, process and understand human language.

network analysis: the process of investigating structures through the use of networks and graph theory.

robotic process automation: partial or full automation of manual, rule-based and repetitive human activities by robotics software or "bots".

self-organising maps: a type of artificial neural network that is trained using unsupervised learning to produce a low-dimensional, discretised representation of the input space of the training samples, called a map, and is therefore a method of performing dimensionality reduction.

smart contracts: programmable applications that can trigger financial flows or changes of ownership if specific events occur.

text mining: the process of exploring and analysing large amounts of unstructured text data aided by software that can identify concepts, patterns, topics, keywords and other attributes in the data.

web portal: static file upload via web site with built-in automated validation checks.

web scraper: automated capture of web data by programs or "bots".