

Project Ellipse

An integrated regulatory data and analytics platform

March 2022





Foreword

Ross Leckow

Acting Head, BIS Innovation Hub Regulators need accurate and timely information to assess emerging risks and to make informed supervisory decisions. Technology can help supervisors by giving them access to more data, structured and unstructured, with greater predictive insights than ever before.

Project Ellipse shows how authorities could use a single platform for "on demand" access to timely and integrated sources of data to help support and inform their supervisory assessments.

This project showcases our strong partnership and collaboration with the Monetary Authority of Singapore and we wish to also thank our project partners at the Bank of England, the International Swaps and Derivatives Association and Financial Network Analytics.

I am confident that Project Ellipse will inspire authorities around the world to further consider, explore and collaborate on common solutions that can future-proof the data and analytical capabilities of supervisors.

Hern Shin Ho

Deputy Managing Director, Monetary Authority of Singapore Technological advancements in recent years mean that there is scope to significantly improve supervisory effectiveness and outcomes by leveraging on data and technology.

Project Ellipse is a prototype that shows how technology can be used to help supervisors undertake more real-time risk surveillance, and conduct pre-emptive supervision. It demonstrates how data analytics techniques like natural language processing and sentiment analysis can help supervisors to collate and analyse vast volumes of data from disparate sources to gain more timely situational awareness of adverse events that could impact our financial institutions or system.

I want to thank the BIS Innovation Hub for collaborating closely with the Monetary Authority of Singapore, to envision how technology can transform the way we supervise regulated institutions. Project Ellipse is an excellent example of what we can achieve together through collaboration, and MAS is adapting the prototype for our own supervisory needs. I hope other supervisors will similarly find it useful, and look forward to further initiatives to develop common technological solutions for supervisors.

Contents

Forev	vord	2
Acror	ıyms	5
List o	f tables and graphs	6
Execu	itive summary	7
1	Introduction	9
2	Challenges with regulatory data in the digital age	11
	2.1. Regulatory reporting2.2. Incorporating other sources of data2.3. Rapid development in technology2.4. Towards an integrated regulatory data and analytics platform	12 12 13 16
3	Ellipse Phase 1: Granular data, common data models and machine- executable reporting	19
	3.1. Common data models and data standards3.2. Cross-border data model using retail mortgages3.3. Machine-executable reporting using the Common Domain Model	20 21 22
4	Ellipse Phase 2: Integrating data and analytics onto one platform	25
	 4.1. Identifying user needs Hypothetical credit risk scenario Practical challenges Entity name reconciliation Availability of market and financial analyses Mapping exposures to high-risk/stressed entities and networks Assessing potential impact on exposures based on current events Automation of capital impact under stressed conditions 	26 26 27 27 27 28 28 28
	 4.2. Platform functionalities using a storyboard Reconciling corporate entity names Market and financial analyses Mapping exposures to entities and networks Analytics on exposures ad current events Automating potential capital impact Other functionalities 	29 29 30 32 33 34 35

	4.3. Prototype development	36
	Prototype user features	37
	Home screen	37
	Top regulatory alerts	38
	Top exposure table	39
	Automating potential capital impact	40
	Simulate stress testing	42
	Network analytics	43
	Fire sale contagion	43
	Interbank loan contagion	44
	Supply chain contagion	45
	4.4. High-level architecture for the Ellipse Platform	46
	Data connectors	47
	Text analysis	47
	Data sources	47
	Combining structured and unstructured data	47
	Risk alerts	48
	Stress testing exposures	48
	Network analytics	49
	Data attributes	49
	Fire sale contagion dynamics	49
	Interbank contagion dynamics	49
	Operational considerations	50
5	Findings, opportunities and considerations for supervisors	51
	5.1. Findings from the Ellipse Project	52
	5.2. Opportunities and consideration for supervisors	53
Anne	xes	55
Gloss	ary of terms	63
Defe	,	
keter	ences	64

Acronyms

_

AI	Artificial intelligence
API	Application programming interfaces
BIS	Bank for International Settlements
ВоЕ	Bank of England
CDM	Common Domain Model
CET1	Common Equity Tier 1 (capital)
CRM	Credit risk mitigation
EAD	Exposure at default
EBIT	Earnings before interest and tax
FNA	Financial Network Analytics
GDP	Gross domestic product
ICR	Interest coverage ratio
ISDA	International Swaps and Derivatives Association
MAS	Monetary Authority of Singapore
ML	Machine learning
NLP	Natural language processing
отс	Over-the-counter
PD	Probability of default
РоС	Proof of concept
RWA	Risk-weighted asset

List of tables and graphs

Tables

Table 1	Challenges of regulatory reporting
Table 2	Possible solutions explored in Project Ellipse

Graphs

Graph 1	Overview of the Project Ellipse phases
Graph 2	Overview of attributes for retail mortgages
Graph 3	Common Domain Model extended for retail mortgages
Graph 4	Simulated retail mortgage regulatory reports referencing the same data model
Graph 5	Visualising the Ellipse dashboard
Graph 6	Display of regulatory alerts
Graph 7	Visualisation of network analytics
Graph 8	Display of exposures from a single supervised entity
Graph 9	Integrated risk analytic tools
Graph 10	Systemic stress test capabilities on the banking system
Graph 11	Integration of regulatory reports into the Ellipse dashboard
Graph 12	Login screen to the Ellipse dashboard
Graph 13	Ellipse home screen
Graph 14	Real-time feeds for top regulatory alerts
Graph 15	News alert search function
Graph 16	Display of largest exposures by bank
Graph 17	Overview of bank's top corporate exposures
Graph 18	Information on a corporate entity
Graph 19	Real-time display of entities flagged by risk type
Graph 20	Estimating impact of corporate default on banks and the banking system
Graph 21	Estimating potential contagion risk to related companies
Graph 22	Information on exposures by sector and by entity in the loan book
Graph 23	Stress scenarios to estimate possible impact on banks
Graph 24	Display of network effects through fire sale contagion
Graph 25	Display of network effects through interbank loan contagion
Graph 26	Display of network effects through supply chain contagion



Executive summary

A transformational shift in the volume, speed and variety of data is driving the innovative use of financial technology, leading to rapid changes in the financial landscape. At the same time, regulatory authorities still rely on the collection of template-based supervisory data, which has remained largely unchanged. Supervisors are faced with the challenge of needing to assess rapidly evolving risks to business models and technology-driven changes that may affect financial stability, with regulatory data that are infrequent and collected according to legacy frameworks.

In January 2021, the BIS Innovation Hub Singapore Centre and the Monetary Authority of Singapore (MAS) launched Project Ellipse. With the support of the Bank of England (BoE), the International Swaps and Derivatives Association (ISDA), Financial Network Analytics (FNA) and Accenture, Project Ellipse explores how technology solutions could enable supervision to be more forward-looking, insights-based and data-driven, using an integrated regulatory data and analytics platform.¹ Importantly, the Ellipse prototype combines both structured and unstructured sources of data that are relevant to current events in real time. Advanced analytics are then applied to those integrated data sources to provide supervisors with early warning indicators, analytics and prudential metrics.

Project Ellipse was undertaken in two phases. In Phase 1, the project investigated how data-driven supervision could be enabled by machine-executable digital reporting, using a cross-border common data model. Our exploration found that

¹ We would like to extend our sincere thanks to the many public authorities and private sector participants for their time, insights and feedback throughout the project. A special note of thanks to the teams at the MAS, BoE, ISDA and FNA who provided significant support towards the development the Ellipse prototype. The project also benefited enormously from the extensive feedback and insights provided by participants from the Australian Prudential Regulation Authority, Basel Committee on Banking Supervision, BIS Innovation Network, Bank Negara Malaysia, Bank of England, Bank of Japan, Bangko Sentral ng Pilipinas, European Central Bank, Federal Deposit Insurance Corporation, Federal Reserve Board, Federal Reserve Bank of New York, Financial Services Authority of Indonesia, Financial Services Agency of Japan, Financial Stability Institute, Hong Kong Monetary Authority, International Institute of Finance, Monetary Authority of Singapore, Oesterreichische Nationalbank, Office of the Superintendent of Financial Institutions in Canada, Reserve Bank of India, South African Reserve Bank, Swedish Finansinspektionen, Swiss Financial Market Supervisory Authority and Swiss National Bank.

regulatory reporting requirements can be expressed in unambiguous machinereadable logical reporting instructions underpinned by a consistent data model. Programmatic specifications of the steps for generating regulatory reports can also be published alongside regulations to ensure a clear understanding of the expected data at the most granular level. With additional logical instructions based on the same data model, supervisors could also automatically query the underlying transaction data and generate regulatory metrics referencing that standardised data. Phase 1 illustrated the possibilities and the efficiencies that could be gained if machine-executable reporting using common data models were to be adopted. This could also increase the volume of granular data available to supervisors, as needed to enable the use of advanced analytics.

In Phase 2, the project took existing large exposures regulatory data and integrated these with unstructured data. Advanced analytics such as machine learning and natural language processing were applied to these data sources to make risk correlations and to analyse sentiment, alerting supervisors in real time of issues that might need further investigation. Network analytics were also used to demonstrate how exposures could be mapped, indicating possible systemic risks to the banking system. The Ellipse platform prototype was developed, which can extract insights from the mined data and display these via dashboards as early warnings for supervisory attention. The second phase of Project Ellipse demonstrates how a single platform could be built so that authorities could benefit from "on demand" access to timely and integrated sources of data to help support and inform their supervisory assessments.

The BIS Innovation Hub's Project Ellipse is a prototype that authorities can test in their own environments and which may help them to explore new solutions. It also presents an opportunity for the global regulatory community to further consider, explore and collaborate on common solutions to future-proof the data and analytical capabilities of supervisors.



1. Introduction

In recent years, technology and innovation have spurred rapid and exponential changes in the financial landscape. These have been enabled in part by a transformational shift in the volume, speed and variety of data, driving the innovative use of financial technology.² At the same time, the collection of supervisory data through regulatory reporting has remained largely unchanged and is still submitted infrequently using fixed, aggregated templates. Supervisors also manually scan through vast volumes of market and news data to assess early indicators of emerging risks. At times of heightened risk, the need for up-to-date data increases but, given the fixed static nature of regulatory reports, supervisors may not have the timeliest data to make informed judgments. In this environment, supervisors therefore face the dual challenges of needing to assess the forward looking impact that these technology-driven changes may have on business models and risk profiles, with regulatory data that have been collected according to legacy frameworks.

To meet these challenges, in January 2021, the BIS Innovation Hub Singapore Centre and MAS launched Project Ellipse. With the support of the BoE, ISDA and FNA, Project Ellipse explores how technology solutions could enable supervision to be more forward-looking, insights based and data driven. The Ellipse proof-of-concept (PoC) is an integrated regulatory data and analytics platform that combines structured granular data and unstructured sources of data that are relevant to current events in real time. Advanced analytics are then applied to those integrated data sources to provide supervisors with early warning indicators, analytics and prudential metrics.

The project has proceeded in two phases. In Phase 1, Ellipse investigated how data-driven supervision could be enabled by machine-executable digital reporting, using a cross-border common data model. Phase 1 illustrated the possibilities and the efficiencies that could be gained if machine-executable reporting using common data models were to be adopted. This could also increase the volume of granular data available to supervisors, as needed to enable the use of advanced analytics. In Phase 2, advanced analytics such as machine learning and natural language processing were applied to unstructured data and granular reporting data to make risk correlations and to analyse sentiment, alerting supervisors in real time of issues that might need further investigation. Network analytics were also used to demonstrate how exposures could be mapped to indicate possible systemic risks to the banking system.

This report provides a detailed description of the project and outlines the features and functionalities of the Ellipse prototype. Section 2 sets out the challenges supervisors face with regulatory reporting, the rationale for the project and the scope and objectives of the Ellipse PoC Section 3 describes the use case in Phase 1, while Section 4 outlines the development of the Ellipse platform prototype. Section 5 concludes with the project findings, opportunities and further considerations for supervisors.

² Also commonly referred to as the three "Vs" - volume, velocity and variety.



2

A

2. Challenges with regulatory data in the digital age

2.1 Regulatory reporting

A core principle for the effective supervision of banks and banking systems requires the supervisor to develop forward-looking assessments of banks' risk profiles, to identify and assess risks that may emerge from banks and the banking system, and to intervene early if needed.³ To do this effectively, supervisors rely on various sources of information to assess and evaluate the safety and soundness of banks, including prudential reports, statistical returns and publicly available information. Much of this information is collected from regulatory reporting, which provides supervisors with data to undertake these assessments on individual banks (microprudential risks) and on the broader financial system (macroprudential risks).⁴ These data are supplemented with other qualitative information, such as from management reports, news and market developments.

Supervisors today therefore still rely heavily on regulatory reporting to inform them of potential risks that may be forming within regulated entities, which can have broader implications for the financial system. However, there are challenges with the information that supervisors receive from regulatory reports. This information is often template-based, meaning that it is fixed by use case, and hence the data received cannot easily be reused for other purposes. Reporting data are also often sourced from legacy corporate data systems that are not always interconnected, which results in data points for any given product or transaction being described differently, both within a bank and across different banks. These reports contain aggregated data, which means that potentially important insights into risks can be lost when data are shown only at a portfolio level. More importantly, regulatory reports are submitted from reporting entities infrequently (eg every month or quarter), so that, when received, the information is dated and backward-looking.

At times of heightened risk, supervisors need information that is current and up-to-date but in many cases it is difficult to obtain data quickly from reporting entities. This is because supervisors must create reporting requirements and templates each time new information is needed. Given the challenges of compiling and receiving regulatory data, supervisors may not have the timeliest data when taking decisions. Moreover, as the pace of change continues to impact the rapid development of new digitally-enabled financial products and services, static, infrequent and aggregated reporting does not support new tools or applications that could enable real-time or early warning analytics.

2.2 Incorporating other sources of data

In addition to regulatory data, supervisors use other sources of information to help inform their prudential assessments. These include management reports, financial and regulatory disclosures, external credit assessments, market data and news. This type

³ Basel Committee on Banking Supervision (2012).

⁴ Statistical and regulatory data constitute the largest category of information that authorities collect from financial institutions: see Crisanto et al (2020).

of information can be described as unstructured, as it is not organised in a specific or unified format, unlike regulatory reporting which is based on data that are structured using reporting templates. Information for regulatory reporting is often linked with these other types of unstructured information that may help to support or validate the assessment of emerging risks. For instance, information sourced from market data and news often provides the first indication of emerging risks but it is difficult for supervisors to scan through the vast volumes of market and news data and correlate these developments manually with every regulatory return received.

The challenge, therefore, is how supervisors can form an accurate picture of exposures and gain predictive insights into emerging risks using these disparate data sets. This is further compounded by newer sources of unstructured data that have not been systematically used and incorporated into prudential assessments by supervisors. Information from web content, emails, images or social media posts, for example, could validate or even contradict information being received from regulatory reporting. However, gathering intelligence from these unstructured data sources in addition to regulatory reporting requires a fundamental shift away from current manual processes to ones that are enabled by technology.⁵

2.3 Rapid developments in technology

As data sources are transformed in both volume and variety, there is considerable opportunity to complement these changes by applying new analytical tools, particularly in relation to unstructured data. There is increasing interest globally in the use of innovative technology by financial authorities, otherwise referred to as suptech. Such applications include artificial intelligence (AI) and machine learning (ML), application programming interfaces (APIs) natural language processing (NLP) and text analysis to name but a few.6 The adoption of these tools could enable a shift from the traditional rules-based and principles-based approaches to supervision towards a system that is data- and insights-driven.⁷ Supervision that is insights-based using a suite of new technologies could make it possible for supervisors to use multiple sources of both structured and unstructured data sets to drive predictive analytics. Early warning and monitoring of risks could enable supervisors to make earlier interventions if needed. The adoption of such technology by supervisors could also enable better oversight of its use by regulated entities.

Nevertheless, while many authorities recognise the enormous process efficiencies and analytic capabilities that could be gained by adopting these new tools, a large majority remain in what has been described as the "first and second generations" of technology use.⁸ Such early generations of data architecture support mostly descriptive and diagnostic analytics, such as "what happened and why did it happen?". This is in stark contrast to the "third and fourth generations" of technology adoption, which enable the use of big data and AI to power predictive and prescriptive analytics, namely "what will happen and what should I do about it?"9

Toronto Centre (2018).

Financial Stability Board (2020).

World Government Summit (2018).

Newer sources of unstructured data will require an architecture that can support the collection, storage, analysis and visualisation of these data sets. Such architectures include technology stacks that support data of higher granularity, diversity and frequency. The end-to-end data pipeline, from input, consolidation, interrogation and visualisation, would then be integrated to support advanced tools such as ML with the aim of connecting, merging and querying correlations from various data sources. NLP, drawing on unstructured news flow, could also help in entity resolution and network mapping to detect possible transmission channels via common exposures or common borrowers. These insights could then be displayed on intuitive dashboards, tailored for specific user needs and used to support further supervisory investigation (**Table 1**).

Table 1 — Challenges of regulatory reporting

Template based, aggregated



CHALLENGES FACED

Regulatory requirements are often template-based and call for aggregated data, meaning that data sets are fixed to a use case and hence the data received cannot be easily reused for other purposes. New reporting requirements are needed whenever additional or ad hoc information is needed.

Infrequent, backward looking

Regulatory reports are submitted to supervisors from reporting entities

on an infrequent basis (eg every month or quarter). At times of heightened risk, the need for upto-date data increases but, given the static nature of regulatory reports, supervisors may not have the timeliest data to make informed judgments.

Data are inconsistently described

2



Reporting data are often sourced from reporting firms' legacy data systems, which may not be integrated. This often results in heterogeneity of data for any given product or transaction – both within a bank and across different banks – as different systems will describe these data differently.

Different sources of data are not integrated

Information contained in regulatory reports is often linked to other



types of information that may point to emerging risks, but these sources of information are not connected. For instance, information sourced from market data and news often gives the first indication of emerging risks but it is difficult for supervisors to scan through the vast volumes of market and news data to assess which point to a need to take early action.

Table 2 — Possible solutions explored in Project Ellipse



Granular data

The collection of granular data from reporting entities could replace the need for authorities



to request information using templates. It could also enable authorities to reuse those data for different use cases. Supervisory metrics could also be derived using granular data, as opposed to requiring reporting entities to aggregate the required data prior to submission.

Common data 2 models

Differences in the description of data similar for products and transactions across banks can be addressed using data standards and common data models. Granular reporting requires a common understanding by authorities and financial institutions of what these data are, so that financial institutions can map their operational data to a common "input" before the required data can be reported. Supervisory metrics could then be derived using programmable rules that reference machine-readable and machineexecutable common data models.

SOLUTIONS EXPLORED

Real-time information



Real-time insights advanced using analytics could be derived from large volumes of unstructured data that would supplement the granular reporting available. This

would provide supervisors with additional indicators and early warnings of any at-risk exposures of reporting entities.

Integration of structured and unstructured data

Integrating granular data from reporting entities with other sources of

unstructured information such as news and market data on the same platform would obviate the need for supervisors to spend time manually scanning for information. Advanced analytics such as AI and ML could be used to make risk correlations and analyse sentiment, alerting supervisors in real time of issues that may need further investigation.

2.4 Towards an integrated regulatory data and analytics platform

The primary objective of Project Ellipse was to explore and demonstrate how an integrated data and analytics platform solution could enable supervisors to digitally extract, access and analyse large and diverse sources of data that are relevant to current events in real time. A digitally enabled data architecture solution also significantly increases the possibilities for testing and applying Al/ML tools for supervisory use, which then can help to build capacity for supervisors to oversee Al/ML use in the financial sector. This project also aims to pave the way for the development of a common platform solution that has the potential to be shared and scaled across regulatory authorities around the world. The greater the number of authorities that converge on a common solution, the greater the potential for there to be convergence on common data inputs, data standardisation and sharing of information to support financial stability.

To meet these challenges in the digital age, Project Ellipse is therefore premised on the assumption that supervisors could benefit from "on demand" access to timely and integrated sources of data to help support and inform their supervisory assessments. Several possible solutions were therefore explored in this project, namely the collection of granular data, the use of common data models, having access to "real-time" information, the integration of structured and unstructured data and the use of a modular data architecture that is flexible enough to accommodate different use cases **(Table 2)**.

Granular data

The collection of granular data from reporting entities could replace the need for authorities to request information using templates. It could also enable authorities to reuse those data for different use cases. Supervisory metrics could also be derived using granular data, as opposed to requiring reporting entities to aggregate the required data prior to submission.

Common data models

Differences in the description of data for similar products and transactions across banks can be addressed using data standards and common data models. Granular reporting requires a common understanding by authorities and financial institutions of what those data are, so that financial institutions can map their operational data to a common "input" before the required data can be reported. Supervisory metrics could then be derived using programmable rules that reference machine-readable and machine-executable common data models.

• Real-time information

Real-time insights using advanced analytics could be derived from large volumes of unstructured data that would supplement the granular reporting available. This would provide supervisors with additional indicators and early warnings of any at-risk exposures of reporting entities.

Integration of structured and unstructured data

Integrating granular data from reporting entities with other sources of unstructured information such as news and market data onto the same platform means that supervisors would not have to spend time manually scanning for information. Advanced analytics such as AI and ML could be used to make risk correlations and analyse sentiment, alerting supervisors in real time of issues that may need further investigation.

Modular data architecture

A technology stack that enables an integrated end-to-end data pipeline where both structured and unstructured data sources are input, consolidated, analysed and displayed using a customisable dashboard user interface.

To demonstrate these possible solutions, the project was undertaken in two phases. Starting in January 2021, Phase 1 explored how authorities could increase the availability of granular reporting data by using common data models. It then extended this concept in a cross-border context to explore whether two authorities in different jurisdictions could use the same common data model to generate reporting requirements. It further investigated the efficiencies gained when adopting machine-executable reporting using machine-readable common data models. Phase 1 was exploratory, to understand how a move to granular data could be undertaken in discrete projects alongside current regulatory reporting frameworks.

Building on this phase of work, Phase 2 commenced in September 2021 and explored the possibility of real-time information and early warning capabilities. As granular regulatory data are not yet widely available to supervisory authorities, the project took existing regulatory data with sufficient levels of granularity and integrated these with unstructured data, using AI and ML to extract insights from these data sources to highlight correlations between current events and supervisory metrics. A platform prototype was developed, which could then extract insights from the mined data and display them as an early warning for supervisory attention via dashboards (**Graph 1**). Our findings and lessons learned are set out in this paper for authorities who wish to explore these solutions in more detail. Graph 1 — Overview of the Project Ellipse phases



Ellipse Phase 1

3

Granular data, common data models and machineexecutable reporting

Ellipse Phase 1: Granular data, common data models and machine-executable reporting

3.1 Common data models and data standards

Data-driven insights start with large quantities of structured granular data. A key component of an Ellipse-type platform that could enable supervisors to be datadriven is a system for the digitally-enabled collection and processing of granular and standardised data. This means that those data are: (i) consistently understood by all stakeholders; (ii) can be repurposed for different use cases; and (iii) represented in such a way as to allow programmable code to reference those data to generate the reporting of regulatory metrics.

Several regulatory authorities have been exploring different ways to replace templatebased, aggregated regulatory reporting with granular data and digital regulatory reporting.¹⁰ The vision of regulatory reporting using granular data removes the need for multiple templates, allowing supervisors to constantly repurpose common data points for different analytical use cases. Nevertheless, granular reporting requires a common understanding by authorities and financial institutions of what those data are, so that financial institutions can map their operational data to a common "input" before the required reporting can be generated. Important initiatives around data standards, taxonomies and data models are being developed for this purpose.¹¹ However, data standardisation more generally remains nascent globally.

This concept is also important in a cross-border context. Currently, global banks must submit regulatory data to multiple authorities on financial products that are similar in many jurisdictions but must report these data in different ways and for different templates. Hence efficiencies would also be gained if authorities in different jurisdictions could also come to a common understanding of data for regulatory reporting purposes. This is because reporting platforms built on common data models might allow global financial reporting entities to fulfil different cross-border reporting obligations using a common input layer. This could reduce compliance burdens for financial institutions in responding to template-based regulatory reporting requests from different supervisory regimes. It could also enable the home and host supervisors of these global reporting entities to compare exposures in a more consistent and transparent way, without the need to directly share confidential regulatory information.

Data standards, taxonomies and data models are therefore integral for digital reporting and for enabling data-driven analytics. A move to machine-executable digital regulatory reporting is therefore closely linked with the need for data standards. There are significant benefits in the medium to long term for financial institutions and authorities to move towards a digital reporting framework based on data standards.

¹⁰ eg the Bank of England, www.bankofengland.co.uk/prudential-regulation/transforming-data-collection; the Oesterreichische Nationalbank (OeNB), www.oenb.at/dam/jcr:d9cdbe0a-a6d4-409a-8ac5-670cad2619b0/05_Kienecker_ Statistiken_3_18.pdf; and the UK Financial Conduct Authority, www.fca.org.uk/innovation/regtech/digital-regulatoryreporting.

¹¹ See eg in Europe, www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/reporting/html/bird_dedicated. en.html and www.ecb.europa.eu/stats/money_credit_banking/anacredit/html/index.en.html and in Italy, www. cooperazionepuma.org/.

However, the initial investment and costs needed to achieve this outcome can steer decisions based on business incentives and risk appetite. While data standards initiatives would ultimately benefit if financial institutions and authorities were to adopt a greenfield approach, and do indeed require them to do so, the perceived scale of such an exercise could also deter stakeholders from embarking on it.

In the absence of global data standards, Project Ellipse explored practical ways in which to tackle the need for granular data. As a first step towards bringing stakeholders closer to a common understanding of data needs, we discussed with authorities if there were similarities or commonalities in the type of data being collected. Second, we looked at industry initiatives to see if data models and taxonomies used for specific product domains could be explored for wider use. Third, we established a small narrowly scoped use case to see if we could find common data attributes and agree their definitions "in principle". Finally, we explored ways in which the requirements for the reporting of those data can be published in programming languages.

3.2 Cross-border data model using retail mortgages

To demonstrate that the utility of common data standards and models extends beyond nationally specific contexts, Project Ellipse explores how stakeholders from different authorities and jurisdictions can gain an understanding of the requirements needed to shift to digital reporting, without needing to invest significant resources upfront. The BIS Innovation Hub partnered with the MAS and the BoE to explore the feasibility of finding a common understanding of data collected for regulatory purposes across reporting regimes. To develop the concept, the Project Ellipse team used the reporting requirements for retail mortgages in the United Kingdom and Singapore.

Retail mortgages were selected for this exercise because, while the collection of these data is common across authorities globally, the templates and formats used to collect them are very different. To illustrate this, Ellipse based its modelling exercise on two regulatory reporting requirements in the United Kingdom and Singapore on retail mortgages. In the United Kingdom, this is referred to as *Product sales data reporting (Mortgages) PSD001* (PSD001) and in Singapore, it is referred to as *MAS Notice 645 Computation of Total Debt Servicing Ratio for Property Loans* (MAS 645). The former is an example of reporting requirements presented in a template, while the latter is an example of requirements that are integrated within prudential instruments.

The aim of the initial exercise was to determine whether common baseline requirements could be established between these respective reporting requirements. This required a common understanding of the rationale for the data collections, the type of information that is needed for supervisory purposes and what information is provided on an optional basis. Despite the varied and sometimes bespoke nature of some mortgage products, the Ellipse team found that mortgage transaction data and the respective reporting requirements aligned around three main components: (i) the loan; (ii) the borrower; and (iii) the collateral. Based on these components, as well as taking into consideration the data that are needed for both authorities for supervisory purposes,¹² a minimum common set of baseline requirements was developed for the purposes of our exploration. Taking the common baseline set of requirements, we were able to drill down further into the three main components to reflect what were considered the key attributes of retail mortgages that were relevant for this PoC **(Graph 2).** Importantly, these data attributes were also intended to be used to calculate or aggregate supervisory metrics. For example, loan size and value of property would be the key inputs for deriving the loan-to-value measure. Other types of aggregation would also be enabled, such as loan-to-income, total loans issued by postcode or total loans issued by purpose (eg owner-occupied property or buy-to-let investment).

Graph 2 — Overview of attributes for retail mortgages



However, agreement on common data attributes necessitates an agreement on their definitions. During the process of identifying attributes for retail mortgages, definitions were sourced from the respective jurisdictional frameworks or from third-party sources where existing definitions did not meet the needs of the respective reporting requirements. These definitions were used solely to inform the attributes of the PoC and are not reflective of the actual prescribed definitions found in either reporting regime. However, it was intended to illustrate the exercise of agreeing to a definition "in principle" for the modelled attributes, as well as to highlight the importance in a cross-border context of definitions that are applicable in other jurisdictions for similar products. Converging on a common understanding of products and their definitions is an important step towards normalising their use in a standardised manner. A list of definitions that we used for this PoC can be found in **Annex 1**.

3.3 Machine-executable reporting: using the Common Domain Model

Digital reporting requires data attributes to be represented in a way that allows functional programming to reference and automate the reporting of regulatory metrics, ie so that they are "machine-readable and executable". Project Ellipse sought to explore

¹² Legislative or regulatory requirements that were specific to a jurisdiction were excluded from this exercise.

whether existing industry data standardisation initiatives could be repurposed for different use cases, such as retail mortgages. Of particular interest were those initiatives that were open-sourced and geared towards being machine-readable and executable.

The Common Domain Model (CDM) developed by the International Swaps and Derivatives Association (ISDA) is an open-source, standardised, machine-readable and machine-executable blueprint for how financial products are traded and managed across the transaction lifecycle.¹³ The product scope of the CDM includes over-the-counter (OTC) derivatives, cash securities, securities financing and commodities. To ensure reusability across different markets, the CDM is designed as a composable model whereby financial objects can be constructed bottom-up based on building block components.

For these reasons, we modelled the PoC's mortgage attributes using the CDM. Testing the CDM's use and extensibility for other product domains such as mortgage loans was key to the PoC, as globally applicable common data models that can be used across products could reduce the number of data models in use by financial institutions and ease the mapping burden. In addition, as the CDM is open-sourced, this allows the model to be more widely accessible for testing within existing environments.





To achieve standardisation across products and asset classes, the CDM identifies logical components that fulfil the same function and normalises them, even when those components may be named and treated differently in the context of their respective markets. For instance, the CDM identifies that, regardless of the asset class or product type, a financial transaction always involves two counterparties trading (ie buying or selling) a certain financial product in a specific quantity and at a specific price. This approach means that a single logical concept such as quantity can represent concepts that may be named and captured differently across markets (eg notional or principal amount).

We followed the CDM design principle of normalising concepts such as quantity, price and party in the representation of financial transactions. Following this approach,

¹³ Documentation for the CDM can be found at https://docs.rosetta-technology.io/cdm/readme.html.

we took the modular components that exist in the CDM and reused concepts such as product, loan, price, collateral and party to reflect the mortgage attributes that characterised loan, borrower and collateral. We also extended the model to reflect attributes that were specific to the PoC around collateral (eg property) or the credit profile of the borrower (eg credit details) **(Graph 3)**.

To simulate the creation of retail mortgage regulatory reports for Singapore and the United Kingdom referencing the same model, we created a dummy mortgage transaction with the following parameters: (i) new/first retail mortgage loan; (ii) owner-occupied; (iii) one main borrower; (iv) a 25-year loan, fixed rate for first five years; and (v) reversion to the firm's standard variable mortgage rate after five years.

Taking this dummy transaction as a proxy data feed, reporting logic was generated based on each of the jurisdictional requirements that referenced the Ellipse mortgage model. Using the CDM, executable code automatically generated from the model definitions enabled us to simulate the creation of retail mortgage regulatory reports for Singapore and the United Kingdom referencing the same model **(Graph 4)**.¹⁴



Graph 4 — Simulated retail mortgage regulatory reports referencing the same data model

Phase 1 of our project illustrates the potential for process efficiencies that could be gained when adopting machine-executable reporting using common data models. It could also increase the volume of granular data available to supervisors, which is needed to enable the use of advanced analytics. Further exploration between regulatory authorities would, however, be needed to validate these findings and to see whether the exercise could be extended to other reporting use cases. Importantly, authorities exploring digital reporting would be encouraged to engage with reporting entities and other authorities responsible for the collection of data. This is because issues related to governance and legislative considerations would need to be addressed for digital reporting to be implemented within respective jurisdictions.

¹⁴ Detailed demonstrations of the modelling, programming and updating of the regulatory reports using the Ellipse mortgage CDM can be found at https://ellipse.bisih.org/.

Ellipse Phase 2

Integrating data and analytics onto one platform

Ellipse Phase 2: Integrating data and analytics onto one platform

4.1 Identifying user needs

The challenges we had identified around regulatory reporting could be addressed through "on demand" access to timely and integrated sources of data to help support and inform supervisory assessments. Some possible solutions were therefore proposed, such as integrating granular data from reporting entities with other sources of unstructured information such as news and market data onto the same platform and using advanced analytics to provide supervisors with real-time insights and early warnings of at-risk exposures of reporting entities. If this solution were viable, it would mean that supervisors would not have to spend time manually scanning for this information. ML could also be used to make risk correlations and analyse sentiment, alerting supervisors in real time to issues that may need further investigation.

To understand if our proposed solution could meet the needs of supervisors, we adopted a technique known as design thinking to help us build a prototype of the Ellipse platform. Design thinking is an innovation process centred on understanding the needs of end users together with solutions that are technologically feasible and economically viable.¹⁵ This process begins with identifying which users would most likely benefit from the solution being developed, gathering information from those users on how they would most benefit from a technology solution and then coming up with ideas that are further developed.¹⁶

Hypothetical credit risk scenario

In this hypothetical case, supervisors have received regulatory data on the large exposures that Bank A, Bank B and Bank C have to Corporate B, which is reported in the news to be facing financial difficulties. However, to the regulator, Corporate B appears to be three different entities as it is not reported and formated in the same way in Bank A, B and C's reports. Supervisors also have information in a separate report on collateral pledged against some of these exposures to Corporate B from Corporate B's parent company, Parent C. The supervisors scan market news and articles daily, and there is news that Parent C is also experiencing losses in a particular sector. The supervisors would need to assess whether the losses experienced by Parent C could impact (i) the collateral pledged to loans provided by Banks A, B and C with respect to their exposures to Corporate B; (ii) the quality of the exposures to Corporate B; and (iii) the potential impact of losses on the banks' balance sheets and capital. This information would be needed for discussions on potential implications with the three mentioned banks.

^{15,16} See eg What is Design Thinking? Design thinking is a process for creative problem solving, www.ideou.com/blogs/ inspiration/what-is-design-thinking.

Practical challenges

Our participants confirmed that, as a starting point, supervisors use regulatory reports containing aggregated information as the basis of their offsite assessments of the risk profiles of regulated entities. This information is submitted to them on a periodic basis (eg every month, quarterly). Information contained in the reports can be linked with other information, such as other regulatory reports, management reports, financial and statutory disclosures, stress testing, horizontal industry analysis, news and market data. However, these sources of information are not systematically integrated in current reporting systems. As a result, supervisors are analysing these sources of information individually and manually for the most part, to derive insights that are used as part of on-site assessments, supervisory ratings and possible supervisory actions.

With this starting point, when presented with the hypothetical credit risk scenario, our participants highlighted five main challenges that they would typically face in such a scenario: (i) corporate entity name reconciliation across different reports; (ii) having access to latest market and financial analyses that are linked to corporate entities in a large exposures database; (iii) mapping and aggregating system level exposures to entities and related entities that are deemed high-risk or under stress; (iv) assessing the quality of exposures and any credit risk mitigation based on current events; and (v) calculating the possible impacts to bank capital and other risk parameters under such possible stressed scenarios.

Entity name reconciliation

A key challenge that participants highlighted was that the names of corporate entities were often provided by banks in different formats, making it difficult for supervisors to immediately reconcile system-level exposure to the same corporate. While unique legal entity identifiers could be used if available and entities are located within jurisdictions, global identifiers are not universally available. The same was true for reconciling those entities to their corporate group structure and other related entities. Taking the example of the hypothetical scenario, it may not be immediately apparent to supervisors if Parent C is of material concern to the creditworthiness of the Corporate B, if information on that parent is not provided or available to the supervisor. Likewise, it may not be immediately apparent if several banks have large exposures to the same Corporate B, unless a supervisor maintains a master list of entity names that enables aggregation of system level exposures.

Availability of market and financial analyses

Central banks and authorities have different departments that undertake market analysis and economic research on an ongoing basis to inform policy and market operations. However, this information is often provided only for specific functions and is generally not shared more widely across departments, such as for microprudential supervision. Further, this information does not generally support analysis on a real-time basis but is typically based on historic time series data. Market operations within central banks would conduct their own assessments nearer to real time but, again, this information tends not to be shared more systematically across the organisation (in the case of an integrated central bank and supervisor). In the case of prudential authorities, offsite horizontal teams would typically scan for market developments but the frequency of updates would not be typically shared on an emerging developments or "real-time" basis. Our participants indicated therefore that technology could help the flow of information and tailor market and economic data for various uses and purposes.

Mapping exposures to high-risk/stressed entities and networks

In a similar manner, feedback suggested that mapping total system or bank-wide exposures to specific corporates that are identified as high-risk or stressed would be challenging under most data collections. While not entirely impossible, it would require a significant degree of manual reconciliation to aggregate the potential total exposures to corporates, and potentially their related parties. Feedback also indicated that, while supervisors are scanning news and looking at information that may indicate that corporates are under stress or are in "high-risk" sectors, it would require supervisors to proactively see if regulated entities had large exposures to these firms.

Supervisors could be supported if there were automated applications for making a link between corporates that are under stress and the extent to which regulated entities have exposures to these corporates. The use of technology could also help to trim down the list of affected banks when there is an adverse event by picking up only those with exposures. Further linking any news to these entities, as well as the countries and sectors where they may operate, to the reporting entity could also enable network analysis and potential concentration risk analysis.

Assessing potential impact on exposures based on current events

To understand the overall impact on a bank, or even a group of banks, if a corporate entity were to become stressed or insolvent, several factors would need to be taken into account, such as changes in collateral value, other credit risk mitigation (CRM), and provisioning. These data are not reported dynamically but can be affected if conditions are stressed. Information on credit risk mitigation that is applicable against reported large exposures is not always contained in the same reports. Furthermore, as these data are not updated frequently, supervisors may not have visibility on whether current events are impacting the quality of the collateral.

Our participants suggested that supervisors could therefore be supported if information were available linking exposures with any applicable credit risk mitigation. Importantly, as these data are not updated frequently, insights on potential changes to these values could be supported using advanced analytics that gather these exposure and collateral data, adjust them with any latest financial market data and compute to see if there could be expected losses arising in a forecasted period. This automation in near real time would enable supervisors to focus on what insights are generated with a view to challenging the banks and their risk assumptions.

Automation of capital impact under stressed conditions

An important regulatory metric for supervisors is understanding the capital resilience of banks to withstand unexpected losses and shocks. Metrics that are reported help supervisors to monitor this resilience include capital ratios such as Common Equity Tier 1 (CET1), risk-weighted assets (RWA), probability of default (PD) and exposure at default (EAD) parameters. If market conditions or credit exposures were potentially stressed, our participants indicated that supervisors would want to see if banks were close to breaching regulatory and or capital ratios. Specifically, they would want the extent of the capital impact calculated automatically and for them to be alerted through early warnings.

4.2 Platform functionalities using a storyboard

Based on these discussions and the feedback provided, we set out to develop certain functionalities that could address our use case for the development of the Ellipse prototype, which is based on the assessment of credit risk exposures to large corporates. We adopted two personas, a line supervisor and a macroprudential analyst, to understand how they would use a data and analytics platform and what information they would want to see using the platform. We identified user needs based around the five challenges supervisors typically face. To get a sense of how the platform might look to a user, we sketched out these ideas using storyboards. These were then used to determine how the data architecture would need to be engineered to enable the user needs identified by supervisors.

A key challenge in Phase 2 was reconciling the need to integrate granular data from reporting entities with the largely aggregated data that many supervisors still receive today. To ensure that the use case for Phase 2 could be applied as widely as possible, we took the large exposures regulatory data collection as the proxy source of granular data that could be used for the Ellipse prototype. This information is collected under the Basel Framework and is therefore widely available to supervisors globally. It also has a sufficient level of granularity where exposures to individual corporates could at least be identified at a minimum.

Taking the ideas developed in response to the user needs identified in this use case, the Ellipse team sketched out the possible functionalities and features on a front-end user interface, using a storyboard tool as outlined below. We wanted to replicate as close to what a prototype could look like, in order to give the development team enough clarity on the purpose of the feature and for them to determine how the back end and data architecture would need to be engineered to build these applications.

Reconciling corporate entity names

Functionality 1:

Reconciling corporate entity names to a single format and master list could enable tracking of system-level exposures to the same corporate **(Graph 5)**. If entity names were reconciled to a single master source, the user could:

- Search for a reporting entity via the dashboard search function to see its top exposures by counterparty/corporate
- View the top exposures to corporate counterparties, ordered by size of exposure and linked to the corresponding reporting entity
- Identify corporate credit exposures of regulated entities in a database of largest counterparties/exposures. This would enable the user to reconcile system-level exposures to the same corporate
- Drill down further to find out more information on the corporate entity including its structure, financials and other market data, by clicking on that corporate entity
- Information associated with the regulated entity such as regulatory returns, management reports and financials could also be pulled from data repositories



Market and financial analyses

Functionality 2:

Real-time alerts based on news and market data feeds that are linked to regulatory databases could enable supervisors to investigate that information further or use that information as the basis of undertaking more detailed analysis **(Graph 6)**. If regulatory data sets and unstructured sources of data were linked:

- When logging into the dashboard, the top news alerts could show any current developments related to corporate entities in the large exposures database
- Users could receive early warning alerts on those corporates if text analytics detect any negative sentiment based on news and market feeds
- The early warnings could be ranked or colour-coded by the degree to which the developments have a possible adverse risk
- These early warnings can also be linked to the bank(s) that have exposures to the corporate and the size of exposures
- The early warning alerts could direct the supervisor to investigate the corporate entity, or related information such as the sector or jurisdiction in which that entity operates
- The news analytic tool could be programmed and filtered to provide the supervisor with the following information:
 - News on financial institutions and credit counterparties
 - Sentiment analysis to be applied for certain text, depending on use case, such as insolvency, default, failed, delayed or missed payments, deadlines, bankruptcy, indebtedness, litigation or even changes in senior management roles
 - Repetition of news to be filtered so that only one alert is displayed based on the same topic
 - Hyperlinks to be provided if the supervisor wants to read the articles in whole

• The user could also click on these companies and extract further corporate or financial information, or any changes to market data on these corporate entities

⊞ Pi	roject Ellipse 📕 Home 🗍 Analytics 🖺 News 🕢 Stress Tost 🔷 Settings	Search 👻	
Top regulatory alerts filtered and colour-coded for	Sustainable Bond Grant Sche ne Scompany Misses Bond pay vents to investors.		
relevance	Legman defaults on debit		
	S Co increasing more flights the faster lane agreement.		

Graph 6 — Display of regulatory alerts

Mapping exposures to entities and networks

Functionality 3:

Mapping exposures to corporate entities identified as high-risk or stressed could reduce the burden placed on supervisors to find and link this information manually. If the platform could identify exposures that were high-risk:

- The user could see the top 20 largest exposures to corporates ranked by size and by corresponding bank
- The user could then click on an exposure or the corporate entity to drill down further into its financials or details of any associated credit risk mitigation
- News alerts could also be tailored to pick up corporate entities where text indicates a negative sentiment, allowing supervisors to investigate the alert to see if further actions are warranted
- Text analytics could also enable further linking of news on these corporate entities to the countries and sectors where they may operate
- The user could also visualise these exposures through network mapping, and supervisors could then be able to see interbank contagion risks
- Text analytics that highlight which sector and country the corporate entity operates in could also pick up other corporate entities that operate in the same sectors and countries, which could indicate possible supply chain or related entity risks
- This could be visualised as aggregate exposures to industries, as well as exposure concentrations to corporates in different jurisdictions
- If there are no "direct" exposures to the corporate, the supervisor could use network analytic tools to investigate "indirect" exposures to the investors/ lenders/supply chain of the distressed corporate (Graph 7)
- Using these tools, the exposures could be visualised through networks
- The relative size of the exposures between the banks to the named corporate could be displayed by size of node (ie bigger nodes indicate a bigger exposure, smaller indicating smaller exposure)
- Network analytics could also help to map the broader corporate structure and related entities, if these data are available
- If unavailable, a proxy indicator could be to assume that investors/banks of the same domicile as the corporate entity have exposures to the corporate. The supervisor could then see if local banks have exposures to those banks/ corporates domiciled in that country



Analytics on exposures and current events

Functionality 4:

Analytics that gather data on exposures, collateral and/or provisioning and current financial market data could provide insights into the quality of the large exposures reported **(Graph 8)**. If advanced analytics could be applied to regulatory data and current market data, the user could:

- Look at how exposures could be affected by current events by viewing the exposures on a dashboard by the following breakdown, if available:
 - the total amount of the loan/exposure
 - when more than one loan, a line-by-line breakout of the loans
 - if secured/unsecured
 - currency denomination
 - maturity (only shown with line-by-line view)
- Identify and map collateral/credit risk mitigation made against exposures by counterparty reference or identifier
- If data are available, the user could then drill down further to look at the type and quality of CRM as well as the provider of the CRM
- A search query or even hyperlinks on the name of the corporate could redirect the user from the dashboard to information sourced from market data
- Information sourced from market data could display the corporate's earnings, ratings, downgrades, financials (profit and loss, balance sheet)

- The dashboard could also enable the user to click on a specific sector, where it could then display a breakdown of the largest exposures to the top 10 counterparties, for example
- Analytics could possibly be then applied to link current market data on corporates to exposures



Automating potential capital impact

Functionality 5:

Automating the potential impact to capital and or other regulatory ratios during periods of stress could be provided through early warning alerts to supervisors so that they can investigate the potential impact further **(Graph 9 and 10)**. If the platform could automate impact on a bank's capital and prudential metrics:

- Early warning alerts could be sent to the supervisor if the capital ratios of banks were close to breaching a pre-programmed threshold
- This alert would be based on linking any adverse developments that are picked up in the news on corporates with any relevant large exposures that banks may have to those corporates, and stressing those exposures under certain assumptions (eg changes to interest coverage, probability of default, EBIT, currencies, macroeconomic forecasts and GDP etc)
- The early warning alerts could also potentially be linked to the supervisory rating of the regulated entity
- If the user was concerned about either the quality of the large exposure or the regulated entity, they could also initiate their own stress testing by clicking on different pre-set scenarios or manually input stressed parameters/assumptions to see how these could impact the banks' capital or other metrics of concern

Project Ellipse Hon	ne 🐥 Aler	ts Analytics	🖹 News 🕕 Str	ess Test 🛛 O Settings	Search	
DATE TIME \checkmark	RISK TYPE 🗸		MATCHED ENTI	TY V TEXT EXTRAC	T V	
Today at 1:05 PM	Insolvency 🕅	Fantastic	FANTASTIC HOLDIN CO LTD	GS Hours before Fant rating on Fantastic extremely high risk READ MORE	astic's disclosure, ABC Ratir by four notches to CCC-, r c of default. The global ratin	ngs cut its eflecting an g firm
				IN CASE OF DEFAULT	Simulate	ed stress
BANKING SYSTEM EXPOSURES		Name of Financial Institution	Capital Ratio (Pre-stress)	Capital Ratio (Single-firm stress)	impact on c	apital ratios
SGD 230,000,000		Bank A Ltd	20%	10%	STRESS TEST	
DP 3 BANKS BY EXPOSURE: . BANK A LTD . BANK F LTD		Bank E Ltd	70%	60%	STRESS TEST	
. BANK E LTD		Bank F Ltd	40%	25%	STRESS TEST	
) By % of Loan Book		Total (Weighted)	32%	23%	STRESS TEST	
ELATED CORPORATES	_		11	N CASE OF CONTAGIO	N	
SGD 1,760,000,000	N	ame of Financial stitution	Capital Ratio (Contagion stress)	Exposures to related corporates		
DP 3 RELATED COMPANIES	В	ank A Ltd	5%	Grand Co Ltd SGD 1,665,000	Garden Group Ltd SGD 1,443,000	Green Holding SGD 1,221,00
. GRAND CO LTD . GARDEN GROUP . GREEN HOLDINGS	В	ank E Ltd	45%	Grand Co Ltd SGD 2,665,000	Garden Group Ltd SGD 2,443,000	Green Holding SGD 1,321,00
	В	ank F Ltd	15%	Grand Co Ltd SGD 4,665,000	Garden Group Ltd SGD 2,423,000	Green Holding SGD 2,221,00
	т	otal (Capital Holding)	15%	Grand Co Ltd SGD 8.665.000	Garden Group Ltd SGD 6.443.000	Green Holding SGD 4.221.00

Graph 9 — Integrated risk analytic tools

Graph 10 — Systemic stress test capabilities on the banking system

Project Ellips	se 📕 Home 🌲	Alerts	II Ar	alytics 🖹	News 🕕 S	Stress Test 🔷	Settings		Search		Ť
Г	Balance Sheet Impac	t	Risk	Indicators	Simula	ted stress					
	Search by entity	Q	CET1 Threshold	0.15	impact prudent	on other	a Scer	bles			
	Showing 7 of 158 entities				· ·			_			
	Entity Name 🗢	CE	™ \$	Weighted Probability of Default	÷		Base	Pre	Stressed	Severe	
	✓ Bank A Ltd	0.	15	12%		GDP		0%	-40%	0	40%
	✓ Bank B Ltd	0.	13	12%		Interes	t Rate	0%	-10%	0	10%
	✓ Bank C Ltd	0.	14	12%		USD In	dex	0%	-60%	0	60%
	V Bank F Ltd	0.	12	12%							
	✓ Bank E Ltd	0.	12	12%							
	✓ Bank D Ltd	о.	11	12%							

Other functionalities

Based on our discussions with participants, other functionalities to help support authorities in the use of such an interactive platform could include:

 Modular user authentication, given the sensitivity of supervisory data, to allow only authorised users in the organisation to access the platform and applications specific to the user's access rights

- Easy access to regulatory reports, including the latest returns, management reports, and other documents that relate to the regulated entity (Graph 11)
- Overview access to the credit exposures of regulated entities, showing relative size of exposure to certain corporates or sectors, without detailed access to the underlying data



4.3 Prototype development

The concepts developed using the storyboard provided the basis for building the Ellipse prototype. The prototype was guided by the desire to input both structured regulatory information and unstructured data to a single platform so that the user could:

- easily access information on regulated entities and their counterparties
- be alerted on news or real-time developments that could potentially affect firms and the broader financial system
- understand whether regulated entities were robust to withstand stress and shocks; and
- see if there were indirect contagion risks to regulated entities and their networks

The Ellipse prototype was developed around the following design principles: (i) modularity; (ii) scalability; (iii) deployability; (iv) security; and (v) extensibility.

Modularity: The prototype is built up from separate microservices – each performing a specific function – that are progressively integrated.
 Scalability: By employing container technology, orchestrators can be used to scale up and down based on usage – computing power can hence be allocated to services that most require it to ensure (micro) services run smoothly.

- **Deployability:** Open-source components are used to build up the prototype, so that interested parties can easily adopt / utilise the source codes for their own purposes.
- Security: Significant thought was put into accommodating security protocols given the sensitive nature of supervisory data. Even in the prototype, components can be introduced in a modular fashion, to ensure that the prototype can be integrated with an institution's prevailing authentication systems.
- **Extensibility:** Finally, each of the components are designed for further customisation, recognising that potential users will have slightly different needs based on their structure.

Prototype user features

Login

Starting from the log in screen, the prototype already takes into account the need for the application to be secure, given that supervisory data can be sensitive. Hence, a modular user authentication system was developed that integrates with enterprise systems.¹⁷ This means that only authorised users in an organisation can access the platform **(Graph 12)**.







Home screen

For this prototype, we focused on the type of information that would be relevant to supervisors when assessing corporate credit risk exposures. On the home screen, the user can see an overview on supervised entities from this screen. This includes the top regulatory alerts, news related to supervised entities or their counterparties, an overview of the total corporate exposures from the perspective of the overall banking system as well as individual banks **(Graph 13)**.

¹⁷ For eg, active Directory, which is a directory service developed by Microsoft for Windows domain networks. It is included in most Windows Server operating systems as a set of processes and services.

Graph 13 — Ellipse home screen

Project Elipse Hitters & Nets & Chris Offens	Test 🗗 Young Aller 👹 Analysis		Seech. 0 lag@d
T	o Regulatory Alets • ••••arcs to Sense the same sense the sense were too in the • ••••arcs to Sense to the Sense to the Sense the sense were at the • ••••arcs to Sense to the Sense to the Sense to the Sense to the • •••arcs to Sense to	na accord algage af the investor's process, "School Reargings from School" Registra	
Banking System	total Loss Toposore \$520.56B	Total Banks 42	Tetel Bernstern 10471
Non And All And All And All Annual Tank 1 and as A shall be famile	Toga Exposures		And values of the second secon

Top regulatory alerts

A key feature of this application is to alert supervisors in real time to developments that affect individual firms or the broader financial system. A list of "top regulatory alerts" is shown and are colour-coded based on the frequency of the alert. The red alerts have had multiple mentions of the risk event across news sources while amber and green alerts are less frequent. This colour coding can be customised to fit supervisors' needs. For this prototype, alerts are automatically flagged to show corporates that the supervised banks have exposures to **(Graph 14)**.



The news alerts tab is another feature of this application that supervisors can browse news alerts in real time. Supervisors can also filter entities and risk types and so on **(Graph 15)**.

Graph 15 — News alert search function

📫 Project Ellipse 💠 Home 🏚 Alerts 🔥 Charts 🚯 Stress Test 📑 News Alert 📷 Analytics		Search	θ	Log Out
Date Range Jurisdiction News Agency Industry Entities 🖉 Remove Duplicates				
A \$2 Billion Reminder to Doubt Your Audit Client	By Entities	Filter by		
Bource Prem: Boonterg Today at 7:04 AM			*	
Endless mentioned	By Risk			
			*	
An investment in clean indoor air would do more than neip us tight CUVID – it would neip us concentrate, with lasting benefits Burlef And The Convestion Today at 448 AM	By News Source			
Entities mentioned			٠	
Google	By Date Range			
Can 'plastic credits' help solve the waste crisis? Source from : Con Duriness Todaw # 67 04	mm/dd/yyyy			
Entities mentioned	Jurisdiction			
DANONE SA			٠	
Dry Cargo Supports Polluter Pays Language in EU ETS Amendments Social Prot: Martime Executive Tarlary 2013AN	Clear Apply			0
Endles mentioned.				

Top exposure table

The Home screen provides an overview of system-wide exposures, but it also provides a detailed breakdown on largest exposures by bank, their exposure, collateral and provisions (**Graph 16**).

Graph 16 — Display of largest exposures by bank

		Top 20 Exposures		
Creditor	Borrower	Total Exposure	Secured Exposure	Provisions Made
RACK & XENON BANK	SIGNIFY NV	\$18.328	\$18.328	\$0.119
GANNET & YOKE BANK	ASIATIC GROUP (HOLDINGS) LTD	\$12.108	\$1.218	\$0.005
MARLIN & WOLF BANK	SIA ENGINEERING COMPANY LTD	\$12.028	\$1.318	\$0.453
MARLIN & WOLF BANK	BW LPG LTD	\$11.378	\$8.288	\$0.012
RACK & XENON BANK	JP NELSON HOLDINGS	\$11.198	\$1.128	\$0.711
RACK & KEMP BANK	JAMF HOLDING CORP	\$10.008	\$10.008	50
GANNET & YOKE BANK	LABRADOR IRON ORE ROYALTY CORPORATION	\$10.008	\$10.008	\$0
MARLIN & WOLF BANK	UIPATH INC	\$10.008	\$10.008	50
MARLIN & WOLF BANK	A D S DIAGNOSTIC LTD	\$10.008	\$10.008	50
GANNET & YOKE BANK	CEC ENVIRONMENTAL PROTECTION CO LTD	\$10.008	\$10.008	\$0.005

By clicking on a bank name, the supervisor can get an overview of the bank's top corporate exposures and other relevant information, such as the industry and a breakdown of the loan exposure **(Graph 17)**.

Graph 17 — Overview of bank's top corporate exposures



The table also enables the user to look further into the counterparty, if available. By clicking on one of the corporates, the platform will also provide an overview of the corporate financials in this tab **(Graph 18)**.

Graph 18 — Information on a corporate entity



Automating potential capital impact

The application will flag that a corporate entity to which banks have exposures has been in the news. Here in the Alerts tab, a list of mentions and the entity will be shown in real time with an extract of the text that is most relevant to the banks **(Graph 19)**.

Graph 19 — Real-time display of entitles hadded by risk ty	Graph 19 -	 Real-time 	display of	of entities	flagged	by risk typ
--	------------	-------------------------------	------------	-------------	---------	-------------

1	noject Ellipse III Home	Alerts di Charts 🔘	Stress Test 🔋 News Alert 👩 Analyt	a	e Log Out
					Q. Search X
	Date	Risk Type	Mentioned	Matched Entity	Text Extract
>	02/14/2022	Sustainability	Testa	TESLAINC	The spot speeds up when it makes a pointed callout. No comparing Man*-a not so subtle reference to Testa's Clon Musk and his pursuit of spore registration. It goes on to invoke climate activit Get
>	02/14/2022	Sustainability	Testa	TESLAINC	People though it was a thot at Tenia because its founder and CEO Elice Musk has said he plans to get an orbit to Mark by 2024. The company has also team secured of greenessiting, or neally over setting
>	02/14/2022	Insolvency	K.A.75	ILAFS TRANSPORTATION NETWORKS LIMITED	ILLET's group stacks: ILLET's group would resulter ₹ 55,000 crow in debt by March 2022, the board of the onisis hit company has said in an affaturit find before the National Company Law Applicate Tritoria
>	02/13/2022	Insulvency	Wirecard AG	WIRECARD AG	On 25 June 2020, news of Wirecard AG, headquartered in Munich, Germany, made waves in the financial world when it announced that it would be filing for insolvency.
>	02/13/2022	Insolvency	Wrecard	WIRECARD AG	"It may have been one to three months' worth of salary. If Wirecard in Singapore filed for bankruptcy, it may have been nothing," she said.
>,	02/13/2022	Sustainability	Nasan	NISSAN MOTOR CO LTD	One in three in Southeast Asia are spen to purchasing an electric vehicle (CV) as their next car. They're even ready to pay up to 50 percent more than the price of a conventional car to even an EV,
>	02/13/2022	Sustainability	Debenhams	DEBENHAMS PLC	John Ham's article about Sheffeld and the impact of the closure of Deberhams and John Lewis (The death of the department store, 10 February) raised many questions, not least 'What are obj certer
>	02/13/2022	Insolvency	K. & FS	ILAFS TRANSPORTATION NETWORKS LIMITED	While updating the progress of the resolution progress, the ILMTS band, led by Uday Kotak, said ₹ 55,000 crore debt would be resolved through asset reproductions, restructuring and insolvency proceed
>	02/13/2022	Sustainability	Fujian Haikia Environmental Protection Group	FUUIAN HAXIA ENVIRONMENTAL PROTECTION GROUP CO LTD	Sewage and solid waste processor Fujian Haxia Environmental Protection Group was up 19% last week, while pollution prevention firm GAD Environmental Technology also gained by a similar magnitude.
>	02/12/2022	Insolvency	Hyflun Ltd	HYFLUK LTD	Investment scame socar all over the world and can last far langer than Envy survived, Bernie Madolf ran a giant Panal scheme under the noves of investment and regulators for decades. Skill, the allegat
>	02/11/2022	Insolvency	Logan Group	LOGAN GROUP CO LTD	Concerns over the financial health of Logan Group have shaken investor confidence in what was seen as one of China's stronger developers, deepening the turnoil faced by the crisis hil industry.
>	02/11/2022	Sustainability	FueiCell Energy	FUELCELL ENERGY, INC.	Bloom Energy sells solid solide but cell generators called Bloom Energy Servers which generate electricity from natural gas or blogas via a electrobernical process without combustion. These servers a

If the user wants to find out more about the potential impact on banks, they can expand the alerts. Here, you can see the total banking system exposure to the firm, the individual banks which have exposure and the simulated impact on the Common Equity Tier 1 (CET1) ratio in the event of insolvency. In our synthetic data, only two banks have an exposure to the firm. As this exposure is adequately secured, there is no CET1 impact **(Graph 20)**.

Graph 20 — Estimating impact of corporate default on banks and the banking system

✓ Last Saturday at 11:48 PM Insolvency	FRP	FRP HOLDINGS INC	It appointed administrators from FRF since May 2020.	P, claiming the business was "insolvent" and that it had been in default on loans
BANKING SYSTEM EXPOSURE			IN CASE OF DEFAULT	
SGD 1089000	Name of Financial Institution	CET1 ratio (Pre- stress)	CET1 ratio (Single-firm stress)	
1 Hound & Egg Bank	Hound & Egg Bank	12.81%	12.81%	STRESS TEST
E BOOM O VARIENCE DANK	Loom & Camphor Bank	14.00%	14.08%	STRESS TEST

Further down, the application also pulls out information on related companies, in this case firms from the same country and sector. This is one way that the platform gauges potential contagion risk **(Graph 21)**.

Graph 21 — Estimating potential contagion risk to related companies

Related Corporates			IN CASE OF CONTAGIO	N	
SGD 1904600	Name of Financial Institution	CET1 ratio (Contagion stress)	Exposure to related corporates		
fop 3 COMPANIES BY SECTOR 1. VORNADO REALTY TRUST	Hound & Epg Bank	12.81%	FRP HOLDINGS INC SGD 993600		
2. JONES LANG LASALLE INCORPORATED 3. INVITATION HOMES INC	Ballast & Arrow Bank	24.47%	VORNADO REALTY TRUST SOD 295300		
	Arrow & Quail Bank	18.72%	INVITATION HOMES INC SGD 139100	VORNADO REALTY TRUST SGD 30300	

Simulate stress testing

Another key feature of the application is the ability to drill down to further investigate risk. If the user is concerned by this risk alert and wants to carry out further risk analysis on the bank, they can click on the stress test button. This brings the user to the bank's stress-testing page where they can see the bank's top five exposures by sector and by entity (Graph 22).



Further down, the user will see on the left the different companies that the bank is exposed to and key metrics such the RWA and the PD. On the right, there are a range of scenarios that the supervisor can click on. If the user clicks on the stressed risk analysis button, for example, they can see that the corporate's RWA and probability of default will adjust in response.

They can also click into the risk indicators tab to the right of the balance sheet impact and will be able to see another set of risk indicators that may concern supervisors. In this case the financials of the companies to which the banks are exposed. For example, the EBIT will drop, while interest expense will rise (Graph 23).

Graph 23 — Stress scenarios to estimate possible impact on banks Select a Scenario BALANCE SHEET IMPACT RISK INDICATORS Q Search 40% Entity Name RWA Probability of Default NELNET, INCORPORATION 14,104,804 10% 22.05% -10% 10% Interest Rate 5% HOLDINGS 204,081,389 110% 13.28% 105% -00% 60% USD Index 40% 5.975.771 1 53% 22.05% 1229% VESTMENTS 0.98% 0 TILAK /ENTURES LTD 287,429 155% 17.06% 1203% 5 rows - |< < 1-5 of 408 > >|

		BALANCE SHEET IN	IPACT RISK II	DICATORS			Select a Scenario					
					Q Search	×		BASELINE	STRESSED	SEVERE		
Entity Name	EBIT		Interest Expense		ICR		GOP	-24%	-		47%	
NELNET, INCORPORATION	93,744,137	7 41%	627,713,200	↑ 180%	0.18	↓ 79%			100		1111	
OPPENHEIMER HOLDINGS INCORPORATION	32,245,095	↓ 41%	32,975,800	1 180%	0.98	↓79%	Interest Rate	5%	-		1014	
RAYMOND JAMES FINANCIAL, INC.	221,188,83	8 ¥41%	452,470,400	1 180%	0.48	↓79%	USD Index	40%	40%	•	60%	
SEI INVESTMENTS COMPANY	74,187,151	41%	1,705,200	↑ 180%	43.51	↓ 79%						
TILAK VENTURES LTD	1,845,482	↓ 41%	3,155,040	1 180%	0.58	↓ 79%						-
					5 rows - < <	1-5 of 406 > >I						9

Network analytics

Ellipse can be expanded to provide even more in-depth analysis in this case on potential indirect risk and hidden risk to the banking system. The solution that FNA contributed to the Project Ellipse platform¹⁸ combines advanced analytics of granular loan-level data and sentiment analysis of news data to alert financial authorities in real time to troubled firms and sectors. The individual dashboards provide insights on direct exposures as well as spillovers from indirect contagion channels, such as fire sales, interbank lending and the supply chain impacts.



Fire sale contagion

Enables supervisors to understand about the network structure of common asset holdings and the systemic impact of fire sale contagion stresstest modelling.



Interbank contagion

Advises supervisors about the structure of interbank networks and provides an understanding of the systemic impact of interbank contagion stress-test modelling.



Supply chain contagion

Explores the connection between the production economy and the financial system and enables supervisors to understand the contagion path between the two systems.

Fire sale contagion

This page allows the user to visualise the structure of the common asset holding network between banks in the corporate loan market. More importantly, it enables users to gain an understanding of the systemic impact of fire sale contagion stress-test modelling in the event of the failure of a corporate entity. The network visualisation in the dashboard is automatically constructed from corporate loan data. From left to right, users can observe how the shock spreads throughout the system, impacting even organisations indirectly exposed to the corporate used in the example for illustrative purposes only **(Graph 24)**.

¹⁸ The prototype of the Ellipse platform has been built for the sole purpose of investigating the technological and practical feasibility of particular designs and should not be understood to imply or express any regulatory policy stance or endorsement by the BIS or any other party of the solutions provided during this project.

Graph 24 — Display of network effects through fire sale contagion



A Fire sale losses bar chart

This bar chart shows total systemic losses due to fire sale stress test in the case of the specimen corporate failure. It combines direct losses for the banks directly exposed to the corporate in the loan market and second-round losses due to contagion in the network of common asset holdings.

B Portfolio similarity table

This widget displays the portfolio similarity (common asset holding) value of every pair of banks in the corporate loan market.

C Common asset holding network

The red node in the centre left represents the example corporate. The orange nodes represent banks with direct exposure to the corporate. The blue nodes are other firms in the corporate loan market that borrow some loans from these banks. The yellow nodes represent banks with common asset holdings to orange nodes, in terms of the types of institution that they are exposed to. The size of the node corresponds to the amount of fire sale losses that the bank will suffer, in which larger nodes correspond to higher losses.

D Trends widget

This widget displays trends in the portfolio similarity and diversification value of banks.

Interbank loan contagion

This page informs users on the structure of the interbank network and the possible systemic impact of interbank contagion stress-test modelling in the event of the hypothetical failure of the corporate provided for illustrative purposes only. The network visualisation in the dashboard is automatically constructed from the interbank loan data. It details further potential channels of contagion in the interbank lending market through the interbank lending relationships of the banks. The bar chart on the left summarises the potential losses of each bank in a stress simulation scenario. The application automatically filters the institutions that have been most severely affected by this contagion channel **(Graph 25)**.

Graph 25 — Display of network effects through interbank loan contagion



A Interbank loses bar chart

This bar chart shows the total systemic losses of interbank contagion stress test in the case of a hypothetical corporate failure. It combines direct losses for the banks directly exposed to the corporate in the corporate loan market and second-round losses due to contagion in the interbank market.

B Trends

This widget displays trends in vulnerabilities or systemic losses for the banks.

C Scatter plot

This widget shows the relationships between two node properties, for example between capital and vulnerability of banks.

Interbank network

The red node in the top left corner represents the corporate used for illustrative purposes only. The orange nodes represent banks with direct exposure to the corporate in the corporate loan market. These banks are connected to other banks in the interbank market, which are denoted by orange colour nodes. The size of the node corresponds to the amount of interbank losses that the bank will suffer, in which larger nodes correspond to higher losses.

Supply chain contagion

The supply chain contagion page explores the connection between the production economy and the financial system. Corporations are embedded in a complex web of supplier and buyer relationships. At the same time they also act as borrowers from banks. As a result, their two-layered role makes it possible for shocks in the supply chain to result in a contagion event in the financial system **(Graph 26)**.

Graph 26 — Display of network effects through supply chain contagion



A Fire sale losses bar chart

This bar chart shows total systemic losses due to fire sale stress test in the case of the specimen corporate failure. It combines direct losses for the banks directly exposed to the corporate in the loan market and second-round losses due to contagion in the network of common asset holdings.

B Common asset holding network

The red node in the centre left represents the example corporate. The orange nodes represent banks with direct exposure to the corporate. The blue nodes are other firms in the corporate loan market that borrow some loans from these banks. The yellow nodes represent banks with common asset holdings to orange nodes, in terms of the types of institution that they are exposed to. The size of the node corresponds to the amount of fire sale losses that the bank will suffer, in which larger nodes correspond to higher losses.

C Portfolio similarity table

This widget displays the portfolio similarity (common asset holding) value of every pair of banks in the corporate loan market.

4.4 High-level architecture for the Ellipse platform

The Ellipse platform comprises a front-end user interface (coded using the React framework) which is backed by a suite of back-end components/microservices (coded in Python, using the Django web framework), each integrated in a modular fashion. Each component is packaged as a Docker image, and the resulting containers can be deployed either using Docker runtime or via the Kubernetes orchestration platform. The key back-end components are (i) data connectors; (ii) text analysis and entity resolution modules that help to populate an underlying metadata model; (iii) a search engine for finding and retrieving information; (iv) authentication and authorisation protocols; and (v) advanced analytics (stress testing, risk alerts) modules. In addition, network analytics are provided by FNA using natural language processing and sentiment analysis.

Data connectors

Data connectors are designed to connect to three types of data that supervisors seek to integrate: (i) structured data from third-party providers (eg market data providers etc); (ii) unstructured data from third-party providers (eg news aggregators, or ratings research); and (iii) structured data held in internal databases (eg regulatory reports submitted by reporting entities). Depending on the nature of the various internal/third-party endpoints, data are either loaded into a queue (for endpoints where data streaming is possible) or extracted based on scheduled jobs.

Text analysis

Via a natural language classification script,¹⁹ AI is applied to the news/unstructured data to identify relevant text extracts and bucket them into possible risk taxonomies (namely, insolvency and sustainability risks), which are iteratively introduced into a metadata model that underpin the prototype. ML is then applied, via named entity recognition²⁰ algorithms, which help to match the entities to those in the above metadata model.

Data sources

The prototype used synthetic bank exposure data. The synthetic data set is generated based on sample statistics of actual transaction-level data set on corporate loans offered by banks. The characteristics of the distribution are then used to impute loan amounts to corporate entities based on various corporate attributes such as country of domicile and business sector, in line with published statistics on the distribution of loans to these country/sectors. Other sources of data used include corporate data from actual companies that are extracted from a data vendor and news and market data APIs, where a real-time feed from a news aggregator API is used.

Combining structured and unstructured data

To combine the various structured and unstructured data input into the prototype, components (that analyse/interpret the text and extract metadata) process the data based on a common (yet customisable) metadata model. Metadata extracted include uniquely identifying the entities associated with each unit of information (ie news article, data set), which can then be matched against the synthetic bank exposure data or the corporate data to match risks identified from newsfeed to banking system exposures, and market data of the entity in question.

¹⁹ The purpose of this classification script is to draw on key phrases or terms to identify paragraphs in news articles that could indicate a corporate facing potential insolvency or sustainability risks.

²⁰ Named entity recognition seeks to locate and classify entities mentioned in unstructured text, into categories such as companies. To achieve high accuracy rates, ML algorithms are applied to convert the unstructured text into numeric vectors. For instance, the same word "Apple" in "I ate an apple", and "Apple reported its earnings" would be represented as different numeric vectors, as the ML algorithm has been trained to treat the word differently in different contexts.

Risk alerts

Based on the above AI and ML algorithms for text analysis and named entity recognition, alerts are generated for corporates to which the supervised banks have exposures.

Stress testing exposures

When the application highlights an entity with text extracts that suggest risk events, the entity is flagged. Supervisors would then investigate the possible impact that the entity could have on banks or the banking system as a whole. By matching against the exposures data, the prototype reveals all banks with exposures to the entity, as well as the size of the exposures. The simulated impact on the bank's CET1 ratio in the event of the entity's insolvency (ie the probability of default set to 1) is also reported, taking into account offsets such as credit risk mitigation strategies (collateral). To assess contagion, the application also identifies corporates that are similar to the entity identified (ie from the same country-sector), and the aggregate impact of the failure of the entity and these similar firms are also reported.

The application also contains a stress-testing module at the bank level. Users can adjust a set of macro-financial variables (eg GDP, interest rate, exchange rate) to generate interest rate and earnings shocks to corporates. These in turn impact the interest coverage ratios (ICR) of corporates (based on a simple linear model), which can be mapped to the probability of defaults (via an ICR-credit spreads model) and the RWAs of banks with exposures to these entities. Specifically:

- the percentage change in the ICR is approximated by the percentage change in earnings (measured by earnings before interest and tax) less the percentage change in interest expense. The change in earnings is assumed to be directly affected by the deviation in GDP growth relative to the baseline and inversely affected by an appreciation in a broad USD index. Interest expense rises in concert with an increase in local interest rate, which is further exacerbated by tightening financial conditions associated with broad dollar strength.
- a mapping table²¹ is used to map ICRs to credit spreads, which are used to imply a risk-neutral probability of default that then is used to calculate the RWA based on an assumed loss-given-default (LGD). For counterparties that are under stress but do not default (PD < 1), the impact on CET1 ratio occurs through increases in the RWA for simplicity (ie a denominator increase).
- when assessing the impact of a default (PD = 1) on CET1 capital, a simple formula (max (0, LGD*Exposure At Default (EAD) – amount secured) is used, where the EAD is assumed to be the actual exposure of the bank to the counterparty. This amount is subtracted from the CET1 capital amount to determine the impact of stress on CET1 capital.

²¹ See eg work by Aswath Damodaran at https://people.stern.nyu.edu/adamodar/New_Home_Page/datafile?ratings.html.

Network analytics

In the dashboards provided by FNA, natural language processing and sentiment analysis are used to identify in real time emerging credit risk indicators from news data combined with the network analytics of individual loan data in order to assess direct and indirect contagion channels of credit risks. The solution combines advanced analytics of natural language processing and network analytics, and operationalises different types of both structured and unstructured data, such as news, corporate loans, interbank loans and supply chains.

Data attributes

The input data can be divided into nodes and links. Nodes represent non-relational inputs such as news and entities in our context. Links represent all the relationships in the data sets such as transactions or loans. The properties of both the nodes and links can be found in **Annex 2**.

Fire sale contagion dynamics

The fire sale contagion model is based on the method that has been introduced and widely used in the systemic risk literature.²² The steps of the fire sale contagion model can be summarised in **Annex 3**.

Interbank contagion dynamics

To assess the systemic impact of interbank stress test contagion, we use DebtRank.²³ DebtRank is an iterative algorithm which assumes that stresses in an interbank network propagate due to the credit quality deterioration of the lender's counterparty. First, banks with direct exposure to the specimen corporate will suffer some losses. This causes changes in market perceptions about the creditworthiness of these directly impacted banks in the interbank system. The lenders of these banks will then suffer some reduction in their interbank lending assets as their interbank loans are marked to market.

Technically speaking, the contagion dynamics in the interbank market can be defined as follows:

$$h_i(t) = \min\left\{1, h_i(t-1) + \sum_j \quad W_{ij}h_j(t-1)\sigma_j(t)\right\}$$

 $h_i(t) \leftarrow$ level of distress

 $W_{ii} \leftarrow \text{impact of bank} - j \text{ on bank} - i$

 $\sigma_i(t) \leftarrow \text{only active banks propagate shocks}$

²² eg Caccioli et al (2014), Greenwood et al (2015) and Ramadiah et al (2020).

²³ Battiston et al (2012).

Operational considerations

One key operational consideration is the security of the application/prototype, given that supervisory data can be sensitive. To address this, a modular user authentication system has been developed that can be integrated with enterprise systems such as Active Directory. Only authorised users can access the platform.



Findings, opportunities and considerations for supervisors

Findings, opportunities and considerations for supervisors

5.1 Findings from the Ellipse project

Project Ellipse set out to explore and demonstrate how an integrated data and analytics platform solution could enable supervisors to digitally extract, access and analyse large and diverse sources of data that are relevant to current events in real time. It looked at whether the collection of granular data from reporting entities could replace the need for authorities to request information using templates. This would require data standards and common data models, where supervisory metrics could be derived using programmable rules that reference machine-readable and machine-executable common data models. The project also set out to demonstrate how real-time insights using advanced analytics could be derived from large volumes of structured and unstructured data that could be used to make risk correlations and analyse sentiment, alerting supervisors of issues that may need further investigation. It also sought to display these insights using an integrated end-to-end data pipeline where both structured and unstructured data sources are input, consolidated, analysed and displayed using a customisable dashboard user interface.

The first phase of Project Ellipse was an exploratory PoC looking at the feasibility of cross-border digital reporting. This PoC is intended as a first step towards bringing authorities and stakeholders closer to a common understanding of data that are collected by authorities globally. It was intentionally limited in scope, but an important aim was to see if, for the specific use case, our stakeholders could come to a consistent understanding of the data attributes of retail mortgages and if these data attributes could be represented in a way that allows programmable code to reference them to generate the reporting of regulatory metrics. Our exploration confirmed that:

- regulatory reporting requirements can be expressed in unambiguous machinereadable logical reporting instructions underpinned by a consistent data model
- technical standardised programmatic specifications of the steps for generating regulatory reports can be published alongside regulation and ensure clear understanding at the most granular level of the required data
- executable libraries can be automatically generated and published alongside regulations to assist accelerated implementation
- if a common data standard was agreed to and implemented, financial institutions may no longer need to interpret reporting instructions and submit aggregated data by use case
- with additional logical instruction based on the same data model, supervisors may also be able to automatically query the underlying transaction data and generate regulatory metrics referencing that standardised data

The second phase of Project Ellipse explored how a single platform could be built so that authorities could benefit from "on demand" access to timely and integrated sources of data to help support and inform their supervisory assessments. An important aim of Phase 2 was to show that, as authorities continue to explore digital reporting and the collection of granular data, they could in the meantime use existing regulatory data on the Ellipse platform, such as data collected on large exposures. We also wanted to show that a single platform could be capable of acting as a "one stop shop", where supervisors could find regulatory information quickly and could be guided by insights generated from the analytics running on the platform. Even though the project build was limited to six months, the Ellipse prototype demonstrates that:

- regulatory data that have sufficient granularity, such as large exposures, can be integrated on the same platform with other sources of unstructured information such as news and market data
- the use of advanced analytics can automate the assessment of possible impacts to key prudential metrics, providing supervisors with real-time insights and early warnings of the at-risk exposures of reporting entities
- if supervisors had this information automated in real time, they would be able to investigate the risks further and challenge reporting entities faster, allowing earlier intervention if needed
- functionalities that meet the needs of different units with different supervisory or oversight responsibilities can be built into a single platform, with authentication systems allowing access only to authorised users
- the platform has been built to accommodate multiple applications and therefore has the potential to scale up with larger and more granular data sets

5.2 Opportunities and considerations for supervisors

With these findings in mind, Project Ellipse presents an opportunity for the global regulatory community to further consider, explore and collaborate on common solutions that can future-proof the data and analytical capabilities of supervisors. Our stakeholder engagement with regulatory authorities and international institutions globally over the last 12 months confirms the desire by supervisors to move towards more digitally driven solutions. The greater the number of authorities that converge on a common solution, the greater the potential for convergence on common data inputs, data standardisation and information-sharing to support financial stability.

Digital transformation raises important considerations for authorities that were not within the scope of this project if they wish to explore platform solutions like Ellipse. These considerations include senior management sponsorship and organisational change management strategies to enable a shift towards digitally driven solutions. Such solutions ultimately require resources and expertise, as well as broader training and development of supervisory staff. The use of such platforms nevertheless might meet with scepticism, raising questions such as whether it should be the responsibility of supervisors to be forward-leaning and proactive in dealing with emerging risks or whether this responsibility should remain solely with the supervised entities.

It is important to stress that augmenting the supervisory toolkit with digitally enabled analytics does not replace the need for supervisory assessment and judgment. On the contrary, supervisory instinct will become even more crucial when presented with predictive insights, as a thorough knowledge of supervised institutions will be key to determining if early action is needed. However, without new solutions that harness the power of data, supervisors will spend too much of their time making sense of the disparate data instead of using their expertise and judgment to challenge supervised entities with this information. Authorities are therefore encouraged to start this dialogue within their respective organisations, if they have not done so already.

The BIS Innovation Hub's Project Ellipse is a prototype that we hope will help authorities to explore new solutions, and to investigate for themselves the art of what could be possible. To this end, the Ellipse platform is being developed as a global public good to be shared with authorities globally. Our intention is to use the platform as a basis from which to create a collaboration community so that this solution has the potential to be shared and scaled across regulatory authorities around the world. By joining forces through a collaborative community, central banks and regulatory authorities could work together to create new applications to serve common use cases and priorities, and to reduce duplication of effort. Ultimately by working together, authorities can scale up these solutions faster so that collectively they can prepare themselves better to meet the challenges of the digital age.

Annex 1 Data definitions

	Loan
Attribute	Ellipse PoC definition
Transaction ID	Unique identifier (ID) for each loan: the loan ID should not change throughout the life of the transaction. If the original loan ID cannot be maintained in future, then in this field enter the original ID followed by the new ID.
Purpose	The reason for taking out the mortgage loan, eg to buy a first home, buy another home, remortgage from another lender, etc.
Loan size	The original balance when the mortgage was completed. Amount of loan approved (and currency of loan). In the case of loans where the amount advanced is less than the total amount of the loan which the firm has agreed to lend, this should be the amount of the committed advance (including any committed drawing facilities).
Main/first borrower ID	Unique identifier (ID) per borrower (not showing the real name) to identify the point of contact amongst the borrowers or the largest borrower – in terms of liabilities – in a group of borrowers. Should not change over the life of the transaction. If there is more than one borrower then list the borrower IDs with comma delimitation with primary borrower first.
Date mortgage matures based on contract date	The date a loan contract is due to mature if it goes to full term. (This can then be used to calculate the remaining loan tenure at any point in time.)
Initial rate	The rate of interest reported should be the initial gross nominal rate charged on the loan and should take into account any discount being provided.
Reversionary interest rate	Type of interest rate at the end of initial incentivised rate period. Could be fixed rate, discount, capped rate, standard variable rate, base rate tracker, reference rate tracker, other tracker, other or not applicable.
Date of first drawdown	Date on which the initial tranche or complete amount of the mortgage loan is transferred from the lender to the borrower or their agent.

Collateral			
Attribute	Ellipse PoC definition		
Collateral ID	Unique identifier per property to enable properties with multiple loans in the pool to be identified (eg further advances/second liens are shown as separate entries).		

Country where collateral property is located	The ISO code of the country where the land/property which is subject to mortgage is located.
Property type	For PoC, limited to private residential, which refers to privately owned property for occupation of at least 40% as a dwelling by the borrower(s) and/or dependent(s).
Dwelling type	One option from list below per mortgaged property. A dwelling is a self-contained room or suite of rooms, including cooking and bathing facilities, intended for long-term residential use. A dwelling is private (not generally accessible by the public) and is contained within a building that is an immobile structure. A dwelling may comprise part of a building or the whole of a building. Regardless of whether they are self-contained or not, rooms within buildings offering institutional care (eg hospitals) or temporary accommodation (eg hotels, motels or hostels) are not defined as dwellings.
Bungalow	A single-storey dwelling, usually detached.
Detached	A standalone residential structure of more than one storey that does not share any outside walls with another residential property.
Semi-detached	A residential dwelling of more than one storey in height attached to another building or dwelling by one common party wall
Terraced	A dwelling of more than one storey attached to another house on two sides forming part of a row of similar houses each with its own frontage to a road.
Apartment/flat /condominium	An apartment/flat/condominium is a separate and self- contained premise constructed or adapted for use for residential purposes and forming part of a building from some other part of which it is divided horizontally. Flats have to be contained within a dwelling with at least two storeys. These dwellings do not have their own private grounds and usually share a common entrance foyer or stairwell. Does not include duplexes, townhouses or a detached residence that includes a flat (such as a granny flat) on the same property. Also includes maisonettes - in the case of the United Kingdom - which refers to a set of rooms for living in typically on two storeys as part of a larger building with a separate entrance from the rest of the building. Can be one level or split level. Includes both purpose-built and converted buildings.

Other value of property	Any other type of residential dwelling property. The market value of the property which is subject to the mortgage. The value reported should be based on the surveyor's valuation, a valuation index, or other method that the product provider used to determine the market value. [In the case of staged construction or self-build schemes, value means "expected final value of property at the time the building decision is made".
Purchase price of property	The purchase price of the land/buildings which are the subject of the mortgage, as stated on the mortgage application.
Postal code	Format will be country-specific. In the United Kingdom, report the post code of the mortgaged property. For new build/self-build properties only, firms may report only the first half of the postcode, eg XY45, if the full postcode has not yet been assigned. For all other properties, the full postcode of the property must be reported, eg XY45 6XX.

Borrower				
Attribute	Ellipse PoC definition			
Borrower ID	Unique identifier (ID) per borrower (not showing the real name) to allow borrowers with multiple loans in the pool to be identified (eg further advances/second liens are shown as separate entries). Should not change over the life of the transaction. If there is more than one borrower, then list the borrower IDs and use a comma as a delimiter, with primary borrower first.			
Borrower type	An individual or natural person (may be more than one) – in the United Kingdom, can also have a trustee but this is out of scope for the PoC. An entity/firm/corporate is also out of scope for the PoC.			
Employment status	One option from the list below per individual. Where the borrower has more than one employment status, report status that makes up largest portion of verified income.			
Employed	An individual who works under an employment contract. They are usually required to work regularly for a minimum number of hours and expect to be paid for time worked. A manager or supervisor is responsible for their workload, defining when a piece of work should be finished and how it should be done. Work is done at the business premises or at an address specified by the business. The business provides the materials, tools and equipment for their work.			

Self-employed	An individual is self-employed if they run their business for themselves and take responsibility for its success or failure. They can decide what work they do and when, where and how to do it and can work for more than one client.
Retired	An individual who has reached the official retirement age or other specified age and is usually no longer undertaking paid employment.
Other	Examples include company directors.
Income type	A borrower may have income of more than one type.
from employment	Gross annual income from employment before tax or other deductions. This should be evidenced, either document based (eg payslips, bank statements or tax returns) or derived through the use of automated systems.
from self- employment	Gross annual income from self-employment before tax or other deductions. This should be evidenced, either document based (eg bank statements or tax returns) or derived through the use of automated systems.
other	Income from sources other than employment such as pensions or investments
Income amount	 Gross monthly income, in relation to a borrower, shall be the aggregate of: (a) in the case where the borrower has a fixed monthly income only, his monthly income (in the case of Singapore, excluding any contributions made to the Central Provident Fund account of the borrower by the borrower's employer, where applicable) at the time of applying for a credit facility or a financing; (b) in the case where the borrower has a variable income only, such as commission, bonus or allowance from his employer (i) the average of the monthly variable income earned in the preceding 12 months; or (ii) the employment income reflected in the latest available assessment of taxable income from the tax authority at the time of applying for a credit facility or a refinancing; (c) in the case where the borrower has a fixed and variable monthly income (i) the aggregate of his fixed monthly income as determined in accordance with subparagraph (a) above and his variable monthly income as determined in accordance with subparagraph (b)(ii) above; or

	(ii) the aggregate of his fixed employment income and his variable employment income reflected in the latest available assessment at the time of applying for a credit facility or a refinancing;
	(d) the monthly rental income received by the borrower, if any (in the case of Singapore); and
	(e) the value of the eligible financial assets of the borrower, if any (in the case of Singapore).
Currency of denomination	GBP = United Kingdom pound; EUR = euro; USD = US dollars; JPY = Japanese yen; OTH = other. If more than one applies, report the currency that applies to the largest proportion of the mortgage.
Income frequency	The frequency in which gross basic pay from employment (whether from one or more jobs), gross income from self- employment insurance or pensions and state benefits is paid.
Type of debt	Report only where the borrower is consolidating debt into the new mortgage. Type of debt shall be anything where the borrower is obliged to make payments in the future.
Debt servicing amount	Refers to the amount of outflows per period that the borrower is subject too at the point of applying for a credit facility or a refinancing.
Frequency of payments	The frequency in which the bank is deducting money from the bank account provided to serve the mortgage loan.

Annex 2 Data attributes

Unstructured data

	News
'IES	Headline
ERT	Full article
ROP	Source
DE P	Date
NOL	Time
	Language

Structured data

	Corporate loans	Interbank loans	Supply Chain
LINK PROPERTIES	Lender ID	Lender ID	Supplier ID
	Borrower ID	Borrower ID	Buyer ID
	Annual interest	Outstanding exposures	Transaction values
	Duration (months)	Interest rate	Transaction number
	Outstanding exposures	Reference rate	Transaction values on trade credits
	Outstanding exposures (exposed	Maturity rate	Transaction period
	Loan quality	Currency	Currency
	Maturity date		
	Currency		

NODE PROPERTIES

Banks	Firms
Capital	Sector
Total liabilities	Industry
Total assets	Total corporate loans
Total corporate loans	EBITDA
Total interbank loans	Supply chain purchases
Total interbank liabilities	Supply chain outputs sold to other firms
Location	Location
	Annual turnover



Annex 3 Fire sale contagion dynamics

Glossary of terms

Advanced analytics	The autonomous or semi-autonomous examination of data or content using sophisticated techniques and tools to discover deeper insights, make predictions, or generate recommendations. Advanced analytic techniques include those such as data/text-mining, machine learning, pattern- matching, forecasting, visualisation, semantic analysis, sentiment analysis, network and cluster analysis, multivariate statistics, graph analysis, simulation, complex event processing and neural networks.
Aggregate data	Data, typically numerical, that are collected from multiple sources and/or on multiple measures, variables and combined into data summaries, typically for the purposes of public reporting or statistical analysis.
Artificial intelligence	The theory and development of computer systems able to perform tasks that traditionally have required human intelligence.
Data attribute	A data attribute is a single field representing a certain feature, characteristic, or dimensions of a data object.
Data model	Refers to a semantic data model that is a method of organising data which reflects the basic meaning of data items and the relationships between them. This organisation makes it easier to develop application programs and to maintain the consistency of data when they are updated.
Data standards	Data standards are the rules by which data are described and recorded in a consistent way. In order to share, exchange, and understand data, the format and meaning must be standardised.
Exposures	In finance, exposure refers to the amount of money that an investor has invested in a particular asset and also represents the amount of money that the investor could potentially lose on an investment.
Granular data	The data points and data formats that firms use in their internal books and records for financial and business purposes. Data defined at lowest appropriate level possible for a given data set, for example relating to individual contracts or transactions.
Machine learning	A method of designing a sequence of actions to solve a problem that optimises automatically through experience and with limited or without human intervention.

Glossary of terms

Open source	Open source describes software that comes with permission to use, copy and distribute, either as is or with modifications, and that may be offered either free or for a fee. The source code must be made available.
Predictive analytics	The use and examination of data to predict patterns of activity. Predictive analytics may involve technologies such as machine learning or visualisation tools and is characterised by techniques such as regression analysis, forecasting, multivariate statistics, pattern-matching, predictive modelling and forecasting.
Proof of concept	A proof of concept (PoC) is a demonstration of a product, service or solution in a sales context. A PoC should demonstrate that the product or concept will fulfil customer requirements while also providing a compelling business case for adoption.
Reference implementation	For this PoC, we refer to reference implementation as a point of reference rather than being put directly into productive use.
Regulatory reporting	Data received to fulfil mandated functions such as those of a supervisor, regulator, macroprudential authority or resolution authority.
Reporting requirements	The description of which firms need to provide data, what data they need to provide, how they need to provide it and when they need to provide it. Requirements may include rules, instructions and technical specifications.
Structured data	Information that has a predefined data model or is organised in a predefined manner.
Suptech	Any application of fintech used by regulatory, supervisory and oversight authorities.
Technology stack	A list of all the technology services used to build and run a given application.
Unstructured data	Information that either has no predefined data model or is not organised in a predefined manner.

References

Bank of England (2022): Transforming data collection, updated 22 February, www.bankofengland. co.uk/prudential-regulation/transforming-data-collection.

Basel Committee on Banking Supervision (2012): Core principles of effective banking supervision, September, www.bis.org/publ/bcbs230.htm.

Battiston, S, M Puliga, R Kaushik, P Tasca and G Caldarelli (2012): "DebtRank: Too Central to Fail? Financial Networks, the FED and Systemic Risk", *Scientific Reports*, vol 2, no 541.

Caccioli, F, M Shrestha, C Moore and J Farmer (2014): "Stability analysis of financial contagion due to overlapping portfolios", *Journal of Banking & Finance*, vol 46, pp 233–45.

Castri, S, S Hohl, A Kulenkampff and J Prenio (2019): "The suptech generations", *FSI Insights*, no 19, October.

Crisanto, J, K Kienecker, J Prenio and E Tan (2020): "From data reporting to data-sharing: how far can suptech and other innovations challenge the status quo of regulatory reporting?", *FSI Insights*, no 29, December.

European Central Bank (2022): *Banks* Integrated Reporting Dictionary (BIRD), www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/reporting/html/bird_dedicated.en.html.

------- (2022): AnaCredit, www.ecb.europa.eu/stats/money_credit_banking/anacredit/html/index. en.html.

Financial Conduct Authority (2022): *Digital regulatory reporting*, www.fca.org.uk/innovation/regtech/ digital-regulatory-reporting.

Financial Stability Board (2020): The Use of Supervisory and Regulatory Technology by Authorities and Regulated Institutions: Market developments and financial stability implications, 9 October, www.fsb.org/wp-content/uploads/P091020.pdf.

Greenwood, R, A Landier and D Thesmar (2015): "Vulnerable banks", *Journal of Financial Economics*, vol 115, no 3, pp 471–85.

International Swaps and Derivatives Association (2022): Common Domain Model, https://docs.rosetta-technology.io/cdm/readme.html.

Kienecker, K, G Sedlacek and J Turner (2018): *Managing the processing chain from banks* source data to statistical and regulatory reports in Austria, www.oenb.at/dam/jcr:d9cdbe0a-a6d4-409a-8ac5-670cad2619b0/05_Kienecker_Statistiken_3_18.pdf.

PUMA (2022): Unified Business Matrix Procedure, www.cooperazionepuma.org/.

Ramadiah, A, D Fricke and F Caccioli (2022): "Backtesting macroprudential stress tests", *Journal of Economic Dynamics and Control*, (forthcoming).

Toronto Centre (2018): SupTech: *Leveraging Technology for Better Supervision*, July, https://res. torontocentre.org/guidedocs/SupTech%20%20Leveraging%20Technology%20for%20Better%20 Supervision%20FINAL.pdf.

World Government Summit (2018): *Regtech for Regulators: Re-architect the system for better regulation*, February, www.worldgovernmentsummit.org/api/publications/document?id=5ccf8ac4-e97c-6578-b2f8-ff0000a7ddb6.



Bank for International Settlements (BIS)

ISBN 978-92-9259-541-8 (online)