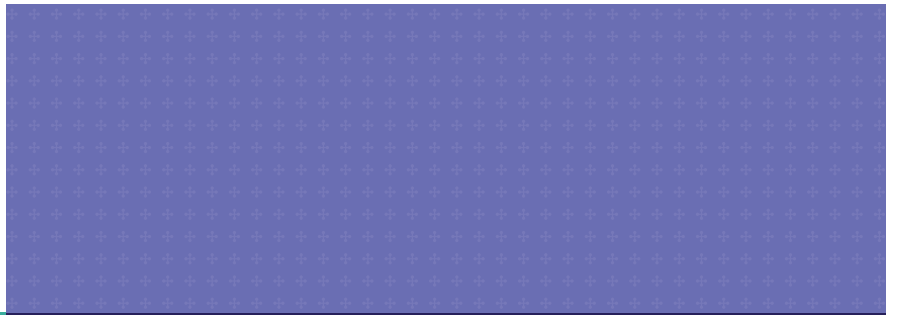




Discussion Paper

Transforming data collection from the UK financial sector

January 2020



Discussion paper | Bank of England

Transforming data collection from the UK financial sector

January 2020

By responding to this discussion paper, you provide personal data to the Bank of England. This may include your name, contact details (including, if provided, details of the organisation you work for), and opinions or details offered in the response itself.

The response will be assessed to inform our work as a regulator and central bank, both in the public interest and in the exercise of our official authority. We may use your details to contact you to clarify any aspects of your response.

The discussion paper will explain if responses will be shared with other organisations (for example, the Financial Conduct Authority). If this is the case, the other organisation will also review the responses and may also contact you to clarify aspects of your response. We will retain all responses for the period that is relevant to supporting ongoing regulatory policy developments and reviews. However, all personal data will be redacted from the responses within five years of receipt. To find out more about how we deal with your personal data, your rights or to get in touch please visit bankofengland.co.uk/legal/privacy.

Information provided in response to this consultation, including personal information, may be subject to publication or disclosure to other parties in accordance with access to information regimes including under the Freedom of Information Act 2000 or data protection legislation, or as otherwise required by law or in discharge of the Bank's functions.

Please indicate if you regard all, or some of, the information you provide as confidential. If the Bank of England receives a request for disclosure of this information, we will take your indication(s) into account, but cannot give an assurance that confidentiality can be maintained in all circumstances. An automatic confidentiality disclaimer generated by your IT system on emails will not, of itself, be regarded as binding on the Bank of England.

Responses are requested by 7 April 2020.

Please address any comments or enquiries to:

Angus Moir

Bank of England

Threadneedle Street London,

EC2R 8AH

Email: DatacollectionDP@bankofengland.co.uk

Foreword

Financial firms depend critically on good information. Fuller, faster, more accurate and more insightful information can make the difference in how competitive they are. As a result, we are now seeing significant investment in new technology to improve the analysis and storage of large volumes of data.

The Bank of England also depends on good and timely information. We cannot fulfil our objectives of monetary and financial stability without an accurate picture of the economy, the financial system and firms we regulate. And, just like those regulated firms, the Bank too has ambitious plans to use new technologies to improve the collection and analysis of data, including from its regulatory returns.

Earlier this year, in its response to the Van Steenis “Future of Finance” report, the Bank committed itself to the development of a “world-class regtech and data strategy”. The ultimate aim is to make data collection significantly more efficient for firms while at the same time improving the Bank’s ability to use what we do collect more effectively.

Because it determines what information is required of regulated firms, the Bank has a significant influence on their data strategies and plays an important part in shaping how firms approach their own data. There is therefore the potential to make improvements that also support firms’ own use of data, making them more productive and competitive.

This paper seeks to lay the foundations for joint work to reform data collection over the next decade. The long timescale recognises that this will be a substantial challenge, and that we will need time to identify and implement the right solutions. It is also designed to encourage us to explore more radical changes, without the constraints that come with shorter timescales.

This needs to be an inclusive exercise, which looks across the full range of firms and collections. We want to hear from anyone with an interest, not just the largest banks and insurers. There will not be a one-size-fits-all solution, so it is vital that we build a full understanding of the diversity of experiences with data collection across the financial system.

We also know that data collection by the Bank represents just one aspect of the reporting challenge that firms face. So we are committed to identifying where there are overlaps and opportunities with other authorities in the UK and internationally, and pursuing closer collaboration as we all respond to the opportunities offered by the new data landscape. For example, we are working closely with the FCA on data initiatives such as the pilot on digital regulatory reporting. The Bank and the FCA will continue to work closely together to ensure that reforms to data collection are aligned.

We are pleased to launch this discussion paper with the aim of improving the efficiency and effectiveness of the Bank’s data collections, and supporting wider developments in data use.



Ben Broadbent

**Deputy Governor, Monetary Policy
Bank of England**

Sam Woods

**Deputy Governor, Prudential Regulation
and Chief Executive of the Prudential
Regulatory Authority**



Contents

Executive summary	1
Questions	6
1 Scope and structure of the discussion paper	7
2 The Bank of England’s collection of data from firms	9
3 Costs to firms	15
4 Challenges for users	22
5 What might a new reporting approach look like?	25
6 Common data inputs	30
7 Modernising reporting instructions	36
8 Reporting architecture and governance	42
9 Next steps	46
Appendices	48

Executive summary

Background and aims

- i. The Bank of England’s mission of maintaining monetary and financial stability depends on access to high quality data.¹ Since the financial crisis, the range and volume of data collection have grown, driven in particular by new regulations and improved monitoring of micro- and macro-prudential risks. This is a vital resource, but one that is costly for the Bank to collect and firms to supply.
- ii. Technological developments have placed data at the heart of the modern economy. The ability to access, store, and analyse large volumes of data efficiently and effectively will continue to grow in importance for financial firms and authorities. This was one of the trends explored by the Future of Finance report, commissioned by the Bank and published in June 2019. The report recommended that the Bank develop a new digital data strategy, to improve the efficiency of the overall system and the effectiveness of the Bank’s data analysis.²
- iii. In response, the Bank committed to ‘launch a review in consultation with banks, insurers and financial market infrastructures to explore a transformation of the hosting and use of regulatory data over the next decade. The review will seek ways to decrease the burden on industry and to increase the timeliness and effectiveness of data in supporting supervisory judgements’.³
- iv. This discussion paper marks the launch of that review. Its scope covers all of the Bank’s structured data collections from firms, not just regulatory data – though these do represent the bulk of collections. It invites firms – both large and small – to work with the Bank to develop a new approach to reporting, with the aim of improving system-wide efficiency and the effectiveness of analysis. Close partnership with firms will be vital to building a common understanding of issues with the current approach, and assessing the benefits, costs and feasibility of potential improvements. Such partnership will also seek to complement firms’ own data strategies and leverage investment efficiently.
- v. The Bank, through its role of defining reporting across the financial sector, plays an important part in shaping how firms approach their own data. Transforming our data collections would offer an opportunity to support wider improvements to the quality and to usability of financial sector data, for example if this initiative can provide a lever to drive the development and adoption of data standards. Common data standards, widely used, represent a public good with wider benefits than just reporting efficiency, and can support private innovation. Reforms to data collection could therefore complement other initiatives such as the Bank’s renewal of its Real-Time Gross Settlement Service, which we are using to drive the adoption of payments messaging standards and legal entity identifiers;⁴ or the Bank-chaired Post Trade Market Practitioners Panel that seeks to address the coordination challenges holding back improvements in the efficiency and resilience of post-trade processes and operations.⁵
- vi. The Bank intends this paper to be the start of a dialogue with regulated firms and solution vendors to shape the evolution of reporting over a 5-10 year horizon. This timescale is intended to

1 Use of ‘the Bank’ in this document includes the Prudential Regulation Authority, save where the context requires otherwise.

2 The future of finance report, June 2019: Recommendation 9: <https://www.bankofengland.co.uk/report/2019/future-of-finance>.

3 The Bank’s response to the van Steenis review on the future of finance, June 2019: Priority 4: <https://www.bankofengland.co.uk/report/2019/future-of-finance>.

4 <https://www.bankofengland.co.uk/speech/2019/dave-ramsdem-afme-12th-annual-european-post-trade-conference>.

5 <https://www.bankofengland.co.uk/research/future-finance/facilitate-firms-use-of-technology>.

encourage an ambitious vision, while recognising that a number of significant challenges and information gaps still need addressing before widespread changes are possible. We intend this to be an inclusive process, and we recognise that it will be important to explore whether different approaches might be appropriate for different sectors, sizes of firm, and types of reporting. For example, we recognise that with the recent implementation of Solvency II reporting, insurers' perspectives may differ from those of banks.

vii. This paper explores and develops the ideas set out in the Future of Finance report, some of which the Bank has already been investigating through its involvement, with the Financial Conduct Authority (FCA), in a pilot on digital regulatory reporting (DRR).⁶ It also builds on messages given by firms to an earlier call for input on data collection by the FCA, and the approach the FCA is taking in the development of its Future Data Collection Platform.⁷ The Bank and the FCA will continue to work closely together to ensure that reforms to data collection are aligned.

Key points

viii. Firms and users face a range of issues relating to the efficiency and effectiveness of how we collect data. The process today is costly, takes time, is relatively inflexible, and involves a degree of duplication. The paper suggests that a few underlying factors may contribute to these effects:

- Heterogeneity in firms' data – for any given product or transaction, different firms might hold and describe equivalent data differently. This makes it hard for the Bank to write a set of reporting instructions that are clear and unambiguous to all firms. In turn, this can lead to “pain points” for firms in interpreting instructions and locating data, which has the potential to cause long timelines and quality issues for the Bank.
- Heterogeneity of the Bank's data needs – reports are designed to address specific use cases. For instance, the Bank often requires data to be aggregated in ways that makes reports hard to repurpose. This leads to more requests for new reports or breakdowns of existing reports than would otherwise be the case. It also leads to redundancy in the reporting process, as firms need to re-assemble the same underlying building blocks in different ways for different reports.
- Duplication of processes across firms – many elements of the production of reports are common across firms. This raises the possibility that further centralising some processes could reduce duplication and improve efficiency for the system as a whole.

ix. Some of these factors are inherent constraints arising from the variety of firms' business models and the Bank's objectives. Nevertheless, initiatives in the UK and around the world offer a range of possible solutions that could help address the problems they give rise to, and deliver improvements to the efficiency and effectiveness of reporting. Some solutions also have the potential to drive wider benefits to industry data beyond just reporting. We group these possible solutions into three blocks, with a range of options in each. These blocks are: (1) common data inputs (2) modernising reporting instructions and (3) changes to the reporting architecture.

x. Developing common data inputs at a more granular level would provide a defined way for firms to record certain data (e.g. data elements for individual mortgages) or capture the key elements in a common input layer. This could provide a more consistent cross-firm foundation from which to build

⁶ <https://www.fca.org.uk/digital-regulatory-reporting>.

⁷ Digital Regulatory Reporting Feedback Statement FS18/2, FCA October 2018: <https://www.fca.org.uk/publication/feedback/fs18-02.pdf>. The DRR pilot was carried out jointly between the Bank, FCA and a small group of firms.

our reports, reducing costs and improving speed and quality. The costs and challenges involved in agreeing and implementing common inputs could vary significantly across firms and data domains. One approach to developing common inputs would be to derive them from reporting needs. An alternative could be to rely on industry data standards, where these exist or can be developed. Doing so may result in broader benefits than just better reporting.

xi. Common data inputs could also form the basis of a move to modernise how the Bank writes reporting instructions. This could include moving from our current natural language approach towards more precise instructions for selecting and transforming the data of interest. Doing so could reduce the cost and time it takes for firms to respond to new requests. At the most ambitious end of the spectrum, writing instructions as code might allow straight-through processing of our requests. Before this could happen a range of technical and legal challenges would need to be overcome. Other changes to reporting instructions, not dependent on common data inputs, could involve annotation to make instructions easier to navigate, use of more standardised language to improve precision, or deeper collaboration with firms when developing instructions.

xii. Common data inputs could also support different architecture solutions such as a ‘pull’ model of data collection. A ‘pull model’ would allow the Bank to be able to query certain data held within firms and generate reports on demand. This could improve speed and flexibility of reporting while reducing the marginal cost to firms of responding to new questions. An alternative could be to establish a central service provider to improve efficiency by carrying out some of the steps that are currently duplicated across firms, such as interpreting instructions or compiling reports from common data inputs. A further option could be for the Bank to specify and collect more data at a granular level, reducing the need to provide reporting instructions but shifting responsibility for calculating transformations of the granular data to the Bank.

Your views sought

xiii. The discussion paper does not put forward a preferred solution, but rather sets out a framework for assessing the issues and an initial range of potential options. The paper provides preliminary views based on available information but does not claim to be comprehensive – we are particularly interested in suggestions for approaches we have not considered here.

xiv. This paper aims to start a conversation on transforming data collection, which should build a clearer understanding of how the current approach affects firms, and which solutions are worth pursuing. Further work with industry will be needed to assess the costs, benefits and feasibility of changes, the range of financial firms and report types that would best fit different solutions, and the appropriate timing and sequencing of implementation.

xv. Responses to the questions posed on page 6 and any other observations that readers may have in response to this DP should be sent by email to DatacollectionDP@bankofengland.co.uk by 7 April 2020. Responses and input are welcome from a wide range of stakeholders including regulated firms, industry bodies, specialist third-party providers, professional advisors, standards bodies and other regulators. The privacy policy is set out at the beginning of this document. Responses may be shared with the FCA.




xvi. As well as seeking written responses to the discussion paper, the Bank intends to establish one or more industry working groups to explore these issues during 2020. Other engagement channels will ensure wider input from those unable to participate directly, and draw on the work of existing

groups working on reporting and data standards. Details on how to express an interest in participating in working groups or other forms of engagement have been published on the Bank's website alongside this discussion paper. The Bank is particularly mindful of the need to ensure solutions can be implemented in a proportional way across all the sectors it interacts with.




xvii. Subject to responses to this paper, the Bank's aim for the working group(s) over the course of 2020 would be to develop a collective vision for data collection reforms over a five to ten year horizon, and proposals for immediate next steps that would move from pilots to live implementation. We expect to publish an update on responses and the proposed next steps during 2020.

Summary of options covered in the paper




Common data inputs

	 What is it?	 Benefits	 Issues
A common input layer	Commonly defined data points, designed to build authority reports.	May decrease cost and time to implement new requirements. Provides a basis for many of the other solutions.	Business case may depend on uncertain future requirements. Hard to define at a granular level for more complex reports.
Industry data standards	Industry standards for operational data, reused as the basis for authority reporting.	Also aligns authority view of data with firm view of data. Possibility of other benefits beyond reporting.	Challenges of getting agreement on industry-wide standards. Possible high cost of adopting new standards.

Modernising reporting instructions

	 What is it?	 Benefits	 Issues
Annotated instructions	Annotation of reporting instructions with standardised tags.	Better search and easier navigation for users, assisting interpretation.	Increases cost of writing instructions for authorities. Impact may be relatively limited.
Standardised language	Stricter standardisation of how instructions are drafted, enforced by technology.	Ensures instructions are drafted consistently making them easier to interpret and use.	Increases cost of writing instructions for authorities. Impact may be relatively limited.
Early engagement	Earlier and more iterative industry involvement in the process for designing reporting instructions.	A clearer understanding of what firms' data and systems look like. A better industry understanding of authorities' data uses.	Need to ensure industry engagement is inclusive. Cost of engagement for industry.
Instructions as code	Publishing a version of the instructions in a programming language, alongside natural language.	Ensures high level of consistency and precision in instructions. Implementation is more streamlined and automated.	Not all instructions may be fully expressed in code. Would need to clarify legal status of code.

Architecture and governance changes

	 What is it?	 Benefits	 Issues
Moving from push to pull	Authorities use APIs to pull data on demand rather than firms submitting reports.	Reduced costs for ad hoc requests. Greater timeliness and flexibility for authorities.	Cost to firms of maintaining on-demand data. Not suitable for some reports. Needs governance changes. Data security risks.
Central service provider	Some reporting processes are carried out centrally by a service provider.	Avoids duplication, potentially saving reporting costs. Increases consistency of data. Could support data publication.	Cost of running the provider. Requires changes to governance. Competition implications. Data security risks.
Collection of granular data	Authority collects granular data in place of reports.	Replaces need for firms to construct reports. Greater timeliness and flexibility for authorities.	Challenges around data quality assurance. Not suitable for some reports. Firms lose sight of data uses. Costs and risks to authority.

Questions

Overall

- A. **Which of the solutions identified (or combination of solutions) do you see as most attractive to explore further as a long-term goal, and why? Are there other promising options we have not considered?**
- B. **What do you see as the most useful actions to take as interim steps towards such a goal?**
- C. **Which sectors / reports should be prioritised, or excluded, in relation to the long-term goal and the interim steps?**
- D. **In what respects do you consider it most important that the Bank coordinates reforms to data collection with other UK or international authorities?**
- E. **What do you see as the most significant wider benefits to firms or to the financial system from improvements to data collection, beyond cost reduction?**

Costs (Chapter 3)

- F. What are the most significant areas of avoidable cost and challenge associated with the current reporting process, and what is the relative burden associated with different steps and types of report, as set out in the discussion paper?
- G. What non-regulatory developments might have a significant effect on reporting costs and challenges over the next decade (e.g. systems redesigns, use of cloud, AI, market developments)?

Common data inputs (Chapter 6)

- H. What are your views on the benefits and challenges from seeking to define a common set of data points as the basis for reporting?
- I. What additional benefits and challenges would arise from seeking to use industry data standards as the basis for defining reporting requirements? What should the role of regulators be in the development and adoption of such standards?

Modernising reporting instructions (Chapter 7)

- J. What are your views on the benefits and challenges of the possible improvements to reporting instructions set out in the paper?

Governance and architecture (Chapter 8)

- K. What are your views on the benefits and challenges of the possible changes to architecture and governance set out in the paper – in particular moving to a “pull” model for certain types of data, or moving some functions to a central service provider?

1 Scope and structure of the discussion paper

1.1 This discussion paper aims to set out the issues facing the current system of data collection from financial firms, then identify and explore a series of potential solutions in order to provide a structured framework for feedback and further discussion with industry.

1.2 Chapter 2 sets out why the Bank of England collects data from financial firms, the characteristics of the data it collects, and how the current system of data collection functions. It demonstrates the high level of diversity that exists across these features, indicating that potential solutions may need to vary accordingly.

1.3 Chapter 3 considers data collection from a firm perspective, seeking to identify where and why costs arise as firms go through the process of responding to a data request. It particularly explores challenges firms face in interpreting natural language reporting requirements and sourcing the data to fulfil them.

1.4 Chapter 4 sets out some of the limitations with the current system of data collection from a Bank of England user perspective including limits to analytical flexibility in data currently collected, data quality, and costs and delays involved in obtaining new views to address new needs;

1.5 Chapter 5 presents a high-level overview of potential reforms that could help address the issues identified in Chapters 3 and 4. It identifies a range of initiatives in the UK and other jurisdictions, and how these compare to the aspirational vision set out in the Future of Finance report involving common data standards, reporting instructions issued as code, and architecture that allows rapid and flexible collections.

1.6 Chapter 6 considers how developing and referencing detailed common data inputs could improve the way regulators ask for data and how firms go about sourcing it from within their systems. It looks at potential approaches for developing common data inputs, and some of the challenges in achieving these across the range of data underpinning the Bank's collections.

1.7 Chapter 7 considers possible ways of improving how the Bank goes about asking for data. Solutions range from incremental changes to how the Bank writes instructions, to make them clearer and easier for firms to process, through to providing instructions in code, which firms could execute automatically. The chapter sets out some of the conditions and potential challenges involved in such changes.

1.8 Chapter 8 looks at different ways of structuring how data are held and accessed as part of the reporting process. It considers the pros and cons of options including methods for authorities to "pull" data on demand, possible roles for central service providers, and the direct collection of underlying granular data.

1.9 Chapter 9 summarises the next steps.

1.10 Annex 1 provides a glossary of terms.

1.11 Annex 2 summarises the main data collections in scope of this paper.

Issues outside scope

1.12 Where relevant, the paper references other initiatives and factors shaping the environment in which any changes to data collection would take place. However, the focus of this paper is how the Bank collects data, so some examples of issues outside scope are discussed briefly below.

1.13 **Improvements to analytical tools using data.** The Future of Finance report identified the potential for the Bank to make better use of data by applying new analytical tools, particularly in relation to unstructured data and big data sets. The Bank's response committed to identify and implement improvements in the use of data, including the use of machine learning and AI to free up supervisors to spend more time making judgements. That initiative is outside the scope of this paper. However, the paper will consider to what extent changes to the current data collection system could complement and support new analytical techniques.

1.14 **Changes to the Bank's requirements for data.** The Bank's requirements for data have evolved significantly over the past decade and are likely to continue to change, driven by factors such as new analytical techniques, changes in the regulatory landscape and developments in the financial system. This paper aims to explore solutions to collections that would make future changes to data requirements more flexible and efficient, whatever form those changes might take.

1.15 **Data collected from financial firms by other authorities.** The scope of this paper is data collected by or for the Bank of England, including collections that are conducted by the FCA and used by the Bank. Various reports collected under Bank powers also currently involve submissions to an FCA data platform, and the FCA accesses and uses some of these data. The paper does not directly consider collections conducted by the FCA solely for its own purpose or by other authorities such as the ONS. The Bank intends to coordinate closely with other bodies in developing and taking forward any changes to data collection, in line with HM Treasury's ongoing review aiming to enhance coordination between authorities in the financial sector. This will include drawing lessons from HM Treasury's recently concluded consultation, which asked "how might firms and the regulators take advantage of new technology to make supervisory reporting more efficient, flexible and less burdensome".⁸ More widely, the Bank and the FCA are also preparing to implement enhanced coordination mechanisms in response to the industry's request for better "air traffic control".

1.16 Many UK firms also face data collections from authorities in other jurisdictions, often with overlapping but diverse requirements that impose additional cost. As Chapter 5 notes, a number of other authorities are also taking forward initiatives to reform data collection, so there is an important opportunity for greater international coordination in this area. We are interested to receive evidence from firms based on their experience of data collection and change initiatives in other jurisdictions, and suggestions of where future coordination will be most valuable.

8 Financial Services Future Regulatory Framework Review, HMT, July 2019: <https://www.gov.uk/government/consultations/financial-services-future-regulatory-framework-review>.

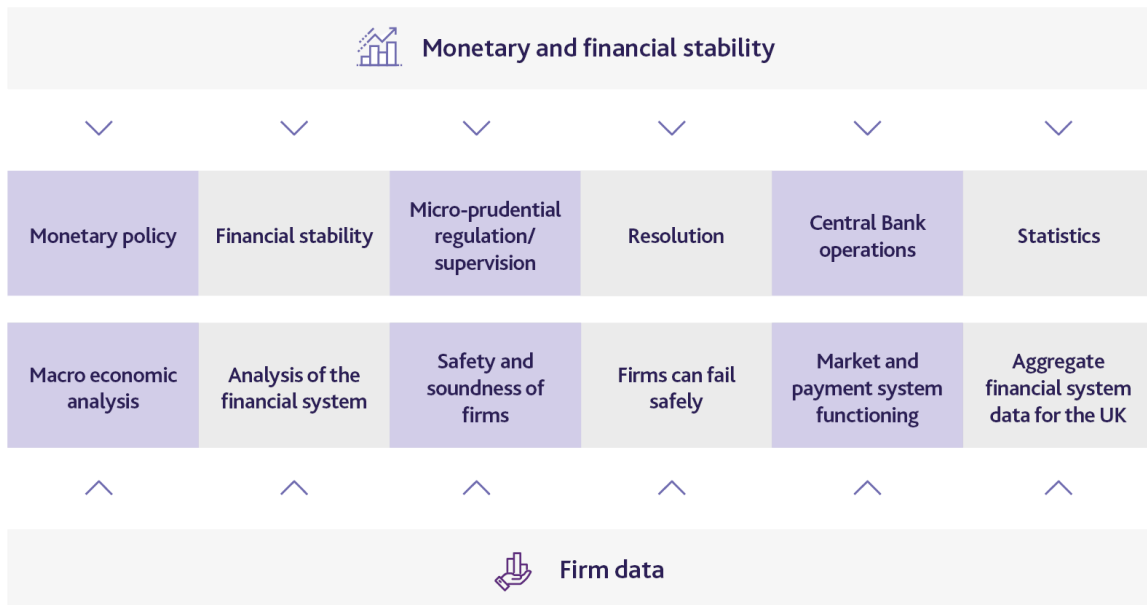
2 The Bank of England’s collection of data from firms

2.1 This chapter provides an overview of data collection by the Bank of England, as the basis for understanding the issues and potential solutions raised in subsequent chapters. It sets out why and how the Bank collects data, and explains why a diverse range of data collections have arisen to serve the various analytical uses underpinning the Bank’s objectives.

Why we collect data

2.2 The Bank of England’s mission is “to promote the good of the people by maintaining monetary and financial stability”. This mission spans a broad range of activities. Access to timely, high quality data is essential to the Bank’s ability to deliver its mission effectively across all of these, as shown in Figure 1. The Bank collects data from financial firms to make sure it has the full set of information it needs.

Figure 1: The Bank’s activities rely on data collected from financial firms



2.3 The Bank’s data requirements vary widely across these activities. Some activities require aggregate data on the financial system, including regular briefings for the Monetary Policy Committee (MPC) and Financial Policy Committee (FPC). The Bank also compiles data for the general public and other authorities, including the Office for National Statistics (ONS) and Bank for International Settlements (BIS). The FCA also uses data collected under Bank powers in respect of dual-regulated firms.

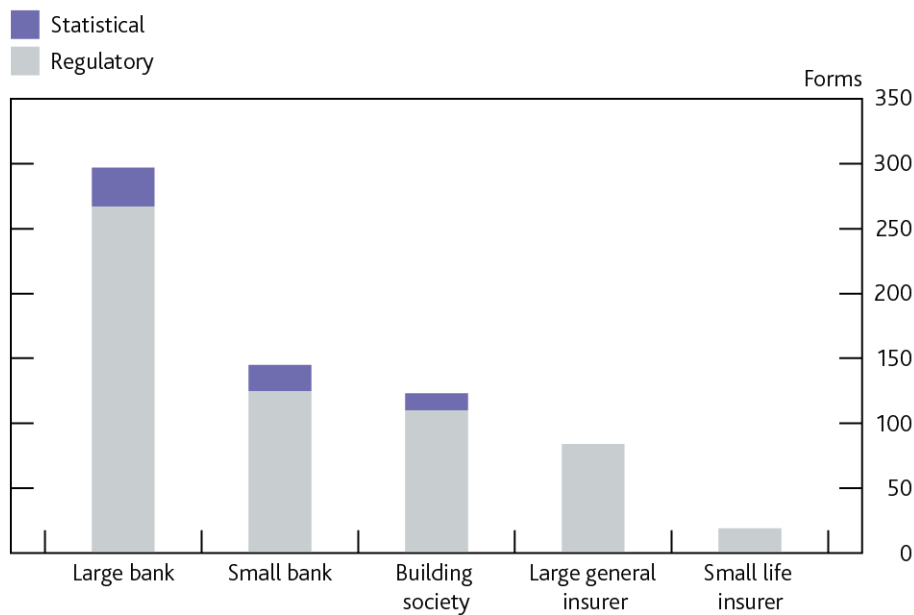
2.4 Other tasks, such as analysing risks for the FPC and the Prudential Regulatory Committee (PRC), require more detailed data on particular markets, such as mortgage lending. While aggregate data can help identify high-level macroeconomic trends, less aggregated data can help identify trends or risks in areas of the economy that could be of policy interest, such as the risks posed by highly indebted mortgage borrowers.

2.5 A third set of tasks – including the Prudential Regulation Authority’s (PRA) work as microprudential supervisor and the Bank’s work as resolution authority – require data about individual regulated firms. This ranges from high-level information about firms to information about specific activities they undertake, including critical regulatory data on firms’ solvency and liquidity. Data are also sometimes aggregated across firms to provide a lens on wider developments across the sector, if information can be collected on a consistent basis.

What we collect

2.6 This paper focusses on structured data collected directly by the Bank, as well as some data collected under FCA powers but shared with and heavily used by the Bank.⁹ The paper does not cover firm-specific requests arising in the course of supervisory activities, internal management information (MI) obtained from firms,¹⁰ or data collected in the course of the Bank’s market and banking operations. It also does not consider data the Bank uses but which is collected by others, such as the ONS, third party data vendors (based e.g. on published accounts or market prices) or Trade Repositories (TRs).¹¹

Figure 2: Typical number of reporting forms by firm and by type¹²



2.7 The majority of the Bank’s formal data collections can be categorised either as “statistical” data collected primarily to support monetary policy, or “regulatory” data collected primarily to support supervision, resolution and micro- or macro-prudential policy. Figure 2 estimates the number of regulatory and statistical reporting forms by different kinds of typical firm. The burden a firm faces is not solely determined by the number of forms it completes, but also by the number of reporting data points in each form and the complexity of those data points. As an example, one report

9 Primarily the FCA’s collection of mortgage product data: <https://www.fca.org.uk/firms/regulatory-reporting/product-sales-data-reporting>.

10 The Bank’s response to the Future of Finance noted the potential benefits from using advanced analytical techniques to process unstructured management information. While not the focus of this discussion paper, Box C on page 24 has further information on this area.

11 While TRs are the result of an important regulatory initiative, and the Bank uses the data they collect, the operation of TRs is independent and is the result of internationally-agreed requirements.

12 Source: Bank of England calculations. Certain collections (e.g. COREP) are counted as multiple forms.

collecting liquidity data from banks (PRA110) asks for over 25,000 data points for each relevant currency, whereas a report on bank positions in UK government bonds (Form GT) has fewer than 70 data points. The largest firms also report more regularly, report for multiple entities within their groups, and complete more data points within many of the forms. Large firms also report more ad hoc data, as described in paragraph 2.12.

2.8 The statistical data draw primarily on bank and building societies' finance systems to present balance-sheet information, with a particular focus on break-downs of lending and deposits.¹⁵ Most statistical collections are focused on UK activity. Typically, the data are collected via two-dimensional monthly templates which break down the balance sheet by dimensions of interest, such as product and counterparty.

Box A: Examples of data collections - Large exposures and Solvency II assets

Large exposures data provide quarterly information on firms' largest exposures to a single counterparty or a group of connected clients. Banks, building societies and certain investment firms provide detailed information on each counterparty. In addition to testing compliance with large exposure limits, this allows the Bank to assess certain concentration risks, identify firms' exposures to a common counterparty, and identify country specific concentration risk. The introduction of CRDIV and COREP reporting in 2014 increased the collection of large exposure information significantly: before 2014, firms reported only exposures over 10% of eligible capital; now, certain firms have to report any exposures over €300m on a consolidated basis as well.¹³

In Solvency II Asset data, insurers provide granular data on each of the assets they hold.¹⁴ For each asset firms provide over 30 data points including information about the nature of the asset, the issuer (for securities), economic sector, value, and acquisition price. Firms may provide more than one row of data per asset depending, for example, on whether part of the position has been pledged as collateral or ring-fenced. The data fulfil a number of purposes within the Bank including supporting supervisors' regular risk reviews. As with Large Exposures, these data can identify firms' exposures to a common counterparty; and identify country specific concentration risk. The granularity of the data means they can be used to support in-depth thematic (i.e. cross-firm) reviews into risks posed by particular asset classes without going back to firms with ad hoc requests.

2.9 Regulatory data are the largest and most diverse class of data collected by the Bank. They include collections from a small number of financial market infrastructures and around 1,500 PRA-regulated firms including banks, insurers, building societies, friendly societies, credit unions and designated investment firms. Regulatory data are often highly complex, requiring firms to make judgements and significantly transform their operational, financial and risk data to measure concepts such as capital and liquidity resources, and credit, market, liquidity and operational risk in line with regulatory definitions. These data are captured through templates of two or more dimensions, typically quarterly although some liquidity data can be collected daily from large firms.

2.10 Some of the data the Bank collects are relatively simple but granular. These granular data are closer to what firms hold in their operational systems. We estimate that around 15% of the Bank's

13 <https://eba.europa.eu/regulation-and-policy/large-exposures>.

14 <https://www.bankofengland.co.uk/prudential-regulation/regulatory-reporting/regulatory-reporting-insurance-sector#>.

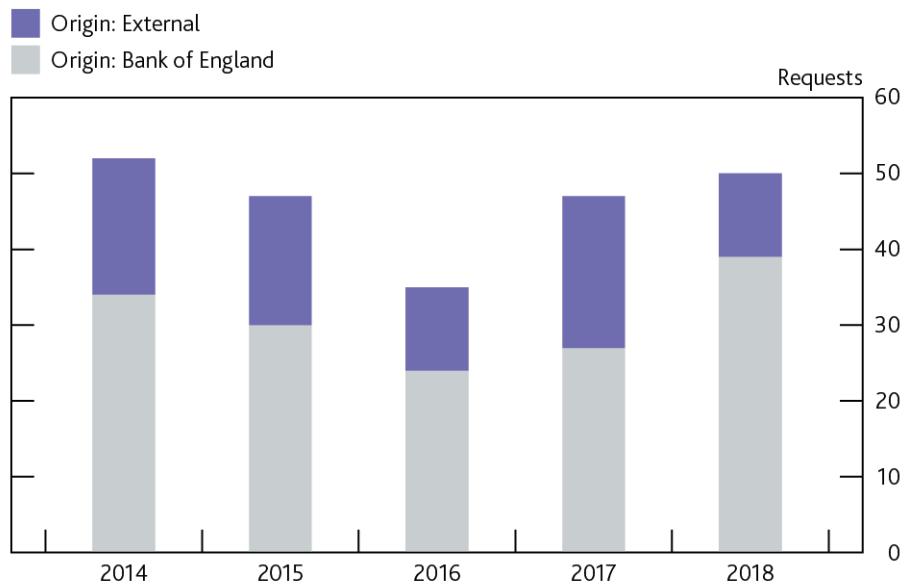
15 The ONS collects statistical data directly from insurers.

collection templates involve granular data. Box A gives two examples of data collections, one relating to transformed data on large exposures, and one relating to granular data on asset holdings.

2.11 The regularity of data requests varies across tasks. Much of what is required is collected on a regular basis, to allow ongoing monitoring. Some of this, such as data on regulated firms’ liquidity, may be collected more frequently in times of stress. Increasingly, the Bank is seeking to define data that may be required in a stress to support firms’ recovery or resolution, without collecting these data on a regular basis. One example would be detailed data on large firms’ trading portfolios which could support their wind-down. The Bank also collects daily transactional data on activity in the secured and unsecured sterling money markets, a subset of which are used to calculate the SONIA benchmark.¹⁶

2.12 Other data needs are harder to anticipate and are instead defined and collected on a purely ad hoc basis, as required. These could include data on an emerging risk for a firm, sector or market, data to help calibrate a new policy, or data to help conduct exploratory stress tests.¹⁷ The majority of cross-firm ad hoc requests are undertaken to support UK or international policy making, or to investigate thematic risks. Ad hoc collections might target a sub-set of firms – often just the most systemic banks or insurers. Some ad hoc requests require qualitative rather than purely quantitative information – for example on operational resilience. Figure 3 below shows the number of ad hoc requests to have varied only modestly in recent years, as well as the proportion that originate from outside the Bank (e.g. as a result of EBA, EIOPA or BIS processes).

Figure 3: Number of ad hoc data requests to firms¹⁸



¹⁶ <https://www.bankofengland.co.uk/markets/sonia-benchmark>.

¹⁷ The Bank conducts two regular concurrent stress test exercises – an Annual Cyclical Scenario (ACS) for the large UK banks and a Biennial Exploratory Scenario (BES). The structured data required for the ACS are largely stable and aligned to regulatory reporting. As the aims and focus of the BES change from test to test, so do the data needs for each exercise.

¹⁸ Source: Bank of England calculations. Excludes firm-specific requests by the Bank’s supervisory and resolution functions.

2.13 The Bank's diverse roles mean that it will sometimes use data collected primarily for one purpose to serve other purposes. An example of this is loan-level data on buy-to-let (BTL) mortgages, which is used by the MPC to understand the mortgage market, by the FPC to understand risks to and within this market, and by the PRA to understand firms' credit risk. At times, however, the fact that the Bank requests data to be aggregated for specific purposes results in a need for separate collections based on the same underlying data – we return to this issue of duplication in Chapter 3.

2.14 The number of collections has grown since the financial crisis. The main changes in recent years include: more granular liquidity information from banks; the Bank's Concurrent Stress Test; and collections relating to the implementation of Basel III, Solvency II and the new resolution regime. The Bank has more recently introduced granular position-level data from central counterparties and the reporting on sterling money market transactions and BTL mortgages mentioned above.

How we collect data

2.15 As the volume of data has grown, the Bank has taken periodic steps to rationalise its collections, improve their coherence and manage the burden on firms. These steps have included a review of data collections and the removal of collections that were no longer required. The Bank has eliminated some returns on intragroup exposures and market risk and aligned other returns (e.g. on profit and loss, balance sheet and data from branches) with wider reporting frameworks (e.g. FINREP) to improve coherence and reduce burdens. Where possible, such as in structural reform and resolution reporting, the Bank has also standardised data collections by building an underlying data point model, leveraging EU work (see Box B).

Box B: Data point modelling

Many reports to the Bank result from EU requirements, coordinated by European Supervisory Authorities (ESAs). Over recent years, the ESAs have led a transformation in how the templates used to collect regulatory data are designed, to improve simplicity and efficiency.¹⁹ Many templates are now prepared and modelled using a standard method called 'Data Point Modelling'. This aims to create a single dictionary of terms, and unique identifiers for those terms. These are associated to data points in reporting templates, as well as defining the relationships between these data points and the validation rules that apply. The resulting Data Point Model (DPM) is represented using the extensible business reporting language (XBRL), and creates reporting taxonomies that software solutions can use to produce reports for regulators.

2.16 Most firm data are submitted to the Bank through one of two data collection systems. The majority of the banking regulatory returns are collected via the FCA's GABRIEL system, which was introduced in 2008.²⁰ The other main collection system is the Bank's Electronic Data Submission (BEEDS) portal, which was introduced in 2015 and upgraded in 2018 with additional capabilities.²¹ Both systems enable modern online data submission, with support for system interoperability, automatic validation checks, and standardised processes for data transmission and storage.

19 <https://eba.europa.eu/regulation-and-policy/supervisory-reporting/data-point-model-dpm->

20 <https://www.fca.org.uk/firms/gabriel>.

21 <https://www.bankofengland.co.uk/statistics/data-collection/beeds>.

2.17 Current technology solutions should not be taken as a binding constraint in the context of this initiative, though leveraging existing investments will be important and the cost of any changes or new systems would be a relevant factor to consider for both firms and authorities. The FCA has an existing programme to replace GABRIEL over the next year, which is out of scope of this paper. The FCA's new platform is already in development and is being built with the flexibility to deal with automated collections, as well as enable additional data collection capabilities to be added over time, as prioritised by the regulators.

2.18 Further development of the ideas set out in this paper, through discussion with industry, would need to identify specific legal constraints that might affect some of the options – for example on writing instructions as code or using a central service provider to collate reports. Although we expect to take the legal and governance framework largely as given, the longer timeframe for this initiative may provide scope to recommend changes to legal frameworks where a case can be made.

2.19 The Bank collects data under various legal powers. The MPC relies on statistical data returns from banks collected under the Bank of England Act 1998. The Banking Act 2009 gives the Bank power to collect data to fulfil its role as Resolution Authority and as supervisor of recognised payment systems. The PRA uses data collected under its Financial Services and Markets Act (FSMA) powers and under EU Regulations, as does the Bank on a similar basis in its role as supervisor for other financial market infrastructures. The FPC relies on data from all of the above, and multiple functions also make use of data the FCA collects under its FSMA powers. The Bank also receives data on a contractual basis from firms through its operation of the Sterling Monetary Framework and the Real Time Gross Settlement system, though such data are not the focus of this paper.

2.20 While the UK remains an EU member, there are constraints on changing regulatory data collections directly mandated by EU legislation, such as CRR or Solvency II. The Government has also committed to ongoing adherence to international standards (such as those agreed by the Basel Committee on Banking Standards) which will continue to shape the approach taken to reporting. Where the Bank collects data under UK powers, it must follow a number of legal requirements when making changes. These include public consultation and cost-benefit analysis, which for PRA regulatory data are set out in FSMA. For statistical data, the Bank sets out its approach in its Statistical Code of Practice. When collecting data on an ad hoc or one-off basis, the Bank and PRA have a rigorous internal challenge and governance processes around the design and proportionality of requests.

3 Costs to firms

3.1 This chapter reviews existing evidence about the costs of reporting and seeks to identify potential “pain points” in the process, which reforms to data collection could address. Later chapters also note some of the wider benefits to firms that could arise from reforms, other than cost savings – for example in relation to industry-wide data standards or greater disclosure of collected data.

Overview of the costs of regulatory reporting

3.2 We need a clear understanding of how reporting costs arise across the financial sector in order to identify which solutions could bring the greatest efficiencies. We also need to understand how great those efficiencies might be, to allow industry and the Bank to determine if the up-front investment needed to implement solutions is worthwhile. Delivering efficiencies would be in line with the PRA’s statutory secondary objective to facilitate effective competition, which is supported by a strategic goal of ‘actively considering the proportionality of our approach as it contributes to the safety and soundness of the UK financial system’.²²

3.3 Any solutions that might apply over a 5-10 year horizon also need to take account of technological change and investment already taking place within firms, which might affect the scale and nature of reporting challenges, and change the cost/benefit calculations. We will need to consider how any new initiative can align with or even magnify the benefits of the changes firms are already making to the way they store and use data.

3.4 A number of surveys and studies indicate that the costs of reporting are significant. For example, in 2013, the PRA published a cost-benefit analysis of regulatory reporting requirements for CRR/CRDIV which estimated one-off implementation costs for banks, building societies and investment firms of £724m and ongoing costs of £956m per annum.²³ Meanwhile the ABI estimated annual ongoing costs for Solvency II reporting at £250m.²⁴ A study by McKinsey and Company in 2019 estimated that regulatory reporting by UK banks costs them £2 billion–£4.5 billion per year.²⁵ In December 2017, the European Commission (EC) launched a public consultation with regard to the on-going fitness of supervisory reporting requirements across the financial services sector.²⁶ Some 35% of respondents estimated average running costs of regulatory reporting at between 1% and 5% of operating costs; 55% estimated these costs at less than 1%. The EC report also identified respondents’ views of the major cost drivers including unclear requirements, lack of technical guidance and an insufficient level of automation.

3.5 Existing surveys have limitations. One challenge is that many do not clearly distinguish the marginal cost of regulatory reporting from the broader costs of managing and analysing data for a firm’s own purposes. For example, firms would need to calculate their solvency metrics to ensure compliance, regardless of any associated reporting requirement. There is also limited information

22 PRA Business Plan 2019/20, April 2018: <https://www.bankofengland.co.uk/prudential-regulation/publication/2019/prabusiness-plan-2019-20>.

23 PRA Consultation Paper CP5/13 ‘Strengthening Capital Standards: implementing CRDIV’, August 2013: <https://www.bankofengland.co.uk/prudential-regulation/publication/2013/strengthening-capital-standards-implementing-crd-4> (page 3 of 3).

24 ABI written evidence to the Treasury Select Committee, 26 January 2017: [http://data.parliament.uk/WrittenEvidence/CommitteeEvidence.svc/EvidenceDocument/Treasury/EU Insurance Regulation/written/47301.html](http://data.parliament.uk/WrittenEvidence/CommitteeEvidence.svc/EvidenceDocument/Treasury/EU%20Insurance%20Regulation/written/47301.html).

25 The future of finance report, June 2019: page 133: <https://www.bankofengland.co.uk/report/2019/future-of-finance>.

26 EC Summary Report of the Public Consultation on the Fitness Check on Supervisory Reporting having taken place from 1 December 2017 and 14 March 2018: https://ec.europa.eu/info/sites/info/files/2017-supervisory-reporting-requirements-summary-report_en.pdf.

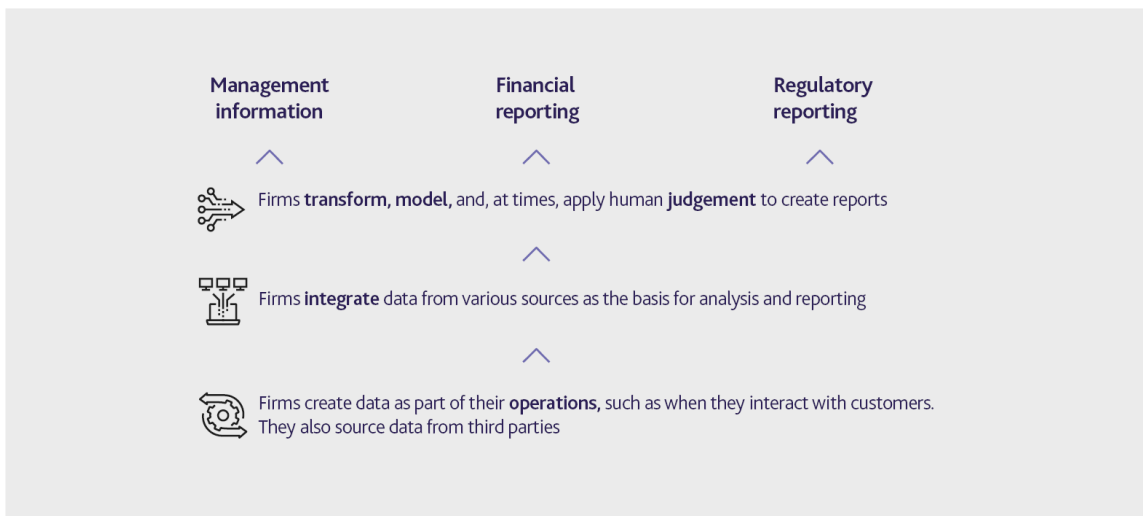
showing which aspects of reporting are the most burdensome, or how impacts vary across firms of different sizes and sectors. While costs are likely to be material for all firms, the nature of the reporting challenges will differ significantly between, for example, a large multinational insurer and a small UK building society.

3.6 The FCA’s calls for input on data collection provided a range of firm views on what drives costs; these inform the analysis in this chapter. The Digital Regulatory Reporting Project (DRR), undertaken by staff from the FCA, the Bank and a number of financial institutions, provides some useful granular information, albeit for a limited set of firms and reports.²⁷ We return to its findings at paragraph 3.10 below and in Chapter 5.

A stylised view of the reporting process

3.7 Firms engage in transformations of operational data for a variety of purposes. Responding to data requests from the Bank will draw on many of the same underlying data handling processes as producing management information or financial reports, as shown in Figure 4. Additional costs arise where the information requested is not used for another purpose, or where it needs to be aggregated in a different way.

Figure 4: Stylised representation of the relationship between firms’ operational and reporting data



27 Digital Regulatory Reporting Feedback Statement FS18/2, FCA, October 2018: <https://www.fca.org.uk/publication/feedback/fs18-02.pdf>.

3.8 Here, we set out a stylised view of the steps involved in creating a new report, or in changing or running an existing report, as the basis for seeking more detailed feedback on costs from firms. The main steps involved in providing data to the Bank are:

- (a) **Interpretation:** firms need to make judgements on the interpretation of reporting requirements, supported by their internal business line, policy, finance and/or compliance functions.
- (b) **Implementation:** firms need to translate their interpretations into technology solutions to source the required data. This includes:
 - (i) Identifying data sources, if available, to meet the request, potentially across multiple operational, risk, finance, purchased third party data sources or other management systems; or setting up new processes to collect the data.
 - (ii) Establishing technology solutions and processes to integrate, cleanse, and enrich existing data to meet the Bank's definitions and expectations of data quality.
 - (iii) Developing reporting and review tools and processes to produce the reporting data points.
- (c) **Process execution including oversight, governance and assurance:** firms need to produce and sign-off reports to meet the Bank's requirements. This includes:
 - (i) Running the technology solutions and processes that produce reports and deliver the data to the Bank, including controls such as validation checks and reconciliations.
 - (ii) Implementing regular management review of data and reports with sign-offs (where applicable) aligned with reporting deadlines (e.g. quarterly). Additionally, firms undertake periodic assurance work through reporting, operational risk or compliance review processes as well as internal or external audit work.
 - (iii) Responding to data queries raised by the Bank.

3.9 Figure 5 illustrates the overall process. All these steps are relevant to introducing new requirements, changing existing requirements or meeting ad hoc data requests, while for ongoing reporting step (c) is the most relevant. Firms of all sizes can make significant use of third parties in each of these steps. Large firms may do more work in-house while smaller insurers, banks and building societies may rely more on third party suppliers to assist with most steps. This may mean that in some cases solution vendors, rather than firms, have a better understanding of the challenges and costs involved.

Figure 5: Stylised process for responding to a data request



3.10 The DRR project considered the costs of the reporting process and attempted to identify the most significant cost drivers across ongoing reporting and new requests. Figure 6 provides its estimate of how costs split between each of the steps in paragraph 3.8, noting the caveat that this is based on DRR information from a limited set of firms for only one reporting domain – mortgages.

Figure 6: DRR estimated costs of each reporting step

	Total
Interpretation	15%
Implementation	28%
Process execution	57%

3.11 The next section considers each step in detail to identify potential ‘pain points’ in relation to new and ongoing reporting.

Challenges in the reporting process

Interpretation

3.12 The first step for firms receiving a new or amended data request is to interpret the authorities’ requirements. The costs of this step can vary significantly between types of data and request. One factor is how similar the data are to those firms use to run their business. More challenging requests require firms to understand and apply complex regulatory concepts or create new models to generate the required data.

3.13 Examples from insurance and banking illustrate the issue. Solvency II reporting on insurers' asset holdings captures basic asset features (such as currency and acquisition value) which firms are likely to use in their own management of those assets. Leaving aside other features of this return, the requirement is, at least, relatively easy to understand. By contrast, regulatory reporting on concepts such as Large Exposures is inherently more complex. It requires firms to understand and interpret the Capital Requirements Regulation, EBA Implementing Technical Standards, Guidance and Q&A as well as PRA Supervisory Statements.

3.14 Complexity arises from this need to review and interpret multiple interlinked sets of requirements. It also arises from the fact that those requirements may imply a complex "decision tree" for firms, which they need to follow to work out what to report. Additionally, interpretation requires firms to make their own judgements at various points about how those requirements apply to their own business. The governance process to agree these interpretations and the complexity of the requirements create costs, as well as risks that the data provided do not have the meaning that regulators intend.

3.15 A further challenge of interpretation is to translate reporting definitions and guidance written in natural language into unambiguous data and technology requirements. As each firm individually makes their own interpretation of definitions and guidance, this may also result in duplicated effort across the system as a whole. The existence of third party providers that specialise in interpreting instructions and provide services to multiple firms may mitigate this to some extent, as does guidance from regulators and standard-setters.

Implementation

3.16 Implementation typically involves significant challenges for all kinds of new or changed data requests. It covers several sub-steps including scoping and costing the changes, performing a gap analysis on information availability, sourcing data, calculating new metrics, and building and testing the processes and systems to extract, enrich and transform data.

3.17 Data sources might be spread across a number of systems, particularly for larger firms. In an EY survey of US banks' reporting to the US Federal Reserve Board, 72% of firms cited their largest technology challenge as multiple independent data sources, which lead to inconsistent information across the organization.²⁸ Data must be aggregated and transformed from customer-facing operational, management and financial reporting systems. Challenges increase further where firms themselves do not employ a consistent data model across their different business units and systems.

3.18 Data quality is a particular challenge for larger firms. Recognising this, the Basel Committee on Banking Supervision (BCBS) issued its 'Principles for effective risk data aggregation and risk reporting' in January 2013 ('BCBS 239') for systemically important banks (SIBs). BCBS 239 recognises that data management is not just a compliance exercise and stresses the importance for firms of sound, reliable and timely risk data. Large UK banks have therefore already been investing significant amounts to make improvements in this area. The BCBS has reported on compliance of firms with these principles, and concluded that 'banks have found it challenging to comply with the principles, due mainly to the complexity and interdependence of IT improvement projects'.²⁹ In the Bank's

28 Optimizing the regulatory reporting function, EY, 2018: page 14: [https://www.ey.com/Publication/vwLUAssets/ey-2018-federal-reserve-regulatory-reporting-survey/\\$FILE/ey-2018-federal-reserve-regulatory-reporting-survey.pdf](https://www.ey.com/Publication/vwLUAssets/ey-2018-federal-reserve-regulatory-reporting-survey/$FILE/ey-2018-federal-reserve-regulatory-reporting-survey.pdf). Although the survey is based on US firms, it corroborates our understanding from firms in the UK.

29 <https://www.bis.org/bcbs/publ/d443.htm>.

experience of implementing new data collections, other firms including insurers, smaller banks and building societies also struggle to some degree with data aggregation.

3.19 When the Bank implements a new request, firms may need to go back to underlying source systems, even where that request is similar to an existing collection. To take a simple example, the Bank collects various data on banks' activities in relation to Small and Medium Enterprises (SMEs) for both statistical and regulatory purposes. In these data requests, the Bank employs three different definitions of SME to meet different objectives, including the need to harmonise with wider (non-bank) statistical reporting and harmonised EU capital reporting. Meanwhile, none of these definitions may be the same as the one banks use when targeting and servicing small business customers. For each request, the firm may need to establish a new process to query data in underlying systems on customers' turnover, assets or employees to flag which are 'SMEs'.

3.20 A final data preparation challenge is the level of certainty over future requirements and the optimum level of investment to make in automation. If requests are ad hoc, significant investment is unlikely and processes are most likely to be manual. By contrast, firms might seek to automate significant ongoing reporting requirements as far as possible. However, as demands on data continue to evolve, committing expenditure to fully automated solutions is not always feasible. A common feature of the current implementation of reporting requirements is that the more changeable the Bank's requirements are, the greater firms' reliance on manual processes or tactical solutions. The EY survey indicated that in 2018, only 43% of the US firms sampled characterized their reporting process as "highly" or "mostly" automated.³⁰ Meanwhile, firms have highlighted to PRA supervisors that ineffective control of manual adjustments to their data and their end user computing solutions are key sources of operational risk.

Process execution including supervision, governance and assurance

3.21 One challenge within this step is that reliance on manual processes and tactical solutions, for the reasons set out above, increases execution costs on an ongoing basis. Manual steps in data provision also create risks to data quality. These costs and risks are exacerbated if time available for quality assurance is constrained, for example because of long processing times needed to bring together data from multiple sources.

3.22 As with other steps, the challenges are greater for data sets that require significant transformation and aggregation. The incremental costs of monitoring and assuring such data are higher than for data that more closely resemble those which firms capture in their operational systems and which they use to run their business.

3.23 The control, governance and assurance of data also represents a material cost for firms. Firms have top-down frameworks to enable effective control and assurance of their data. PRA-regulated banks and insurers must now identify a senior management function holder who is responsible for regulatory reporting. Some high impact data issues (e.g. relating to solvency ratios) may also be reviewed by the Board or its subcommittees. As proper assurance over data is vital, opportunities for savings in this area may be limited.

³⁰ Optimizing the regulatory reporting function, EY, 2018: page 16: [https://www.ey.com/Publication/vwLUAssets/ey-2018-federal-reserve-regulatory-reporting-survey/\\$FILE/ey-2018-federal-reserve-regulatory-reporting-survey.pdf](https://www.ey.com/Publication/vwLUAssets/ey-2018-federal-reserve-regulatory-reporting-survey/$FILE/ey-2018-federal-reserve-regulatory-reporting-survey.pdf). 'Highly' and 'mostly' automated are defined as having at least 70% of reporting from automated feeds as opposed to manual data inputs.

Questions

- F. What are the most significant areas of avoidable cost and challenge associated with the current reporting process, and what is the relative burden associated with different steps and types of report, as set out in the discussion paper?

- G. What non-regulatory developments might have a significant effect on reporting costs and challenges over the next decade (e.g. systems redesigns, use of cloud, AI, market developments)?

4 Challenges for users

4.1 Chapter 2 set out how the Bank seeks to use the data it collects from firms to support its objectives. Here we set out how the current approach to data collection can limit its effectiveness for those uses. The limitations relate to the timeliness, flexibility and quality of data; the Bank's operational costs; and the Bank's ability to benefit from new types of analysis in future. Many of these limitations arise because of the difficulties firms face in reporting data, as set out in Chapter 3.

4.2 The Bank's use of data has expanded over time as it has taken on new tasks such as annual concurrent stress testing and resolution planning. These new tasks have led to new data demands, which at times have accentuated the challenges users face under the current approach.

4.3 This chapter focusses on challenges to users in the Bank. Users within the firms may experience similar challenges, particularly when trying to use existing operational data for new purposes. We are also keen to understand firms' perspectives on the opportunities to improve data so it better supports their own needs. This will be an important step in understanding how we can work with industry towards solutions of mutual benefit.

Timeliness and flexibility

4.4 Whether taking decisions on interest rates or addressing a crisis at a firm or risks across the wider financial system, access to timely data is critical to the Bank. Given the challenges of compiling data, as set out in the previous chapter, timelines for submission of data can be long, particularly for some regulatory data. Statistical data are typically provided two to three weeks after the end of each month. Reflecting its greater complexity and wider scope, regulatory reporting is typically provided four to six weeks after a quarter's end, although in some cases this can be longer. This means that the Bank, whether through its policy committees or as the supervisor of an individual institution, may not have the most timely data when taking decisions.

4.5 At times of heightened risk, the need for up-to-date data increases. The Bank specifies some contingency data needs in advance. For example, larger banks can be required to report granular liquidity data daily with a one-day lag in a time of stress. However, in other cases the Bank faces limitations to the data it can obtain quickly. For example, concern in the market about certain counterparties may give rise to an ad hoc request to firms for data on their exposures to those counterparties. While firms can provide such data quickly, they may do so based on their own management information, rather than the regulatory templates that can be more easily aggregated. Similarly, we need to know what information large trading firms will need to carry out an orderly wind-down of their books in a recovery or resolution scenario. Currently it can take several months for firms to compile this information accurately.

4.6 The current approach to data collection relies heavily on using pre-defined templates. This can restrict the flexibility of what analysis the Bank is able to perform, since information is lost when firms aggregate data to meet the Bank's specifications. As a result, although reports meet the Bank's immediate needs, over time we may need to go back to firms for alternative cuts of the data to support specific purposes, for example assessing the impact of a change to policy or assessing a new risk.

4.7 As an example of the timeframes involved in changing or establishing new data collections, in 2016 the FPC assumed new powers over the buy-to-let mortgage market, and wanted to understand

better the risks it presents. The Bank already received a wide range of data on the mortgage market, which allowed analysis of various metrics of interest, such as on the distribution of loan-to-value ratios for the flow of new mortgage lending. Existing reports also allowed the Bank to see what proportion of overall mortgage lending was attributable to buy-to-let lending. However, because of the way the data were aggregated, they did not permit analysis of the loan-to-value ratios of buy-to-let mortgages. The Bank, therefore, decided to require a new report to permit the analysis needed in support of the FPC's new powers. We spent six months working with firms to define the reporting requirement, followed by a two-year roll out period. Only as of Q2 2019 are firms reporting good quality data with a minimal number of resubmissions. This two-to three year lag between a policy change and receiving good quality data is typical.

4.8 Even for ad hoc requests, timelines can be relatively long to get consistent data from firms. This reflects the time needed to explore with firms what data are available, specify a sufficiently clear data request and allow time for firms to produce and sign-off the return. For example, given the Bank's concern on growing vulnerabilities in the leveraged lending market from mid-2018, the Bank sought to collect data from the major UK banks to inform its ongoing assessment. The Bank shared the data request with firms in October 2018 and received the data in January 2019.

Quality

4.9 The quality of most data received by the Bank is fit for purpose. For example, statistics published by the Bank meet standards set out in its Statistical Code of Practice.³¹ However, where errors occur, they can have a significant impact on the ability of the Bank to fulfil its responsibilities. Complete, timely and accurate prudential data are fundamental to the PRA's supervisory approach, for example. Mark Zelmer's Independent Review of the Prudential Supervision of The Co-operative Bank³² recently recommended that the PRA should consider introducing more formal third-party reviews of key prudential information supplied by banking groups through their regulatory data returns. At the end of October 2019 the PRA wrote to firms stressing the responsibilities of firms and their Senior Managers to ensure their data are accurate, and indicated that it would be commissioning more reports from skilled persons on larger firms' reporting.³³

4.10 The Bank has a higher tolerance for quality issues where data are collected ad hoc or on a "best efforts" basis, but such data nonetheless can present significant challenges for analysis. Due to differences in the availability of data within and across firms, analysis can require the Bank to make a number of assumptions to plug data gaps and allow data to be compared or aggregated. This may undermine the robustness of conclusions drawn from the data. It can also be resource intensive for both users and firms. As noted in the 2019 Independent Evaluation Office Report on the Bank's stress testing approach, "substantive iteration between the Bank and the participating firms is required during the analysis phase to understand the supplied data and refine them further."³⁴

4.11 As noted in Chapter 3, firms face greater challenges when generating data just for the regulator, in a form which differs from how they would use the information themselves. For

31 Statistical Code of Practice, Bank of England, 2013: <https://www.bankofengland.co.uk/statistics>.

32 Independent Review of the Prudential Supervision of The Co-operative Bank Plc, March 2019: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/789802/38336_U_BoE_Independent_Review_Hi_Res.pdf.

33 Dear CEO letter: Reliability of regulatory returns, PRA, 31 October 2019: <https://www.bankofengland.co.uk/prudential-regulation/letter/2019/reliability-of-regulatory-returns>.

34 Evaluation of the Bank of England's approach to concurrent stress testing, April 2019: <https://www.bankofengland.co.uk/independent-evaluation-office/ieo-publication-on-the-boe-approach-to-concurrent-stress-testing-and-boe-management-response>

example, when reporting Solvency II Assets data, firms need to identify the economic sector of the issuer by using standardised economic classifications. This enables the Bank to compare peers and aggregate data in ways that can provide powerful insights into individual firms and the financial system as a whole. However, firms that did not use the same classifications in running their business needed to add this information to their existing data, which may have contributed to the initial poor quality of returns when this collection was introduced.

Operational costs

4.12 The Bank itself faces significant operational costs to request, receive and store data from firms. These include the costs to develop and maintain reporting forms and guidance, implement and maintain technology systems, run systematic data quality assurance processes and liaise with firms. A new approach that resulted in streamlined processes and higher quality data could present an opportunity to reduce these costs.

Adapting to technological change

4.13 Advances in analytical techniques, such as machine learning, have the potential to generate new insights on the financial system. The predictive power of these models is enhanced when they are trained on large volumes of data. So leaving aside other considerations, such approaches would be more likely to deliver benefits if they had greater access to granular data from firms. Balancing that consideration are other advances in analytical techniques, namely natural language processing (NLP), which suggest that less processed and less structured data could be collected from firms in the future (see Box C). Improvements to analytical techniques themselves are outside the scope of this paper, but the Bank will need to take account of these developments when considering changes to data collection, given the timescales involved.

Box C: The role of natural language processing in the analysis of firms' data

The Bank, particularly PRA supervisors, receive large amounts of unstructured data (text) from firms – so-called firm MI (management information). These include documents sent to firms' boards and other committees. These unstructured data often provide more timely and more detailed information to supervisors than structured regulatory data collections.

Advances in natural language processing (NLP) techniques mean that these data can be analysed in increasingly powerful ways. In recent years, the Bank has undertaken various pilots and projects using NLP to better interrogate firm MI. Applications include comparing the extent to which boards of different firms are concerned about the same issues, and identifying inconsistencies between firms' regulatory reporting, MI and published accounts. The Bank has also used these techniques to better understand how it communicates with firms it supervises.³⁵

NLP is a promising area, but the current state of the art suggests that firm MI will continue to supplement, rather than substitute for, structured data collections. This is because considerable effort is still required to extract insightful information from text. Moreover, the nature of the MI that the Bank receives varies from firm to firm. As a result, cross-firm comparison and aggregation using firm MI is difficult.

35 Staff Working Paper No. 688, October 2017: <https://www.bankofengland.co.uk/working-paper/2017/sending-firm-messages-text-mining-letters-from-pra-supervisors-to-banks-building-societies>.

5 What might a new reporting approach look like?

5.1 Chapter 3 explained where some of the costs in reporting currently lie for firms. It suggested that there are particular ‘pain points’ associated with the interpretation of reporting instructions, the implementation process of finding the necessary data in firms’ systems and how to extract them, and the manual processes involved in executing reports that add to the ongoing run costs of each report.

5.2 Chapter 4 identified several issues that users of data face, including limitations on how flexibly existing reports can be interrogated, the long process for requesting new reports, and the quality and usefulness of data supporting analysis.

5.3 While there are a range of underlying causes of the problems identified in Chapters 3 and 4, some factors seem particularly relevant:

- Heterogeneity in firms’ data – for any given product or transaction, different firms might hold and describe equivalent data differently. This makes it hard for the Bank to write a set of reporting instructions that are clear and unambiguous to all firms. In turn, this can lead to “pain points” for firms in interpreting instructions and locating data, which can cause long timelines and quality issues for the Bank.
- Heterogeneity of the Bank’s data needs – reports are designed to address specific use cases. For instance, the Bank often requires data to be aggregated in ways that makes reports hard to repurpose. This leads to more requests for new reports or breakdowns of existing reports than would otherwise be the case. It also leads to redundancy in the reporting process, as firms need to re-assemble the same underlying building blocks in different ways for different reports.
- Duplication of processes across firms – many elements of the production of reports are common across firms. This raises the possibility that further centralising some processes could reduce duplication and improve efficiency for the system as a whole.

5.4 Some of these factors are inherent constraints arising from the variety of business models and the Bank’s objectives. This chapter introduces a range of potential approaches that could help address the problems they give rise to, drawn from initiatives around the world and the Bank’s own experience with the digital regulatory reporting pilot.

Future of Finance

5.5 The Future of Finance report commissioned by the Bank recommended ambitious goals for the collection and storage of data. It argued that requests for structured data should be ‘machine-executable’ – coding the requests such that firms could automatically pull the necessary data from their systems. It also suggested that rather than having firms push data to the Bank on a regular basis, the Bank should have on-demand access to a shared data lake.

5.6 Noting the ambitiousness of these goals, it suggested a range of steps towards achieving them. These include improving reporting consistency via a common taxonomy, accessing firm data directly through application programming interfaces (APIs) for high-value use cases, or building a common data utility to hold granular data that could be accessed and analysed in near real-time.

International experience

5.7 Various central banks and financial regulators around the world have initiatives to improve the effectiveness and efficiency of data collection. These span a number of different approaches, and range from fully-implemented projects to those which are only at a conceptual stage. This section considers a few of the more developed projects in detail to identify their core components and what lessons the UK could draw from their experience. While these projects tend to focus on data collection from banks, many of the lessons would also be relevant to other sectors such as insurance.

BIRD, IReF and AuREP

5.8 The ‘Banks Integrated Reporting Dictionary’ (BIRD) is a collaboration between the European Central Bank (ECB), European national central banks (NCBs) and commercial banks to reduce reporting burden and improve the quality of data reported to the NCBs and ECB.³⁶ It recognises that each bank has to interpret reporting instructions and map from their own data to the reporting data points. It tries to reduce that burden by bringing banks and reporting authorities together to work on the transformations required to construct the reporting data points. Practically, it seeks to agree on a common ‘input layer’ with banks, sitting above their operational data layer, from which a common set of transformations can be constructed collaboratively to produce reporting data points. The initiative began in 2015 and has worked through a number of statistical and regulatory reports since then.

5.9 A related initiative is the European System of Central Banks’ (ESCB) Integrated Reporting Framework (IReF), which is under consultation.³⁷ This initiative aims at integrating into a standardised reporting scheme a wide range of existing statistical reporting requirements that are currently applied differently by NCBs in the Euro area. It would define a sufficiently granular set of requirements, which banks would report. The BIRD input layer would be extended to also cover IReF requirements. The hope is that the reporting requirements collected under IReF would be fairly stable through time, such that future statistical reporting requests could be met by specifying transformations from them, materially reducing work for firms in handling new demands.

5.10 A similar approach has already been implemented in Austria, with the additional step of a central utility – AuRep – which has been built by a group of large commercial banks to manage reporting to the Austrian central bank.³⁸ Under this model, firms populate a broad standardised input layer with mostly granular data. Data in the input layer are sent to a central utility, which performs transformations of the data on behalf of firms to meet a wide range of reporting requirements. Both the input layer and related transformations are developed through close cooperation between banks and the central bank, under the central bank’s lead. The utility then reports directly to the central bank on behalf of the banks. To date, it covers nearly all statistical reporting of banks and includes a number of supervisory reports (eg FINREP solo, resolution planning national supervisory content) with more to follow in coming years.

5.11 These initiatives have the common feature of involving experts to analyse reporting requirements addressed to banks and determine the data and related transformation rules required

36 What is the BIRD, ECB: https://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/reporting/html/bird_dedicated.en.html.

37 The ESCB Integrated Reporting Framework, ECB, April 2018: https://www.ecb.europa.eu/pub/pdf/other/ecb.escb_integrated_reporting_framework201804.en.pdf.

38 Managing the processing chain from banks’ source data to statistical and regulatory reports in Austria, OeNB, August 2018: <https://www.oenb.at/Publikationen/Statistik/Statistiken---Daten-und-Analysen.html>.

to meet them – a task which would otherwise need to be completed individually by every reporting bank. This, in turn, helps ensure that data reported by banks to authorities are comparable, consistent and comply with the requirements.

Rwanda

5.12 The National Bank of Rwanda has established another variant of the granular data extraction model. Pre-defined templates with guidelines have been shared with all reporting institutions to obtain data electronically. The National Bank 'pulls' data based on these templates from firms' core systems into the National Bank's data warehouse, and performs transformations on them in order to meet reporting requirements for internal and external users.³⁹

Other jurisdictions

5.13 A number of other authorities have announced initiatives to reform data collection, which largely remain at an earlier stage of development. These include the US Consumer Financial Protection Bureau (CFPB) and Commodity Futures Trading Commission (CFTC), the Monetary Authority of Singapore (MAS), the Hong Kong Monetary Authority (HKMA), the Japan Financial Services Agency (JFSA), and the Philippines Central Bank (BSP).

5.14 The US bodies are planning Tech Sprints and competitions around reducing reporting burden and increasing automation.⁴⁰ MAS have a roadmap to progressively reduce duplication and automate data submission by financial institutions.⁴¹ HKMA are conducting studies on the feasibility of greater use of automation and granular data in reporting as well as encouraging industry to explore the use of new technology to enhance the regulatory process.⁴² JFSA has announced a proof of concept to reduce reporting burdens on regional banks by accepting granular data held by banks in a flexible manner.⁴³ BSP has undertaken a pilot project to replace a legacy of e-mailed Excel-based templates with a single schema for submission by each bank via API, with the aim of moving to a 'pull' model.⁴⁴

Bank of England data collection from central counterparties

5.15 The Bank supervises the UK's financial market infrastructure including the UK's three authorised Central Counterparties (CCPs). From 2015-2019, the Bank carried out a project to transform the system of data collection from CCPs, which had grown to encompass numerous firm-specific Excel reports and responses to ad hoc queries.⁴⁵ This exercise provided a chance to try out new ways of defining data collections.

5.16 We agreed to receive a regular (in many cases daily) set of 13 highly structured, ISO 20022 compliant XML reports common to all CCPs. These broadly covered every financial risk aspect of a CCP's business, from the products they clear and their clearing member margin requirements, through to their financial accounts and investment activities. Having just a few parties involved allowed us to work more closely with the firms, to ensure we had a clear understanding of how CCPs

39 BNR Annual Report 2016-2017: <https://www.bnr.rw/news-publications/reports/annual-reports/>.

40 <https://www.cftc.gov/LabCFTC/index.htm>; https://files.consumerfinance.gov/f/documents/cfpb_rfi_tech-sprints.pdf.

41 <https://www.mas.gov.sg/news/media-releases/2018/mas-moves-towards-zero-duplication-of-data-requests>.

42 <https://www.hkma.gov.hk/eng/news-and-media/speeches/2018/09/20180927-2/> (Para 28).

43 https://www.fsa.go.jp/en/news/2019/20190703_joubun/01.pdf.

44 See <http://www.bsp.gov.ph/publications/media.asp?id=5127> (paras 5-7).

45 The Bank of England's supervision of financial market infrastructures — Annual Report 2017-2018, Section 4.4.1: <https://www.bankofengland.co.uk/news/2018/february/supervision-of-financial-market-infrastructures-annual-report-2018>.

themselves were storing data. This allowed us to take a flexible and iterative approach to the design of the collections, in order to meet our needs while minimising the potential for confusion.

5.17 As an example, we wanted to create flexible views of which clearing members and clients were exposing the CCP to risk. To do this while avoiding full details on every position in every product for every client, we asked firms to aggregate the data by each product's maturity date. This one degree of aggregation preserved much of the granularity that supported flexible analysis, while greatly reducing the amount and sensitivity of the data we received. The simplicity of the aggregation also made it easy for the CCPs to code the change in their systems.

Digital Regulatory Reporting

5.18 The Bank and the FCA, in partnership with a small group of firms, recently concluded a pilot of technology to explore the possibility of Digital Regulatory Reporting (DRR).⁴⁶ This followed earlier calls for input by the FCA on data collection. The pilot has explored the feasibility of regulators publishing a code version of their instructions – so called “machine executable regulation”, which firms could automatically convert into code that runs directly in their systems – thereby greatly streamlining the interpretation step. The pilot tested the concept using synthetic mortgage data supplied by the firms in a standardised format, running on a dummy system. Regulators had the functionality to schedule or run reports as required, or pull data directly from the firm via an API.

5.19 Some of the conclusions from the calls for input and DRR pilot were that:

- Firms participated in the pilot on the basis that they saw scope for improvement in regulatory reporting processes, both in reducing the costs involved and improving data quality.
- Focus areas for improvement could be broken down into four stages: addressing ambiguity in reporting requirements; building a common data approach; mapping requirements to firms' internal systems; and submitting data to the regulators. Many firms highlighted the particular importance of building a common data approach.
- The team developed a process for building a granular data model that could be used as the foundation for a number of existing reports. However, the team recommended this model requires further validation before being used in a production environment. Replacing the reports with data defined at a granular level could reduce the complexity of reporting for firms and authorities.
- The team explored a number of solutions for expressing regulation in code. None currently met all the DRR team's requirements, but this might be possible with further time and investment.
- The team used distributed ledger technology (DLT) to test its compatibility with some aspects of the DRR vision. This demonstrated how some aspects of reporting could be automated using DLT. However, there were significant challenges identified in the potential use of DLT as a full-scale solution.

⁴⁶ <https://www.fca.org.uk/digital-regulatory-reporting>.

A framework for considering solutions

Common data inputs

5.20 All of the initiatives rely, to some degree, on agreeing what we might call a ‘common input layer’ – a standardised set of data inputs across firms, which can then be transformed into reports in a common way across firms. This should reduce the cost of implementing new reporting requests, to the extent that they can be met from transformations of existing data in the input layer. On the authorities’ side, it should increase the set of available data and reduce the time delays in meeting new requests.

5.21 The standardisation of input data is a starting point for a range of possible broader solutions, including various forms of automation. However, some forms of data may be more amenable to standardisation than others. Many initiatives have focused on statistical data, which tend to aggregate information on things like mortgages or commercial lending that share many similarities across firms and can thus be standardised at a relatively granular level. However, in its role as micro-prudential regulator, the Bank may be interested in other kinds of data such as outputs from a firm’s risk management processes, which may differ across firms and cannot be standardised as easily. Chapter 6 explores these issues in more detail.

Making the interpretation of reporting instructions more efficient

5.22 Most of the solutions that build a common input layer then use this to support a more efficient approach to reporting instructions. The approaches range from collaborating on the transformations on pen and paper, but leaving firms to work out how to execute them; to actually writing the transformations as code that firms can execute directly. Taking a collective approach to producing and interpreting reporting instructions reduces burdens by avoiding each firm having to undertake this process in isolation.

5.23 Again, this may be more feasible for some types of reporting than others. Where the transformations are deterministic aggregations of data from a clearly defined input layer, it should be easier to agree on the transformations and even to write them as code, facilitating automation. However, where the transformations reference concepts that require expert interpretation – such as dynamic references to accounting principles and regulatory constructs like capital metrics – it may be harder to agree on transformations and those transformations may require expert human inputs, such as judgements on classification or valuation. It is perhaps for these reasons that many of the initiatives above have focused on statistical reporting, at least initially, rather than regulatory reporting. Chapter 7 explores the issues around these options in more detail.

Changing the architecture and governance of reporting

5.24 Finally, some solutions go beyond having firms execute standardised transformations on standardised data inputs and move that work to a central utility or even to the reporting authority. They also open up the possibility of moving away from a regular cadence to reporting, instead allowing a reporting authority to reach out for data when required. These ideas are explored further in Chapter 8.

5.25 An important point to note in relation to any solutions involving changes to architecture is that firms will continue to have responsibility for ensuring that they comply with their prudential requirements. It will therefore remain necessary for them to ensure key prudential metrics are calculated accurately, regardless of any changes to the architecture of regulatory data collection.

6 Common data inputs

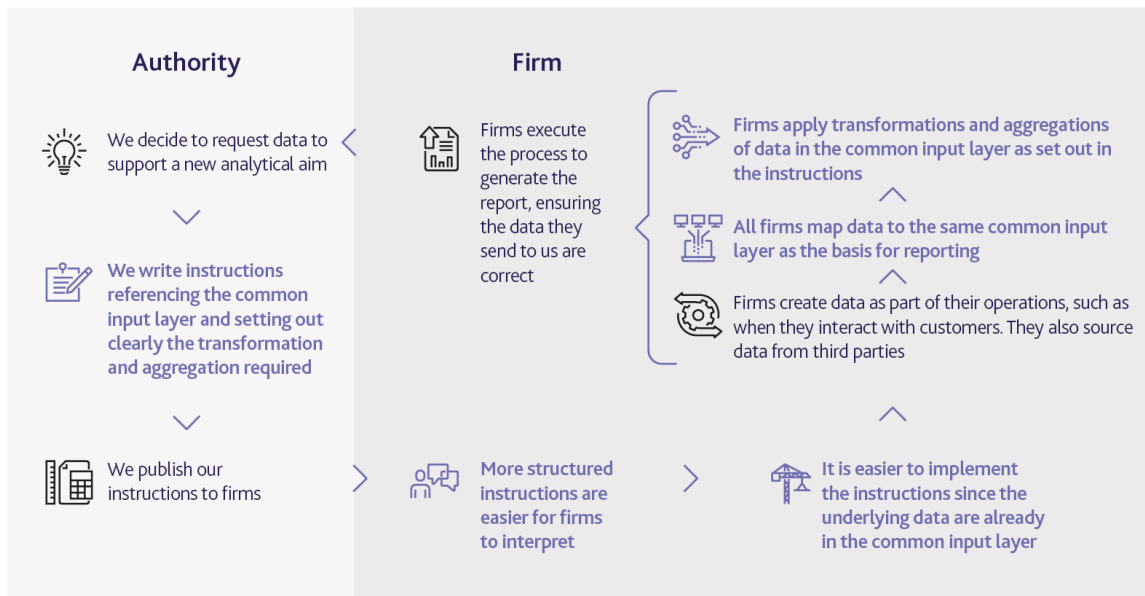
6.1 As set out in Chapter 3, one challenge that firms face when responding to a new data request is understanding what source data they need to use as the basis of their response. Every firm will have their own way of defining and storing most types of data, so there will be differences in how data points are labelled, formatted, and even whether they are collected and stored.

6.2 As a result, there is often no way for an authority such as the Bank to describe the data needed to compile a new or updated report in a way that universally maps to every firm’s systems. An authority cannot therefore easily start from the point of considering what data are available within firms and construct reporting instructions from the bottom up. Rather, a new request will more typically start from the analytical goals, specifying data points at a high level and leaving it to individual firms to work out if they collect the necessary components to fulfil the request, and where and how they are stored.

Specifying data requests directly from common data points

6.3 To address this challenge, most of the data collection initiatives reviewed in Chapter 5 involve some form of a commonly defined set of data points from which reports can be built. We refer to this as a “common input layer”. Firms would undertake an exercise to map data in their systems to the common input layer and establish a process to keep it populated with data. Reporting instructions could then be written to reference this input layer directly, in a clear and unambiguous way, as shown in Figure 7 below (note that the improvements to reporting instructions mentioned here are explored further in the next Chapter).

Figure 7: Adding a common input layer and improving reporting instructions



Benefits and costs

6.4 Many firms already individually build the equivalent of input layers to generate individual reports, so direct efficiency savings in relation to executing established reports may be limited. Others use third party solution vendors to do something similar on their behalf, with the vendor then mapping from their own input layer to reporting requirements. Nonetheless, where multiple reports rely on the same underlying data, this approach could eliminate some duplication by mapping to a consistent, underlying input layer. For instance, in the DRR pilot, it was estimated that the approximately 860 data points collected in a set of UK mortgage reports – many of which are aggregated data points – could be built from a granular layer of 160 data points. There could also be ongoing savings to management and maintenance costs from having simpler systems.

6.5 A greater cost saving from this approach could come from making new requests cheaper and faster to respond to, since firms would avoid having to identify and extract data from elsewhere in their systems if it existed in a common input layer. Most of the improvements to reporting instructions considered in Chapter 7, and potential changes in architecture in Chapter 8, also rely on some form of a common input layer.

6.6 Defining and creating a new common input layer clearly involves up-front costs. The return on that investment would depend on firms being able to meet new requests by reference to data already in the input layer. New requests that required data outside this would still generate additional costs, and require time to source and integrate the extra information. One way to tackle this would be for the input layer to hold a broader set of data points than is needed to meet current reporting requirements. However, predicting future data needs may be easier in some areas than others, and an overly precautionary approach would involve more up-front costs. The common input layer may therefore have to evolve over time as needs change, but having a consistent starting point which covers most data needs could nevertheless be valuable.

Other issues and challenges

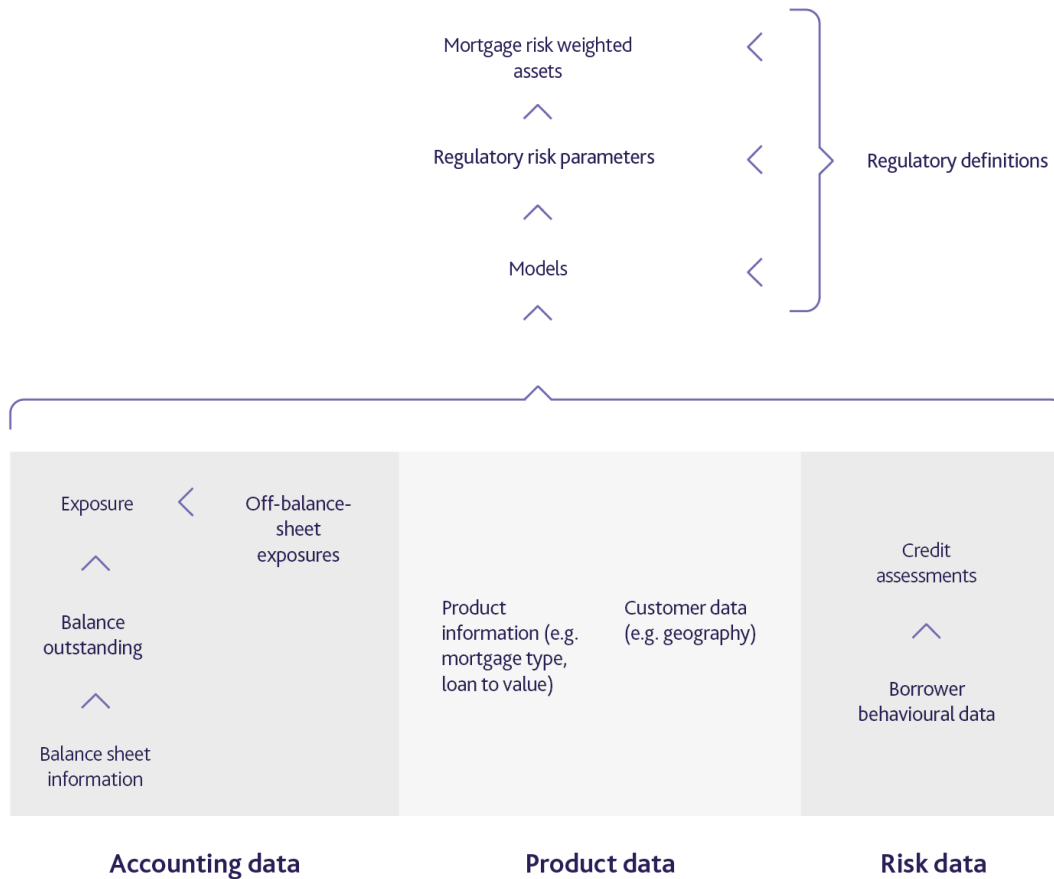
6.7 For some products, there can be a lot of heterogeneity in the way firms record data about a product, including whether they even capture certain information. This can affect calculations about the optimal level of granularity to aim for in the common input layer. More granular definitions would normally do more to reduce ambiguity, provide a better basis for reporting instructions to reference, and widen the variety of reporting data points that can be constructed from the input layer. However, heterogeneous starting points will make it harder to agree granular standards and more costly for firms to implement them. The alternative could be to use a higher level of abstraction for the input layer, despite the reduced benefits.

6.8 Further, not all reported data points are straightforward functions of granular data. Many data points depend on firm specific pricing, risk calculations or forecasts. Some refer to external concepts, such as accounting or regulatory standards that themselves may require judgement. Calculating them may also involve manual steps which are harder to automate, and which might only be carried out periodically. Large firms may face additional complications in aggregating certain data across multiple legal entities, and the need to take account of factors that apply across a group such as master netting agreements, set off rights or portfolio credit mitigation.

6.9 As an example of the breadth of data types involved in certain types of report, Figure 8 illustrates the main components and steps that might go into producing risk weighted assets for mortgage exposures. Starting in the bottom left corner, accounting data provide information on

balances and exposures. In the middle, data about the product and customer will often be recorded in a separate system. On the right, risk data about borrowers' behaviour will feed into credit assessments. All of these in turn feed into a firm's models, which together with regulatory definitions feed into risk parameters and mortgage RWAs – itself just one component of the overall capital calculation.

Figure 8: Stylised representation of inputs to mortgage RWA calculation



The role of data standards

6.10 So far, we have described an input layer in terms of reporting needs. An alternative would be to take an approach based on industry-wide data standards.⁴⁷ Where they exist, or could exist, industry-wide data standards for operational data could greatly simplify the cost of building and maintaining a regulatory input layer, since data could be pulled directly from a firm's operational systems, with minimal processing needed. Such an approach could also improve the quality of reporting data by improving consistency across a firm's systems and aligning reports more closely to the way a firm stores and uses data for its own purposes. Industry data standards could also have wider benefits, including for managing information within firms and across firms. For example, the

⁴⁷ Note that by 'data standards', here we mean a standardised description and identification of the content of data expressed in a form a machine can understand. This is in contrast to standards in the technologies used to store or transmit data.

Bank's Post-Trade Technology Market Practitioners Panel has identified and is considering the potential for data standards to help address pinch-points in the post-trade process.⁴⁸

6.11 The most straightforward areas in which to promote greater use of standards are again likely to relate to granular contract or transaction data. At present, standards exist, with varying degrees of uptake, for some products. The development of ISDA's Common Domain Model – described in Box D – is a prominent example. However, for other products, contracts may only be standardised at the firm level (and sometimes not even then, for firms consisting of multiple entities).

Box D: The ISDA Common Domain Model (CDM)

ISDA's CDM is a new industry standard for derivatives.⁴⁹ Much like the existing Financial Products Markup Language (FpML) it contains a formal representation of the data contained within a derivative contract. Unlike FpML, and many other existing standards, it also has a formal representation of the events that take place during the lifetime of a derivative contract. This should increase the consistency of derivatives data. In turn this should reduce the cost of reconciling data between parties and systems after lifecycle events occur. The CDM is still a young standard and not yet widely adopted. However, solutions like the CDM could increase the quality and transparency of operational data. In turn this could increase the quality of data reported to the Bank, while allowing for greater precision in reporting instructions.

6.12 Some products contain bespoke features specific to a particular customer, or jurisdiction, which complicates standardisation. Even the definition of commonly used terms such as “default” may differ from contract to contract. If contracts are nonetheless very similar, it might be possible to use a reference contract and build a standardised machine representation of that reference contract. Even in a market with a lot less standardisation, it may still be possible to agree a common abstraction layer above the contracts that can be used as the basis of a data standard.

6.13 Agreeing industry-wide data standards would therefore not be a trivial exercise, and firms would also need to undertake investment to migrate existing data on their systems to the new standards.

Standards for accounting, regulatory and statistical data

6.14 Figure 8 showed that the production of some reporting data points requires the interpretation of accounting and regulatory policies. Statistical reporting similarly requires interpretation of national accounts concepts. Many of these concepts are already described in extensive standardised documentation. The meaning of financial accounting concepts are described by standards such as International Financial Reporting Standards (IFRS). Regulatory concepts, such as capital and regulatory liquidity, are defined at a high level by international standard setting bodies such as BCBS or IOSCO, and then in detail in national and international legislation.⁵⁰ If these standardised concepts had clear machine-readable representations, it would be easier to use them in describing

48 The Bank established this panel following the Future of Finance review, to explore how market participants could leverage technological improvements to deliver a more efficient and resilient post-trade ecosystem:

<https://www.bankofengland.co.uk/research/future-finance/facilitate-firms-use-of-technology>.

49 <https://www.isda.org/2019/10/14/isda-common-domain-model/>.

50 Such as the European Union's Capital Requirements Regulation.

data stored in systems as part of the reporting process. It would also be easier to use them in reporting instructions, as described in Chapter 7.

The role of authorities in developing data standards

6.15 Financial authorities can play an active role in developing standards themselves, as they did in the case of the Legal Entity Identifier (LEI) and more recently the Universal Transaction Identifier (UTI) and Universal Product Identifier (UPI) - see Box E. Alternatively, they can use their standing in the financial community to help coordinate others, trying to ensure standards meet the needs of the markets – this fits with the approach the Bank is taking to the use of data standards in post-trade, through the Market Practitioner Panel.

6.16 Authorities can help drive adoption of standards by requiring or signalling an intention to require their use in contexts where they interact directly, such as regulatory reporting or through their role operating payments infrastructure. This could in some cases help to overcome a collective action problem, which may arise since individual firms require an incentive to move to any standard that does not reflect their own existing approach. However, regulators' ability to force the use of standards is limited in processes where they do not interact directly. Those decisions ultimately lie with firms and solution vendors.

Box E: Authorities' role in data standards - the Legal Entity Identifier and ISO 20022

During the financial crisis of 2007-09, it was very difficult to understand the credit, funding and other linkages and exposures between legal entities, both within and between firms. In response, regulators and the industry adopted Legal Entity Identifiers, a 20-character, alpha-numeric code for enabling the clear and unique identification of legal entities participating in financial transactions.⁵¹ Increasing usage of the LEI and the further spread of unique product and transaction identifiers, alongside existing identifier standards such as the International Securities Identification Number (ISIN) for securities, has made the financial system more transparent and robust. They also offer the opportunity to more easily integrate and analyse data relating to the same customers or securities across the systems and entities within a firm, as well as between firms.

Another example where the Bank is using its role to drive wider adoption of an industry standard is in relation to the ISO 20022 messaging standard.⁵² The Bank has collaborated on an ISO 20022 compliant XML-based standard for money market transaction reporting (used in the Sterling Money Market data collection in the UK) with the ECB to enable a common standard to be used, harmonising the way firms report these data. The Bank, alongside other high value payment system operators including the Federal Reserve and ECB, will also move to ISO 20022 over the next few years, as we renew the UK's high value payment system, RTGS.⁵³ By harmonising around an aligned data model for payment messages, central banks as payment system operators can support market demand to move to this standard, and prompt vendors of banking and accounting software to mirror the data elements in their systems.

51 <https://www.fsb.org/work-of-the-fsb/policy-development/additional-policy-areas/legalentityidentifier/>.

52 <https://www.iso20022.org/>.

53 <https://www.bankofengland.co.uk/payment-and-settlement/rtgs-renewal-programme>.

6.17 Where regulators take a role in driving adoption of a standard, there is a case for ensuring it is open and technology agnostic, in order that it can be widely used and not favour particular users or vendors. As in the case of LEIs, UPIs and UTIs, there can be a clear case for international coordination on developing standards if the products span multiple jurisdictions. This would also help ensure that multinational firms can adopt standards on a global basis, avoiding fragmentation and additional cost and improving the potential for data sharing. The costs and challenges of securing international agreement on new data standards may however be significant.

Questions

- H. What are your views on the benefits and challenges from seeking to define a common set of data points as the basis for reporting?
- I. What additional benefits and challenges would arise from seeking to use industry data standards as the basis for defining reporting requirements? What should the role of regulators be in the development and adoption of such standards?

7 Modernising reporting instructions

7.1 As set out in Chapter 3, the current approach to issuing and interpreting reporting instructions can contribute to cost and delay around new data requests. This chapter considers a range of possible steps to address this, up to the most ambitious and challenging option of publishing a code version of the instructions alongside natural language.⁵⁴

Current reporting requirements

7.2 A reporting requirement specifies what data an authority wants, from whom and when. A reporting requirement has three parts:

- **Reporting rules:** require a firm or set of firms to submit data on a certain date or given the occurrence of a certain event. Reporting rules also provide a template that visually specifies the content and format of the required data. For many regulatory reports, these rules will have legal status.
- **Reporting instructions:** precisely explain which data points a particular reporting rule requires and how to produce them. Reporting instructions may be supplemented over time with additional clarifications and “Frequently Asked Question” documents.
- **Technical specifications:** specify the required digital format of the data. For ad hoc requests this is likely to be an Excel template. For regular reports this is likely to be structured using an XML based standard such as XBRL or ISO 20022, or according to a proprietary XML schema. There are also validation rules that check that the submitted data conform to logical and / or mathematical conditions.

7.3 When discussing ‘reporting instructions’ in this paper we refer primarily to the meaning above, but some changes under consideration could also have implications for reporting rules and technical specifications.

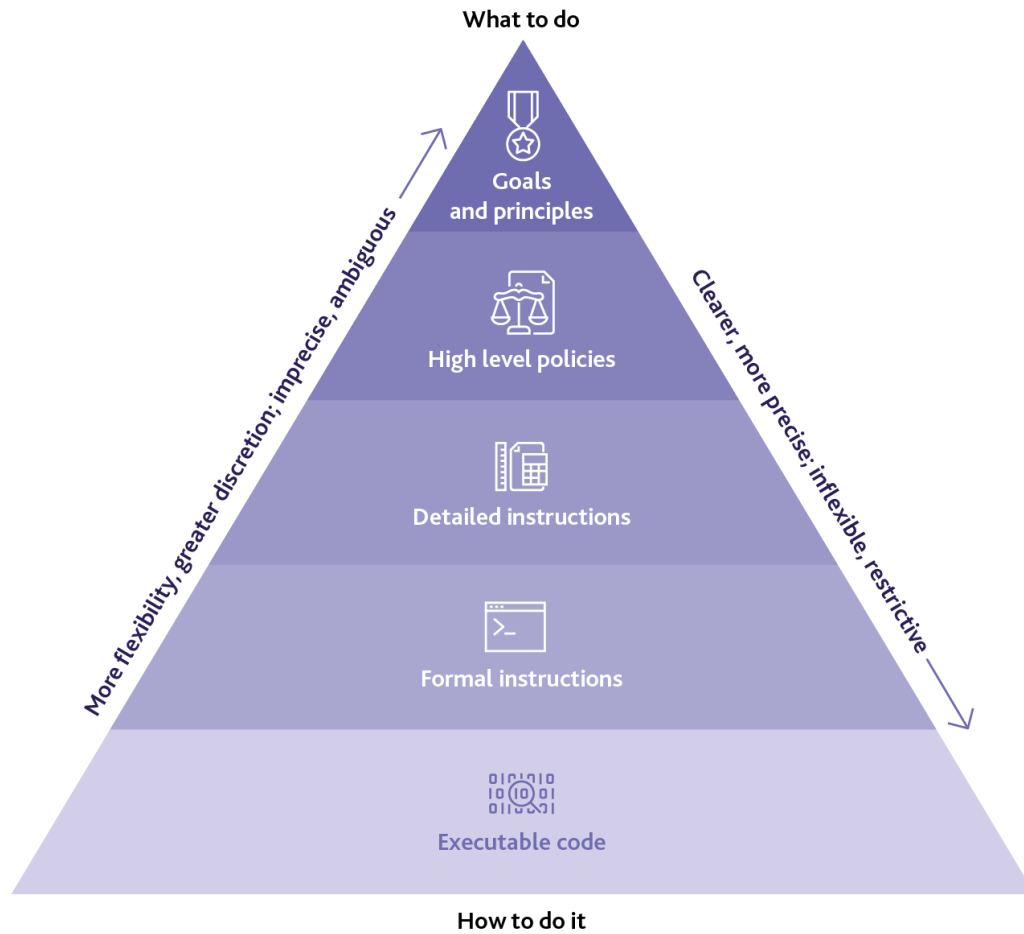
Interpreting and implementing instructions

7.4 The process of writing, interpreting and implementing reporting instructions can be visualised as moving down through the layers in Figure 9 below:

- Data collections start with a high-level objective or goal, for instance to get further transparency about a particular part of the financial sector.
- An authority sets out high-level policies – such as reporting rules – that require a set of firms to submit a report.
- An authority supplements rules with reporting instructions and templates that provide more detail on what is required.
- Firms will often create a more detailed set of instructions that detail how to build the report in the context of the firm.
- Internal technology teams / solution vendors then implement these as code in their solutions, accompanied where needed by manual operating procedures.

⁵⁴ Where data requirements on UK firms have origins at an international level, including from EU Regulations, some of the changes discussed in this Chapter might only be possible if pursued at that level.

Figure 9: Writing and implementing reporting instructions



7.5 At each stage, requirements become more precise and closer to what will ultimately be executed. Costs and delays arise from the number of layers that firms are left to progress through, and the difficulty in moving from one layer to the next. An authority such as the Bank faces a trade-off when defining its reporting instructions: increasing the detail improves their clarity and precision; but going too far may remove flexibility and discretion, and could end up asking for data that firms do not have or excluding desirable data with unanticipated characteristics.

7.6 As set out in Chapter 3, interpreting instructions (understanding what they are asking) and implementing them (establishing processes to find, extract and transform the relevant data) require a number of costly and time-consuming steps. Three features exacerbate this:

- Instructions are written in natural language, requiring interpretation.
- Instructions often draw on other official or third-party documents to define terms fully.
- Instructions are limited in the level of precision they can specify, because firms and the way they hold data are heterogeneous and common data inputs are lacking (as discussed in Chapter 6). This requires instructions to retain a degree of flexibility, which means firms must then undertake extra steps to implement a definitive process for their own data.

Potential improvements to reporting instructions

7.7 Modernising reporting instructions to make them cheaper and faster to interpret and implement could encompass a range of improvements set out below. Each of these improvements comes with its own set of benefits and challenges. What we term “annotated instructions”, or “standardising natural language” both focus on addressing the first two bullets in paragraph 7.6: they are designed to help humans interpret natural language instructions. “Early engagement” on the development of instructions would be an attempt to avoid or reduce interpretation and implementation challenges by collectively working through more of the layers in Figure 9. Common data inputs, as discussed in Chapter 6, are not essential for these three approaches, but could increase their effectiveness. In contrast, many of the benefits from the more radical final option of “instructions as code” seem likely to depend on having well-defined common inputs, in order to address the third bullet of paragraph 7.6.

Annotated instructions

7.8 Technology allows the annotation of reporting instructions in order to make them easier to navigate and understand (see Figure 10 for a stylised example).⁵⁵ Web pages, for example, can have links that allow for easier navigation between and within related documents than an unannotated document such as a pdf. Text can be tagged with a range of metadata in this way, which enriches its functionality by creating the ability to extract and use sub-sets of the information in an automated way. Such steps are sometimes referred to as making documents ‘machine readable’, though this is a term that can encompass a broader range of document types. We already use some annotations to power our online version of regulatory rules: the PRA Rulebook. Similarly, third party solutions that assist firms interpreting regulation will often have a machine readable version of the regulation at the core.

7.9 However, the Bank could go further in this area. It could extend the functionality we already provide for regulatory rules to other aspects of the reporting instructions. The annotations that implicitly power the PRA Rulebook and many vendor solutions do not conform to a formal standard. We could publish a version of our instructions that conform to a new or existing standard. This could enable third parties to create cheaper or more sophisticated interpretation tools. As part of this process, we could extend the list of tags we can use to annotate instructions. For instance by tagging terms with their representation in a data standard; or tagging which bit of the instructions relate to the content or timing of reporting. While these options do not remove human involvement in interpreting instructions, such steps could make instructions faster to interpret and reduce the likelihood of human error. More extensive annotation would involve a cost for the Bank, so the case for this solution would depend how extensive the benefits to firms would be relative to the current approach.

⁵⁵ There are existing standards for publishing annotated machine readable rules such as LegalRuleML.

Figure 10: Rewriting our instructions – illustrative examples

Natural language	Annotated instructions
Constrained by natural language grammar (and some informal legal drafting conventions)	Natural language instructions annotated with standardised machine readable "tags"
Unless otherwise stated, this Part applies only to: (1) an incoming firm; or (2) a third country firm A firm must provide the PRA with information in accordance with the Branch Return Form.	<pre> <applicability> Unless otherwise stated, this Part applies only to: (1) an <definition id=123>incoming firm</definition>; or (2) <definition id=234> a third country firm</definition> </applicability> <rule id="BR_SUB" type="OBLIG">A firm must provide the PRA with information in accordance with the <form id=ABC>Branch Return Form. </form></rule> </pre>
Standardised natural language	Instructions as code
Constrained by natural language grammar + formal drafting + vocabulary restrictions	Use of formal syntax, unconstrained by natural language
A Third Country Reporter is by definition an Incoming Firm or a Third Country firm Each Third Country Reporter must submit a Branch Return Form	<pre> If ("INCOMING_FIRM" OR "THIRD_COUNTRY_FIRM") in firm.type then: firm.type.append("THIRD_COUNTRY_REPORTER"); If ("THIRD_COUNTRY_REPORTER") in firm.type then: firm.submit(BranchReturnForm); </pre>

Standardising natural language

7.10 The Bank already seeks to draft instructions in a clear and consistent manner. However, we could look to use technology to enforce stricter standardisation of how we write natural language reporting instructions. Rather than just attempting to follow guidelines on how to write reporting instructions, we could use technology to check those guidelines are being met. There are existing standards in rule drafting;⁵⁶ if adopted, these or other similar standards could provide a basis for enforcing consistency in how we write reporting instructions (see Figure 10 for a stylised example). Technology options also exist to help check whether natural language rules are compliant with those standards. Standardising drafting would not necessarily make instructions shorter but, as with adding annotations, could make them easier to use and understand, reducing the chances of errors and speeding up the collection process.

7.11 Standardising the drafting of instructions would not require the development of common data inputs. However, standardised instructions could refer to common data inputs if they existed, supporting greater precision and further improving interpretation and implementation steps.

Early engagement on the development of instructions

7.12 The process for finalising reporting instructions already involves a consultation phase with industry. However, it might be possible to deepen the degree of engagement at an earlier stage, so that the Bank and industry engage on interpreting policy requirements with a clearer collective understanding of what relevant data and systems actually look like. This could allow ambiguities to be identified and addressed more quickly, avoiding the need to make clarifications once some firms have already started implementing. From a data collector’s perspective, this might involve investing

56 For example, the Semantics of Business Vocabulary and Rules (SBVR), a standard adopted by the Object Management Group: <https://www.omg.org/spec/SBVR/>.

more time and resource when developing instructions, but the pay-off could be higher quality data and a faster data collection process overall.

7.13 If undertaken alongside an approach involving common data inputs, an early engagement stage could involve ensuring that there is a clear understanding of how such inputs need to be transformed to produce reports, and identifying any areas where common input definitions need to be refined or expanded to support the new report.

7.14 Early engagement would need to respect the policy-making process, and should not be seen as an opportunity to change the objectives of the data request. However, small adjustments could be appropriate if it could be shown these would still deliver policy objectives and do so in a significantly more efficient way. This approach may not work as well in some areas of regulatory reporting, where data requirements flow directly from definitions of other regulatory rules, such as capital rules. Collections that originate from international initiatives would also require that early engagement take place at that level.

7.15 There could be some challenges around ensuring early engagement is sufficiently inclusive. The cost of early consultation would also be a factor – firms would need to recognise the benefits of investing resource before instructions were finalised in order to make later processes faster and cheaper.

Instructions as code

7.16 A more radical approach would be to publish a version of instructions in code, alongside the natural language version (see Figure 10 for a stylised example). This would ensure instructions are expressed with greater precision and consistency, effectively moving down through the layers in Figure 9 above – leaving firms with fewer, easier steps to interpret instructions and implement their reporting processes. Writing instructions as code should make them easier to turn into production code that runs in live reporting systems. For a software developer or business analyst, reading code may even be easier than natural language. Writing a code version of the instructions could also allow the Bank to test the instructions before they are issued.

7.17 Being able to write precise code that produces these benefits will probably depend on it being able to reference common data inputs, as discussed in Chapter 6. Without these, firms would still need to fill the gap between the inputs to the published code and the data in their systems. If it is possible to make common data inputs more detailed and comprehensive, this could support more precise and directly implementable instructions as code.

7.18 The benefits of providing code might be expected to increase for requirements with definitions defined in multiple interlinked regulatory publications, or which involve complex decision trees for firms. Machines can be better than humans at processing such instructions, providing humans can input any judgements required. Code could in principle be written in a way that specifies deterministic transformations alongside some elements that require a judgement to be made. However, code will never remove the need for human judgement completely. Indeed code could make it clearer where human judgement is required as an input to reporting and help ensure that those judgements are implemented correctly and updated as necessary as the firms' business and regulation evolve.⁵⁷

⁵⁷ eg the PRA's Dear CEO letter of 31 October 2019 highlights the need for firms to be able to identify and validate the judgements they are taking: <https://www.bankofengland.co.uk/prudential-regulation/letter/2019/reliability-of-regulatory-returns>.

7.19 One vision for instructions as code would be for the Bank to publish them in a directly machine executable form. This would aim to specify as fully as possible the logical and mathematical relationships between underlying data, judgements, and reportable data points. This could further simplify and speed up the steps for responding to a new request. It is unlikely however that this could extend to a fully automated solution, whereby the Bank could execute code directly on a firm's systems, given diversity across firms and the security and governance issues this would raise. A constrained version of automation, using an application programming interface (API) as a gateway for the Bank to pull data, is considered further in Chapter 8.

Issues and challenges

7.20 For financial authorities to write and publish a code version of instructions is, to our knowledge, almost unprecedented in the world of data collection. The DRR pilot showed that this concept can be successfully executed in a constrained test environment, but identified a number of issues and challenges, which would require further exploration in consultation with industry.

7.21 Some sets of reporting instructions are likely to be easier to express as code than others. Instructions for reports that provide a view on a set of products may be easiest. This includes many of our statistical reports and regulatory reports such as parts of MLAR,⁵⁸ or the securities held by insurers as reported on in the Solvency II assets request. Coding data that require human judgements as inputs (as is the case for much financial and regulatory reporting) may be trickier, and as noted above, cannot remove the need for those judgements to be taken by humans. A code version of instructions merely makes it clear what judgements are required and how they need to be incorporated.

7.22 Publishing a code version of instructions alongside natural language would have legal considerations. There is some precedent here: some jurisdictions maintain multiple versions of the law in different languages; and commercial firms publish plain English versions of their contracts. Nonetheless, such an approach would require clarity on the legal status of the code version, and how to handle any errors caused by faithful implementation of incorrect code within a firm's systems.

7.23 Writing code instructions would require changes to the skillsets of people involved in reporting. Most obviously, the Bank would need staff to be comfortable writing instructions as code, and have appropriate processes and controls. Firms too would need to adapt their people and processes to the change, though vendors could also assist with implementation.

7.24 In pursuing this path, a decision would be needed on what specific technological approach to use, such as what language to write the code in. It would be important to support interoperability and avoid selecting a language that unduly favours a particular firm, vendor or technology. The code would need to be transparent to non-coders, so that they could understand and verify the code is correct.

Questions

- J. What are your views on the benefits and challenges of the possible improvements to reporting instructions set out in the paper?

58 Mortgage Lending and Administration Return.

8 Reporting architecture and governance

8.1 International experience suggests that it may be possible to improve the efficiency and effectiveness of data collection by going beyond standardising data and improving reporting instructions to changing the architecture and governance of data collection. Two themes stand out from the examples surveyed in Chapter 5 and are picked up in both the Future of Finance report and the Bank's response: moving to a more centralised architecture; and moving from regular submissions of data – 'pushing' data to authorities – to having authorities 'pull' data on demand. This chapter lays out some considerations on each before inviting responses.

From push to pull

8.2 One model, as adopted by Rwanda (see Chapter 5) involves a move from firms submitting data using a 'push' mechanism, to authorities requesting data from firms using a 'pull' mechanism. For instance, firms could make their data available via an API. The Bank and other relevant authorities could connect to the API and request data. For this to be manageable for the authorities, it would need to be built on some form of common input layer, as discussed in Chapter 6. As discussed there, that might limit what kinds of data could be held ready in standardised form in order to be pulled.

8.3 The API could restrict access to the data, limiting the data that the Bank or other authorities could pull to the set of data to which they have access rights. It could also set constraints on the requests, such as on the amount of data and minimum level of aggregation. Returning to Figure 7 in Chapter 6, this could be akin to replacing the report preparation step with a process that allowed the Bank to query part of the common input layer directly.

Benefits and costs

8.4 A pull mechanism could have a number of benefits for firms and the Bank. Replacing pushing reports with pulling data may reduce duplication in firm sign-off processes. Firms could sign off one set of data rather than a number of reports. The Bank could move to only collecting the data we need, when we need it, which could materially improve timeliness and save costs on storing and manipulating large datasets. Storing less data may make it easier to meet rules on data security and the treatment of personal data. Finally, data could flow through from operational systems in nearer to real time, increasing the timeliness of data.

8.5 From a firm's perspective, a move from push to pull should decrease the marginal cost of delivering a file to close to zero, but ensuring an API is continuously working and available with the right data behind it would require a certain level of investment and ongoing resource. Whether this would generate overall cost savings may vary across firms depending on how the costs of a pull system compare to the costs of generating and sending files manually, and on the overall number of requests it replaces.

Issues and challenges

8.6 Moving to a pull model could require changes to governance. Currently reporting rules specify when a firm must submit data. Under a pull mechanism, they would need to specify when firms must make data available as well as when, and how often, the Bank could pull data. The rules around resubmissions would need to change; the Bank would no longer receive information of a resubmission by receiving a new file.

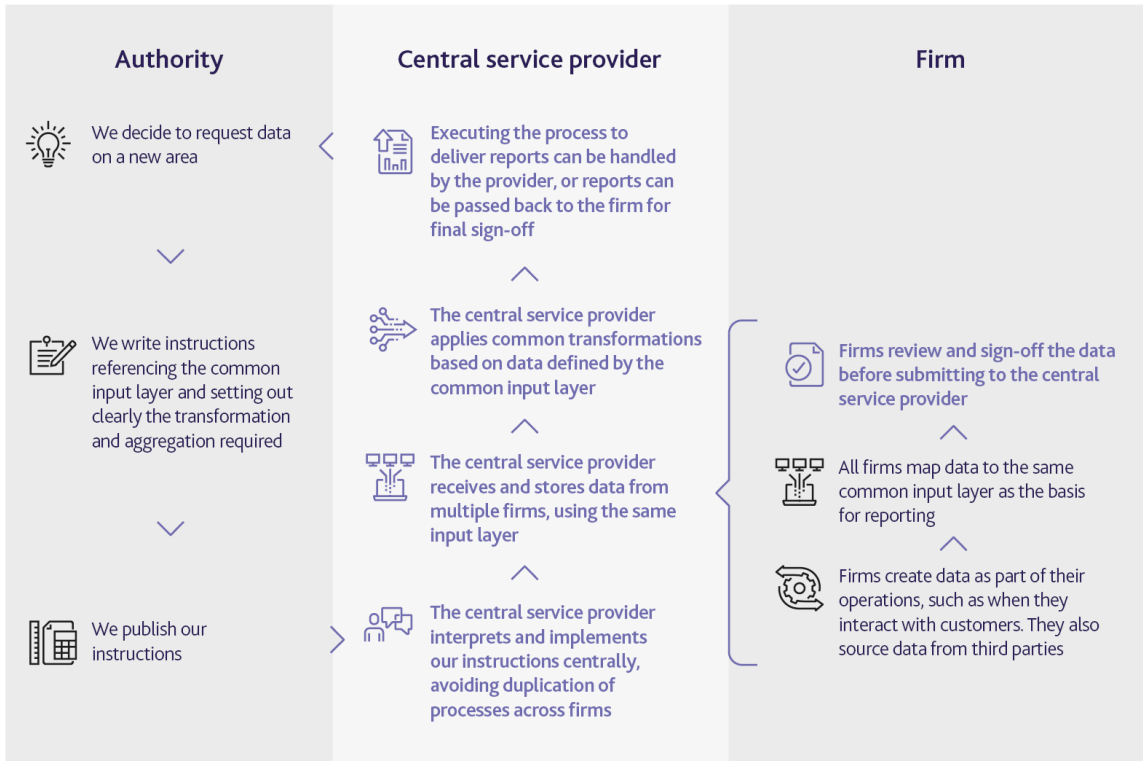
8.7 A pull model may work better for some data collections than others. A pull model could help where we collect the same or similar data for a number of purposes. It also could help mitigate some of the issues of collecting transactional or product level data. Firms would still need to take responsibility for providing correct data, which could raise challenges around providing assurance on constantly changing figures. For aggregate financial accounting or capital data in particular, it is likely to remain more appropriate for firms to ‘push’ data once the figures have been formally signed-off. Firms may want greater control over access to such data, and the scope for reuse of data may be lower.

Central services in reporting

8.8 Chapters 6 and 7 explained that some of the work in mapping from firms’ own data to reporting data points could be done centrally if a standardised input layer existed. Having agreed transformations, firms could implement these themselves. An alternative would be for some of these reporting processes to be carried out by a central service provider. The main international example of this is the system that exists in Austria, as set out in Chapter 5, where the major banks came together to build their own utility, with the endorsement of the central bank. Other ownership and governance models are also possible.

8.9 A central service provider could play a variety of possible roles. For instance, they could transform firm source data into the data needed for reports; or they could actually convert the data into the required reports. The service provider could play a similar role to solution vendors today, but service the whole industry rather than individual firms. The service provider could centralise just a part of the process. For instance, they might only centralise the interpretation of reporting instructions. Alternatively, individual firms could map from their source data to an input layer in a central utility, which could then transform the data and provide reports to the Bank – as shown in Figure 11. This approach could also be combined with an API, to allow the Bank and other authorities to pull data from the utility as required.

Figure 11: Introducing a central service provider



Benefits and costs

8.10 The service provider could save costs across the system by centralising processes (such as the translation of reporting instructions) and avoiding these being duplicated across firms. However, there would be offsetting costs involved in running the service provider.

8.11 By standardising the transformation of data or interpretation of instructions and carrying this out in one place, the quality of reported data may increase. Any errors could be fixed once, centrally, for the whole industry.

8.12 A central service provider may also offer the opportunity to use the standardised data it collects to feed valuable information back to firms, or to improve public disclosure. For example, the provider could publicly publish a subset of data in a central database, in an easily accessible manner, to facilitate peer analysis. This could help ensure that regulatory reporting provides value to firms themselves, as well as investors or policy analysts and researchers. While many jurisdictions provide some transparency around regulatory data, the US Federal Financial Institutions Examination Council provides an interesting example in terms of the depth, frequency and accessibility of the disclosures it produces.⁵⁹

Issues and challenges

8.13 Having an industry-wide central service provider could require changes to governance arrangements. Holding large amounts of granular data in one place would bring data security and management risks. Responsibility for these risks would need to be addressed. The reporting rules

⁵⁹ <https://www.ffiec.gov/npw>.

may need to change to require the use of a utility. Rules may need to consider explicitly what happens if the service provider creates an error in reports. The impact on the broader software service industry would also need to be considered.

8.14 Firms would probably still need to be involved in the sign-off and submission of at least some key reports. Without this, firms may lose sight of the data being used by regulators to assess their firm. This could make it harder for firms to respond to Bank queries. The current sign-off process gives firms a chance to understand the data they are submitting before sending reports to the Bank. Firms may also want to adjust the data before submitting. Firms sometimes manually fix data quality issues once at an aggregate level; with a central provider they could instead be forced to fix manually a large number of underlying data points.

8.15 As with a pull model, some data collections may be better suited to this type of architecture than others. Reports that are simple aggregations of granular source data, often found in statistics reporting, will probably be easiest to provide centrally. Data that require firm-specific judgement cannot be wholly calculated centrally, for instance in estimating provisions.

The Bank or another authority as central service provider

8.16 A potential variant could be for the authority itself to perform some of the roles of a central service provider, in particular the collection of a wider range of granular data. In doing so, the authority would take over the responsibility for transforming granular data, replacing the need for firms to produce and sign-off some reports and potentially increasing the flexibility and speed of analysis.

8.17 Many of the issues for a central service model would remain if the Bank performed such a role, and some could be exacerbated. The Bank is primarily a policy-making and regulatory institution, not a technology or data services provider. Handling and processing large quantities of granular data is historically not its core specialism. Such a role would also pose significant reputational risks to the Bank's wider objectives, for example in case of a security breach. There could be issues around firms losing sight of how their data were being used, posing challenges for data quality assurance. This approach would again not be acceptable for reports that evidence a firm's compliance with regulatory requirements, such as capital ratios. For such reports, the Bank will continue to need to be able to hold a firm accountable for its calculations.

Questions:

- K. What are your views on the benefits and challenges of the possible changes to architecture and governance set out in the paper – in particular moving to a “pull” model for certain types of data, or moving some functions to a central service provider?

9 Next steps

9.1 This paper sets out a range of options to help start a conversation on transforming data collection. It does not identify the Bank's preferred solution and it does not claim to be comprehensive – we are particularly interested in suggestions for approaches we have not considered here. In setting out some options, we recognise that the case for and against each will depend on a range of factors for which we currently have incomplete information including:

- The size and nature of costs associated with the current approach, in particular how these map to different steps of the reporting process and the balance between costs for new reports and run costs for existing reports. Questions in Chapter 3 seek further information on these points.
- Views on the feasibility, benefits and issues around the options put forward. Questions in Chapters 6, 7 and 8 seek initial views on these points, but any promising solutions are likely to require further detailed assessment in collaboration with firms, before proposals could be made.

9.2 Any of the solutions could in principle be adopted for a sub-set of firms or reports, leading to differentiated solutions or a phased approach that could expand over time as experience and evidence of benefits grows. One approach could be to target first those collections where solutions are potentially easier to apply, such as returns based on aggregations of contract level data in contrast to judgement-heavy regulatory returns. An initial focus on such collections would be in line with the approach of other jurisdictions. It would, however, be important to assess at what point a phased approach could start to unlock significant value, and how to maintain coherence across the Bank's collections. In choosing which firms or reports to include, it would also be necessary to be able to demonstrate a proportionate approach consistent with the PRA's secondary competition objective.

9.3 Transformation of the Bank's data collections will be most effective if it complements wider trends and initiatives affecting the storage and use of data in financial firms. This includes international initiatives, since many of the UK's larger financial firms are active in multiple jurisdictions and required to report to other regulators. As part of the follow-up exercise to this paper, the Bank will seek to identify and engage with other relevant UK and international initiatives. This includes continued participation in the Financial Stability Board's work to consider approaches to avoid future fragmentation, which 'include the possibility to promote greater use of common elements in supervisory data'.⁶⁰

9.4 Responses to the questions posed on page 6 and any other observations that readers may have in response to this DP should be sent by email to DatacollectionDP@bankofengland.co.uk by 7 April 2020. Responses and input are welcome from a wide range of stakeholders including regulated firms, industry bodies, specialist third-party providers, professional advisors, standards bodies and other regulators. The privacy policy is set out at the beginning of this document. Responses may be shared with the FCA.

9.5 As well as seeking written responses to the discussion paper, the Bank intends to establish one or more industry working groups to explore these issues during 2020. Other engagement channels

⁶⁰ <https://www.fsb.org/wp-content/uploads/P141019.pdf>; The FSB's Report on Market Fragmentation in June 2019 also noted that 'Significant differences in data reporting requirements and obstacles to information sharing across jurisdictions can increase the compliance cost associated with financial institutions' cross-border operations': <https://www.fsb.org/wp-content/uploads/P040619-2.pdf>.

will ensure wider input from those unable to participate directly, and draw on the work of existing groups working on reporting and data standards. The Bank is particularly mindful of the need to ensure solutions can be implemented in a proportional way across all the sectors it interacts with.

9.6 One focus for early industry discussion will be to inform the scope and aims of the next phase of the Bank/FCA Digital Regulatory Reporting Pilot. The completed pilots have provided valuable information and experience with some of the options under consideration. We would propose the next stage of DRR should aim to inform or align with the likely future direction of both the Bank's and the FCA's data collection strategies.

9.7 Subject to responses to this paper, the Bank's aim for the working group(s) over the course of 2020 would be to develop a collective vision for data collection reforms over a 5-10 year horizon, and proposals for immediate next steps that would move from pilots to live implementation. We expect to publish an update on responses and the proposed next steps during 2020.

Appendices

Appendix 1: Glossary of terms	49
Appendix 2: Bank of England data collections	51

Appendix 1: Glossary of terms

Ad hoc data collections

Data collections that are not regularly recurring, including one-off requests.

Annotated instructions

Text (e.g. a reporting rule) that has been marked-up or ‘tagged’ with specific meta-data represented using a mark-up language to enable a machine to consume, process or present information in way that enables a richer and more efficient user experience for humans to search and navigate that text.

Application programming interface (API)

A set of functions and procedures allowing the creation of applications to connect and share information with another application or service. In the context of this paper, an API could allow the Bank to interact directly with firms’ systems to request and receive certain data in an automated way.

Central service provider

An independent entity that acts as a central interface between firms and an authorities, industrialising the data reporting process. This could include interpreting reporting instructions, providing common interfaces, or transforming firms’ source data into reports or analytical outputs.

Common data inputs / common input layer

A common set of data points, defined and recorded consistently across firms, from which reports and specific data queries required by the Bank can be created. The scope and granularity of the set of data points can vary according to the requirement but could for example be defined at the level of single contracts or deals like loans, deposits and securities.

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. The paper uses this term in particular to refer to industry data standards, where definitions are developed and agreed by industry and used by firms in in both operational systems and for regulatory reporting. Standards could help organisations to consistently use and publish data, as well helping change markets, create open ecosystems and implement policy objectives.

Digital Regulatory Reporting (DRR)

A project undertaken between the FCA, the Bank and firms to explore how technology could make it easier for firms to meet their regulatory reporting requirements and improve the quality of information they provide.

Granular data

Data defined at lowest appropriate level possible for a given data set, for example relating to individual contracts or transactions.

Instructions as code

Reporting instructions represented by a programming language and published in a form enabling integration into and execution by a system.

Operational data

The data produced or used during the day to day running of a firm's business operations, for example recording customer data, insurance contracts or mortgage lending.

Pull method of data collection

The Bank initiates a process for extracting reports from or querying data held at individual firms or at a central service provider.

Push method of data collection

Firms submit reports to the Bank either on a regular scheduled or ad-hoc basis via online form, file upload or business to business transfer.

Regulatory reporting

Data received under a range of UK and EU legislation to allow the Bank to fulfil its functions as supervisor, regulator, macroprudential authority and 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.

Solution vendor

A firm that provides solutions and services to financial services firms. Such solutions or services can integrate with internal systems and data sources e.g. operational to enable the population and creation of reports provided to the Bank.

Statistical reporting

Reporting received under the Banking Act to inform monetary policy. Aggregated and published to meet standards set out in the Bank's Statistical Code.

Appendix 2: Bank of England data collections

The table below summarises the Bank's regular structured data collections, to illustrate the scope of this review. This is not an exhaustive list of data collections, and notably does not include ad hoc collections.

EU mandated collections

Link to forms and definitions

EBA Regulatory reporting data collections for banks, building societies and designated investment firms

<https://eba.europa.eu/regulation-and-policy/supervisory-reporting>

EIOPA Regulatory reporting data collections for insurers

<https://eiopa.europa.eu/regulation-supervision/insurance/solvency-ii>

PRA Rulebook collections

PRA Regulatory reporting data collections for banks, building societies and designated investment firms

<https://www.bankofengland.co.uk/prudential-regulation/regulatory-reporting/regulatory-reporting-banking-sector/banks-building-societies-and-investment-firms>

PRA Regulatory reporting data collections for insurers

https://www.bankofengland.co.uk/prudential-regulation/regulatory-reporting/regulatory-reporting-insurance-sector#pra_reporting

Regulatory reporting data collections for credit unions

<https://www.bankofengland.co.uk/prudential-regulation/regulatory-reporting/regulatory-reporting-banking-sector/credit-unions>

Other collections based on legislative powers

Statistical reporting data collections for banks and building societies

<https://www.bankofengland.co.uk/statistics/data-collection/osca/forms-definitions-validations>

Buy-to-let data collection

<https://www.bankofengland.co.uk/statistics/data-collection/beeds>

Sterling money markets data collections

<https://www.bankofengland.co.uk/statistics/data-collection/sterling-money-markets>

Resolution planning

<https://www.bankofengland.co.uk/prudential-regulation/publication/2018/resolution-planning-mrel-reporting>

Annual concurrent stress testing

<https://www.bankofengland.co.uk/stress-testing>

Other recurring collections

Banks exposures

BCBS – Quantitative impact surveys

Corporate leveraged lending data

EBA – Funding plans projections

Enhanced asset reporting

FSB – data collections for globally systemically important banks (GSIBs)

Hedge fund as counterparty survey

Intra-day liquidity risk questionnaire

Intraday liquidity template

Leverage across securities financing

Longevity improvements data

Monthly treasury asset exposures

Mortgage arrears data

Other systemically important institution (O-SII) identification

Settlement internalisation

Small bank and building society basis risk

Small firms loan book data

Solvent wind-down

UK Central Counterparty (CCP) data

FCA Handbook data collections used by the Bank of England

Mortgage product sales data

<https://www.handbook.fca.org.uk/handbook/SUP/16/11.html>

Second charge regulated mortgage activity

<https://www.handbook.fca.org.uk/handbook/SUP/16/Annex19AA.html>