EBA | EUROPEAN BANKING AUTHORITY

EBA/DP/2021/01

11 March 2021

Discussion Paper

On a Feasibility Study of an Integrated Reporting System under Article 430c CRR

Contents

List of figures

Abbreviations

API	Application Programming Interface
BCBS	Basel Committee on Banking Supervision
BIRD	Banks' Integrated Reporting Dictionary
CDCP	Central Data Collection Point
CoC	Cost of Compliance
Corep	Common Reporting
COTs	Commercial-off-the-shelf
CRD	Capital Requirements Directive
CRR	Capital Requirements Regulation
CSV	Comma separated values
DPM	Data Point Model
EBA	European Banking Authority
ECB	European Central Bank
EIOPA	European Insurance and Occupational Pensions Authority
ESA	European System of Accounts
ESCB	European System of Central Banks
ESMA	European Securities and Markets Authority
ETL	Extraction, Transfer and Load
EUCLID	European Centralised Infrastructure for Supervisory Data
Finrep	Financial Reporting
GDPR	The General Data Protection Regulation
IReF	Integrated Reporting Framework
JC	Joint Committee
JEGR	Joint Expert Group on Reconciliation of credit institutions' statistical and supervisory reporting requirements
MoUs	Memorandums of Understanding
RACI	Responsible, Accountable, Consulted, Informed model
RegTech	Regulatory technology
SDD	Single Data Dictionary
SDMX	Statistical Data and Metadata eXchange

SNA	System of National Accounts
SRB	Single Resolution Board
SSM	Single Supervisory Mechanism
XBRL	eXtensible Business Reporting Language
XML	Extensible Markup Language

Executive Summary

Article 430c of the Capital Requirements Regulation (CRR) mandates the EBA to prepare a feasibility study for the development of a consistent and integrated system for collecting statistical, resolution and prudential data, as well as to involve the relevant authorities in the preparation of the study.

The overall objective of the feasibility study is to identify the feasibility of various ways to streamline the reporting process and increase efficiencies going forward. The report should be neutral in regard to both the options and to the institutional set up and simply identify the various options and the feasibility thereof. It is important that the report does not draw conclusions or make proposals.

To achieve this, the report considers the principle commonly referred to as 'define once, report once', which means the report will consider the feasibility of a common language and dictionary for 'define once' and analysis of architectures to achieve 'report once'. The feasibility study will analyse the feasibility of various options for creating an integrated reporting system as a way forward to increase efficiencies and reduce costs for the entire reporting ecosystem, more specifically:

- reduce the reporting costs for the reporting institutions;

- streamline and increase the efficiency in the reporting processes;

- facilitate data comparability, remove overlaps (following the 'define once' principle);

- facilitate data sharing and increase coordination among authorities (following the 'report once principle').

To achieve these objectives, the analysis and options considered so far by the EBA, described in this discussion paper and which will be further reflected in the final feasibility study, considers the needs of various stakeholders that might be impacted by the possible creation of such an integrated system. The assessment covers the **four core areas**, which are considered key elements in the creation of the integrated reporting system and have already been identified by Article 430c of the CRR:

- the stocktake on the quantity and scope of the current data collected;

- the use of a common data dictionary;

- the feasibility and possible design of a central data collection point;

- the governance and the establishment of a joint committee.

Each of the four core areas are essential components of a reporting process and their design, as part of an integrated system, would affect the way the current reporting process is conducted. To this end, the EBA has assessed the four core areas taking a holistic approach and considering the impact of integration **at each step in the reporting process chain** (data definition, data collection, data transformation and data exploration) **along the different levels of abstraction** (from the high-level business concepts and logic – semantic level, through the formal and standardised formats – syntactic level, to the more tangible components of the technological architecture – infrastructure level, which enables the use of data by final users).

The report outlines the potential benefits and costs that an integration might have at the different steps of the reporting process chain, as well as any possible interdependencies between the different steps of the reporting process chain. Potential benefits might include an increased understanding of the different reporting requirements, reusability and interoperability of the different requirements and having a more efficient process, while on the other side investments in both resources and time will be necessary for all stakeholders to implement and maintain the framework. Besides, this analysis takes also learning curve effects into account. Given the interlinkages between each of the different steps alongside the reporting process, costs and benefits might be different if different components would be integrated together or independently.

A common data dictionary for prudential, statistical and resolution data collections is considered a central piece all along the reporting process chain in this discussion paper. The data dictionary would offer a common understanding of the reporting requirements and support their design in the view of the 'report once' principle. A dedicated section in the report highlights the principal objective of the data dictionary, data comparability, and sets out the characteristics and requirements for the common data dictionary in an integrated reporting system. An assessment of benefits, costs and challenges related to the introduction of the common data dictionary is specified in the discussion paper. However, based on the analysis it is clear that the dictionary should take into account the existing differences between the datasets (such as the different levels of integration and granularity in the data sets) as well as the differences in the underlying regulations.

The topic of granularity can be considered in connection to the data dictionary analysis. Three scenarios for collecting data on a more granular basis were identified and assessed in the report. In general, a more granular reporting might enable the collection of data only once, increase the consistency and reusability of the data and entail efficiency gains. However, when assessing the feasibility of moving towards a more granular data collection, a series of legal and technical challenges should be carefully considered. These challenges include, amongst others, the handling of the transformations, the responsibilities of reporting institutions and authorities in terms of transformations, the interactions between transformation rules and principle-based policy (such as accounting policies), the impact on the data quality assurance processes and proportionality aspects.

An overview of the costs and benefits identified during the abovementioned feasibility assessments on the reporting process chain, the common data dictionary and the granularity can be found in Section 6 of the discussion paper.

The analysis on whether the creation of a Central Data Collection Point (CDCP) is feasible starts from a predominantly technical focus by putting forward possible design choices from a range of system architectures and topologies.

Additionally, the report explores two different data reporting architectures for data collection, the 'Push' and 'Pull' approaches, with the aim of gathering more evidence on the implications and the potential costs and differences of these two approaches.

The assessment is not proposing any specific approach or solution, but rather is limited to describing the various options that could lead to an integrated reporting system and achieving the stated objectives. Key aspects like resources, costs or necessary investments for authorities and institutions will have to be analysed in detail once the specific scenarios for the integration are defined.

The table below provides a summarised view of the analysis and options linked with the different core areas of analysis considered in this discussion paper:

Discussion paper	Analysis and options
Stocktake	- Overview of data requests and analysis on impacts of further integration of data requests
Data dictionary	- Feasibility of a possible design of a unique and common data dictionary that facilitates data comparability and removes overlaps (following the 'define once' principle. - Main characteristics of a common data dictionary. - Differences between data frameworks and the current level of integration. - Data granularity, analysis of different options and the feasibility of increasing the granularity of the data collected. - Assessment of potential costs and benefits.
Central data collection point	- Identification of possible high-level design options - Analysis of these design options against the mandate for the central data collection point. - Assessment of potential costs and benefits.
Governance	- Identification of relevant aspects that would need to be defined and agreed to ensure that any future integrated reporting system could be implemented and managed. - Proposal for a coordination mechanism to facilitate data sharing and increase coordination among authorities (following the 'report once principle'). - An overview of Joint committee role in the integrated reporting system.

The discussion paper serves as a basis for further discussions and aims to gather additional evidence and opinions and does not make concrete proposals. The EBA will prepare the Feasibility Study taking into account the feedback received. The series of questions for which the EBA seeks further evidence are contained in Section 10, and have also been highlighted in the various sections of the report to which they refer. Additional workshops and bilateral meetings may be organised to further clarify certain aspects, if needed.

The final outcome of the feasibility study on integrated reporting will serve as a basis for further decision-making both in terms of orientation and timeline for implementation.

Next steps

The EBA invites stakeholders to share their views and provide feedback on the analysis presented in this discussion paper and on the questions included in Section 9. The feedback sought on this discussion paper will inform the EBA's final report, the Feasibility Study of an integrated reporting system.

Responding to this discussion paper

The EBA welcomes comments on this Discussion Paper on a Feasibility Study of an integrated reporting system and in particular on the specific questions set out in Section 10. The EBA is looking to receive feedback from reporting institutions and other stakeholders that believe they might be impacted by any topic or option outlined in this discussion paper or that might have relevant information that would help to form a complete picture on them.

Comments are most helpful if they:

- respond to the question stated;

- indicate the specific question or point to which a comment relates;

- are supported by a clear rationale;

- provide evidence to support the views expressed / rationale proposed;

- describe any alternative choices the EBA should consider.

Submission of responses

To submit your comments, click on the 'send your comments' button on the consultation page by June 11, 2021. Please note that comments submitted after this deadline or submitted via other means may not be processed.

Publication of responses

Please clearly indicate in the consultation form if you wish your comments to be disclosed or to be treated as confidential. A confidential response may be requested from us in accordance with the EBA's rules on public access to documents. We may consult with you if we receive such a request. Any decision we make not to disclose the response is reviewable by the EBA's Board of Appeal and the European Ombudsman.

Data protection

The protection of individuals with regard to the processing of personal data by the EBA is based on Regulation (EC) No 45/2001 of the European Parliament and of the Council of 18 December 2000 as implemented by the EBA in its implementing rules adopted by its Management Board. Further information on data protection can be found in the legal notice section of the EBA website.

1. Introduction

1.1 Background

1 Since the financial crisis, new reporting requirements have been recognised as key for prudential, resolution and statistical purposes. Reporting requirements provide authorities at EU and national level with the information they need to fulfil their mandates, contributing to the wider objectives of financial stability, market integrity and consumer/investor protection in the EU single market for financial services, conduct monetary policy and provide a macro-prudential oversight of the financial system, as well as ensure the resolvability of institutions when needed.

2 The current EU reporting ecosystem consists of many different actors (reporting entities and authorities) and reporting frameworks (supervision, statistics, resolution, etc.), including the different national, European and international requirements. The separate development of reporting frameworks to fulfil different needs has sometimes led to overlaps in the data requested. In addition, ad hoc requests and national requirements have offset some of the benefits of a single set of harmonised reporting requirements across the EU (Commission call for evidence report, 2017[1]).

3 While harmonised reporting requirements have been welcomed by institutions, the banking industry has also argued that banks are required to fill in multiple templates in which data points partly overlap and definitions differ although they could be harmonised. Furthermore, these requirements have been reported at different frequencies, with a different scope of consolidation and to different authorities. The industry has therefore called for more coordination and data sharing among authorities to avoid overlapping requests.

4 The EBA acknowledges these challenges and is working on improving efficiency of reporting and enhancing proportionality of reporting requirements. The EBA published in November 2019 a Roadmap on Risk reduction measures package[2] outlining an action plan on prudential reporting.

5 The EU legislators have considered these concerns about the reporting costs for reporting institutions and the need to improve the efficiency of the reporting process and introduced amendments to the Regulation 575/2013 (Capital Requirements Regulation 'CRR [3] '). In

---

[1] REPORT FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS, Follow up to the Call for Evidence - EU regulatory framework for financial services, 1.12.2017, https://ec.europa.eu/info/files/171201-report-call-for-evidence_en

[2] EBA Risk Reduction Package Roadmaps

[3] Regulation (EU) 2019/876 of the European Parliament and of the Council of 20 May 2019 amending Regulation (EU) No 575/2013 as regards the leverage ratio, the net stable funding ratio, requirements for own funds and eligible liabilities, counterparty credit risk, market risk, exposures to central counterparties, exposures to collective investment undertakings, large exposures, reporting and disclosure requirements, and Regulation (EU) No 648/2012 (Text with EEA relevance.)

particular, the EBA has been mandated in Article 430c to prepare a feasibility study for the development of a consistent and integrated reporting system for statistical, resolution and prudential data.

6   This discussion paper has benefited from discussions with the ECB, the SRB, competent authorities and the industry, helping to identify the main aspects that should be analysed by the feasibility study (fact-finding workshops and seminars). The EBA has been working together with the ECB and the ESCB to analyse more in depth the topic of data dictionary, granularity and ad hoc requests. The EBA has prepared this discussion paper as a starting point for discussions, in response to the mandate to invite stakeholders to provide their feedback and views on the analysis and proposals.

7   The EBA will prepare the final feasibility study taking into account the feedback received. Additional workshops and bilateral meetings may be organised to further clarify certain aspects, if needed.

1.2   Preparation of the Feasibility Study

1.2.6   Objectives

8   The overall objective of the feasibility study is to identify ways to streamline reporting requests and processes to reduce the financial and administrative burden and increase efficiencies going forward and considering the principle commonly referred to as 'define once, report once'. To achieve this there is a need to have a common language and dictionary for 'define once' and analysis of architectures to achieve 'report once'.

9   To achieve the objective of the feasibility study, the EBA has chosen to follow a holistic approach, as presented in this discussion paper, and analyse reporting integration throughout the reporting process chain from data definition to data exploration to ensure feasibility of scenarios and identify interdependencies.

10   Following the mandate in Article 430c CRR, the EBA aims to identify shortcomings with a view to reduce the administrative and financial costs, both for the authorities and for the institutions and to improve the overall efficiency of the statistical, resolution and prudential reporting process. In addition, the discussion paper assesses the different options for the integration of the different data and the feasibility of the development of a consistent and integrated system for collecting statistical data, resolution data and prudential data.

11   For the Feasibility Study, the EBA will build on the objectives of an integrated system, which include:

- increasing the efficiency of reporting by standardising reporting, reducing redundancies and using common definitions;

- increasing efficiencies for financial entities and authorities;

- facilitating the exchange of data and its usability;

- improving data quality.

12 To achieve these objectives, understanding the cost drivers of institutions' reporting processes and how to improve the usability of data for the public sector are key.

13 This discussion paper, that will inform the feasibility study, does not describe a precise design or implementation of a specific solution or a system but explores a series of possible options (that might increase the efficiency of the reporting process and might be feasible to implement) with the aim of gathering more evidence on their feasibility and the costs and benefits they entail for both reporting institutions and authorities. The final feasibility study will describe pros and cons of options and scenarios accounting also for the feedback received on this discussion paper.

1.2.7 Scope and areas covered

14 Following the mandate, the scope of the feasibility study covers the analysis of prudential, statistical and resolution data reported by institutions. The focus in this discussion paper is on credit institutions' reporting.

15 The actual reporting ecosystem has been analysed in this discussion paper in order to understand the impacts of different types of data in the reporting process. This discussion paper, that will inform the feasibility study, focuses on four core areas of analysis, which are considered key in the feasibility assessment of an integrated reporting system and along which the report aims to tackle some of the issues on reporting costs and the inefficiencies identified by reporting entities.

Figure 1: Overview of core areas analysed in the discussion paper
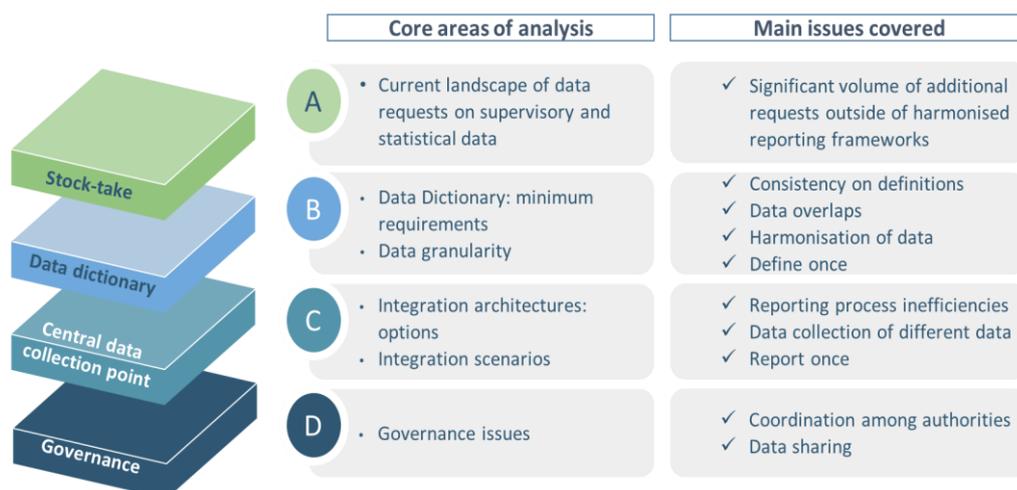
16 The discussion paper covers the different areas under analysis with a holistic perspective, taking into account the impacts of integration in the entire reporting process chain. The methodology used for this assessment is detailed under Section 3.

17 The discussion paper is not proposing any specific approach or solution but rather limits itself to describing the technical requirements in order to achieve an integrated reporting system, analysing it from a technical and architectural perspective and providing different options from a technical point of view. While general governance considerations have been highlighted, further analysis on governance implications should be tackled once a specific path is defined for the common data dictionary and central data collection point.

18 Article 430c stipulates that the feasibility study to be developed by the EBA shall be based on an overall cost and benefit analysis. Impacts on the different stakeholders and the changes to the processes along the regulatory data chain are discussed as part of the individual sections in this discussion paper, providing insights into the costs and benefits for both authorities and institutions.

19 Cost estimates in monetary terms are not feasible at this stage. Costs and benefits are hence described in a qualitative way and at a high level. Further analysis of the costs will need to be conducted at later stages of the project, also taking into account the feedback received on this consultation paper (see the relevant questions in Section 10). In addition, Section 6 summarises the main costs and benefits identified in the various sections of this discussion paper).

1.2.8 Stakeholders' involvement

20 The scope of the feasibility study covers many reporting frameworks, actors and processes that could be impacted following the completion of the feasibility study. Interaction with these stakeholders is of utmost importance for development of the feasibility study. Their experience and views are key in order to understand the different implications across the different steps of the reporting system and to find feasible and efficient ways forward.

21 The EBA is involving authorities, as well as the ECB and the ESCB, in the development of the feasibility study. The ESCB work on integrated data collections has been key in the analysis and its input report published in September 2020. The cooperation with the ECB has been key in order to understand other frameworks under development such as the statistical reporting framework (Integrated Reporting Framework (IReF)) and other initiatives on data dictionaries such as BIRD.

22 For the preparation of this discussion paper, the EBA completed a fact-finding and research phase, which benefited from experiences of various authorities in integrating reporting and private sector stakeholders' views on current challenges and potential solutions.

23 The EBA will further continue the collaboration with the different stakeholders involved and is interested in receiving views from external stakeholders about the assessment proposed in this discussion paper. Similarly, further discussion and exchange of views between competent

authorities, resolution authorities, statistical authorities, as well as Deposit Guarantee Schemes will be of high importance to ensure common understanding and views on feasibility of an integrated system. Stakeholders are invited to provide their response to the questions inserted in Section 10. The EBA will use the responses to these questions to inform its future work and prepare the final report on the feasibility study of the integrated reporting system.

24 The EBA will also build the analysis on evidence gathered during the Cost of Compliance study on prudential reporting.

1.2.9 Links with other EU initiatives and projects

25 Significant efforts are already ongoing to simplify and streamline prudential reporting:

- **EBA ongoing work on data integration:** the EBA is working on a further integration of prudential reporting requirements with related public disclosure requirements within their remit, with a view to reducing the burden and facilitating banks' compliance with both. In addition, the EBA has a mandate to continue enhancing proportionality and perform an analysis on the costs of reporting which will be provided in a Cost of Compliance report[4] with the aim of finding ways to reduce reporting costs primarily for small and non-complex institutions. This study is performed in accordance with the specific mandate of Article 430(8) CRR. Evidence from the cost of compliance study has been used as input for this discussion paper.

- **SRB cooperation:** the EBA and the SRB are cooperating closely on resolution reporting with the aim of creating an integrated and harmonised set of requirements covering both prudential and resolution reporting requirements.

- **ECB initiatives**: In the related area of statistical reporting, the ECB is working on developing common definitions and data models, in particular through two ongoing projects. The first is the ESCB IReF[5], which aims to integrate existing statistical data requirements for banks into a unique and standardised reporting framework that would be applicable across the euro area. It focuses in particular on requirements of the ECB's regulations on monetary financial institutions' balance sheet items and interest rate statistics, securities holdings statistics and bank loan reporting (AnaCredit). The current aim is to implement the IReF by 2024-27. The other project, entitled 'Banks Integrated Reporting Dictionary' (BIRD)[6], aims to help banks organise information stored in their internal systems more efficiently in order to better fulfil their reporting requirements. BIRD is a harmonised data model that precisely describes the data to be extracted from the banks' internal IT systems to derive reports required by authorities. The methodology which serves as the basis for the construction of the BIRD metadata is the

---

[4] https://www.eba.europa.eu/regulation-and-policy/supervisory-reporting/cost-compliance-supervisory-reporting

[5] https://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/reporting/html/index.en.html#IReF

[6] https://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/reporting/html/index.en.html#BIRD

SMCube Information Model[7]. The BIRD currently covers the reporting requirements of AnaCredit, the group module of ECB statistical reporting of securities holdings, as well as financial reporting (FINREP). The coverage of Common Reporting (COREP), asset encumbrance and resolution planning is currently under development.

- **European Commission initiatives**: the Commission has performed a fitness check of prudential reporting requirements in the EU financial services legislation[8]. The fitness check identified a number of areas where there is scope to further simplify and streamline the reporting process such as: i) improve the legislative design of primary legislation; ii) assess the data needs and its uses; iii) greater consistency and harmonisation; iv) governance related to further coordination at earlier stages of the reporting process and data sharing between authorities; v) technological developments that could provide new opportunities to streamline the reporting process.

- The Commission has launched major initiatives around data such as the European Data Strategy[9] and the Digital Finance Strategy[10].

- As part of its data strategy, the Commission is performing follow-up work based on the findings of the fitness check, in order to set out a long-term vision for moving from the current system of prudential reporting to a modern, efficient and effective reporting process.

- The EBA is in close interaction with the Commission in order to provide the report on the feasibility study of the integrated reporting system and contribute to the assessment performed by the Commission on the long-term action plan for an efficient prudential reporting process.

26 The common goal of all these initiatives is to improve the effectiveness of supervision, resolution and statistical data production while also reducing the compliance burden for institutions. The feasibility study will take into account the different existing efforts in order to assess those areas, which could be integrated, and which areas are more challenging from the integration perspective.

1.3   Discussion Paper structure

27 In Section 2 of this discussion paper, a **stocktake analysis** has been included as a key first step in order to gather the necessary information on the current reporting frameworks and to have a good overview of the current reporting landscape in the EU. In addition to the harmonised prudential, resolution and statistical reporting frameworks for institutions, national authorities request data on a regular and ad hoc basis (for either prudential, resolution or statistical

---

[7] https://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/smcube/html/index.en.html)

[8] https://ec.europa.eu/info/publications/191107-fitness-check-supervisory-reporting_en

[9] https://ec.europa.eu/digital-single-market/en/european-strategy-data

[10] https://ec.europa.eu/info/publications/200924-digital-finance-proposals_en

purposes). According to the information provided by national authorities, additional requests for prudential purposes are mostly on a quarterly or monthly basis and principally relate to the areas of credit, counterparty and liquidity risk. The stocktake suggests that there might be benefits in the integration of non-harmonised (national and ad hoc) data requests into an integrated reporting framework.

28 Section 3 of the discussion paper focuses on the **impact of integration at each step in the reporting process chain** (definition, collection, transformation and exploration). Different levels of abstraction are considered, from the high-level business concepts and logic (semantic level), to the formal and standardised formats (syntactic level), to the more tangible components of the technological architecture (infrastructure level).

29 Section 4 covers the common **data dictionary**, one of the core aspects of integration. While the feasibility of creating a common data dictionary was assessed to some extent in Section 3 by looking at its role in integration at different stages in the reporting process. Section 4 is more focused on highlighting the principal objective of data dictionary, data comparability. A common data dictionary would offer a common platform of understanding the reporting requirements and support their design in view of the 'report once' principle. Drawing on the European and national experiences with data dictionaries (analysing their specific purposes, principal components, how they are managed and how they support the data lifecycle), the characteristics and requirements of the common data dictionary are defined. The conclusions of the European Commission's Fitness check on supervisory reporting in the EU, investigating the role of data dictionary to solve the issues identified by financial institutions have been duly accounted for.

30 The topic of **data granularity** has been considered as part of the data dictionary analysis. Section 4.6 investigates the feasibility of increasing the granularity of the data collected with the purpose of: i) meeting authorities' objectives and facilitating institutions' compliance with reporting requirements (increasing the efficiency of the reporting process) and ii) facilitating the integration of prudential, statistical and resolution data. Increasing the granularity of the data collected would be done in the spirit of the 'define once' and 'collect once' ('only once') principle, as one data point would be collected only once and used for different purposes (e.g. templates). This section puts forward for discussion three possible options for the granularity level of the future integrated reporting system. However, moving to a more granular data collection poses a series of legal and technical challenges requiring a careful balancing of the costs and benefits as well as a careful consideration of the possible design of the underlying processes (e.g. transformations). The discussion paper is looking to investigate deeper some of the key points identified for this topic.

31 In Section 5, the discussion paper explores the feasibility of setting up the **Central Data Collection Point (CDCP)** by putting forward for discussion possible design choices from a range of system architecture and topologies. The section looks at how specialised platforms and components could be organised to support the full extent of the reporting lifecycle by taking into account the European regulatory reporting network. The characteristics of each topology is

analysed from the point of view of compliance with the requirements highlighted in Article 430c (collection point, central data register / point of contact, standardised system, support of national reporting, cost-efficiency and technical feasibility). While this section has a predominantly technical focus, the full range of costs and benefits of setting up such a system should account for a broader set of costs and benefits including (but not limited to) governance aspects.

32 Section 6 explores two different architectures for data reporting, **the push and the pull approaches.** While this section is limited to presenting the main differences between the two models, the discussion paper is looking to gather more evidence on the implications of the different designs and to infer the costs and benefits of the two approaches for reporting institutions and competent authorities. A mixed approach may also considered.

33 Section 7 is dedicated to **Governance**, highlighting the relevant aspects that would need to be defined and agreed to ensure that any future integrated reporting system could be implemented and managed. General governance aspects have been considered for data submission, access to data and data sharing. In addition, this section explores how an agile coordination mechanism of data requests could look like asking for additional feedback on its design. Moreover, various processes of the integrated reporting system may be suitable for different approaches in terms of governance: centralised and/or decentralised. Additional work on the legal considerations and appropriate allocation of responsibilities surrounding governance and operational issues would need to be undertaken at a later stage, once a specific path is defined for the common data dictionary and central data collection point system.

34 Considerations on possible roles and tasks for the **Joint Committee** are included, however, the discussion paper points to the fact that setting up the Joint Committee could only be done once the feasibility study has concluded and its role and tasks should account for the conclusions of the study.

Questions to consider

1) Please explain which institutions you think should be considered by the Feasibility Study.
2) Please explain which data collections you think should be considered by the Feasibility Study.
3) Do you consider that the issues identified, the options proposed and the assessment approach taken throughout the discussion paper are relevant and complete? If not, please explain.
4) What do you perceive as the key obstacles and operational challenges to develop an integrated Reporting Framework (for your institution)?

2. Stocktake of current data requests

35 The first step in preparing a feasibility study of an integrated reporting system is gathering the information on the current frameworks, including the reporting systems and data requests. This will allow the EBA to have a good overview of the current reporting landscape in the EU, providing strong support for the further work in the feasibility study and enabling the understanding of the extent to which an integrated reporting framework could benefit both the competent authorities and the banks.

36 Credit institutions, financial and mixed financial holdings, banking groups and branches of EU and non-EU institutions regularly report prudential data within the EU harmonised EBA reporting framework. This framework also partially includes reporting for resolution purposes, which is broadened by the resolution reporting framework developed by the SRB. In addition, based on ECB regulation, institutions in the euro area regularly report statistical data necessary to carry out the tasks of the ESCB.

37 Authorities may collect additional data from institutions to address the data gaps related to a specific area or frequency of reporting or a bank's specific information. This includes data required for prudential, statistical, financial stability or other purposes. The stocktake provides an overview of the current data requests across the authorities in the EU, with the emphasis on those falling outside the scope of the harmonised EU-wide reporting.

38 This stocktake leverages firstly on a Single Supervisory Mechanism (SSM) inventory of data requests to supervised entities for prudential purposes. The exercise was mirrored for non-SSM countries with a similar request from the EBA in order to have a complete and comprehensive picture of the purpose, frequency, magnitude, areas and dispersion of the data requests to supervised entities in all EU jurisdictions. The joint SSM and non-SSM databases[11] are henceforth referred to as 'data inventory'. It should be noted that the data inventory does not allow for a characterisation of either the volume of data, or the number of data points defined in the collection, reported on average by a bank for a given period of time. Thus, the provided analysis is aggregate and does not reflect the situation for all banks in the jurisdiction, as the larger and more complex banks tend to report more data. Additionally, the analysis may be based on differing methodologies, varying by jurisdiction. For SSM countries, Less Significant Institutions' specific requests are not included in the data inventory as this information was not readily available. There may also be differences in interpretation of the scope of the data inventory, as some competent authorities considered mostly microprudential requests, while others examined the entire set of requests to banks. However, in order to reduce the burden on the

---

[11] The SSM inventory has a cut-off date of 24 November 2020, while the non-SSM data refers to a stocktake on 31 December 2019.

competent authorities, the EBA decided to use the available data as an approximation in this note.

39 The statistical requests are described using the information made available to the EBA by the ESCB [12].

2.1 Harmonised reporting frameworks

40 An overview by the number of defined data points of the harmonised reporting frameworks can be found in Table 1: Defined data points by framework It should be noted that due to different ways of collecting data in different frameworks, the numbers are not always directly comparable.

41 The banks' actual reporting in the EBA reporting frameworks differs from the picture presented by the defined data points, sometimes significantly, as around 10% of the defined data points correspond to open tables[13]. Thus, for example, Credit Risk Benchmarking and Large Exposures reporting comprises the majority of the reported values for some banks due to the open data points concept.

42 In terms of the actual values reported in the EBA reporting frameworks, banks on average[14] report roughly 60,000 values each for year-end reference periods, and from 30,000 to 40,000 values for other quarterly periods. Monthly reporting is limited to less than 10,000 values, reported within the liquidity monitoring frameworks. The largest banks might, however, report over 400,000 values per reference period.

Table 1: Defined data points by framework

Framework	Number of defined data points
COREP Own Funds	38,125
COREP Liquidity	26,739
FINREP	13,831
ESCB statistical data collections	8,286
Resolution (EBA and SRB)	3,428
PSD Fraudulent Payments	1,830
Asset encumbrance	1,299

---

[12] Some details of the statistical data requests may be found in the *ESCB input into the EBA feasibility report under article 430c of the Capital Requirements Regulation (CRR2)*, Annex 1, available at https://www.ecb.europa.eu/pub/pdf/other/ecb.escbinputintoebafeasibilityreport092020~eac9cf6102.en.pdf?743bc2d efe61abe865e1857ab1a98337.

[13] Open tables refer to a form of reporting where a bank could report multiple values for a single defined data point, such as for each obligor, currency or country.

[14] The average is referring to the EBA's reporting sample of banks, available here: https://eba.europa.eu/sites/default/documents/files/document_library//882861/SCOP%202020%2023%20rev1%20%2 8Update%20of%20EBA%20List%20of%20Institutions%20for%20Supervisory%20Reporting%29.xlsx. The average takes into account periods from 2014 to 2019.

Framework	Number of defined data points
Funding Plans	1,062
FINREP – COVID-19	600
Remuneration	380
COREP Leverage Ratio	276
Benchmarking of internal models	153
COREP Large Exposures	77
G-SII identification and buffer rates	17
Total[15]	**96,180**

43 Prudential information is collected from EU supervised entities [16] in the scope of the EU harmonised reporting framework, set up and maintained by the EBA. This information is used in banking supervision and also allows authorities to monitor trends and risks in the national and EU banking sectors. Here, as per the CRR, a principle of maximum harmonisation applies, meaning that national data requirements for banks may not deviate from what is prescribed on the EU level by the CRR and the relevant technical standards. The resolution information, collected to aid in resolution planning, overseen by the SRB, is partially included in the EBA harmonised reporting framework. The maximum harmonisation principle does not apply in the same way to resolution data, meaning that in addition to frameworks developed by the SRB and the EBA, national resolution authorities may request further data from the institutions. Constituting a third major pillar of the harmonised reporting frameworks, institutions[17] regularly report statistical data used for the tasks of the ESCB, among which the main function is the conduct of monetary policy for the euro area.

2.2 Additional data requests for prudential purposes

44 Additional data requests for prudential purposes are recurring or one-off data requests to institutions triggered by supervisors to address data gaps or monitor emerging risks. Some additional data requests are subsequently included in the EU-wide harmonised reporting, however as it stands, there are legal and practical constraints for such inclusions, limited by the link between harmonised EBA reporting and CRR /CRD, and the burden of additional reporting of data that might not be pertinent to the majority of EU banks.

45 Additional data requests are described and enumerated in the following chapters using the information provided in the data inventory. Here, any requests originating from the EU harmonised prudential reporting framework or authorities not acting as the institutions' supervisors, such as the EBA, are excluded. Where possible, a comparison is provided between the additional requests for prudential purposes and the EU harmonised reporting framework.

---

[15] The actual total is lower – the EBA and SRB frameworks together (without the ESCB statistical data collections) amount to 81,743 defined data points. This is because same definitions (and thus data points) are reused across frameworks.

[16] Credit institutions, financial and mixed financial holdings, banking groups and branches of EU and non-EU institutions.

[17] Statistical data is collected from credit institutions and other monetary financial institutions, such as central banks, money market funds and other financial institutions with monetary liabilities.

2.2.1 Number of additional data requests for prudential purposes

46 When considering simply the number of additional data requests, most of them, roughly 61%, are of a quantitative nature. Qualitative information is collected in 39% of the cases[18].

47 The inventory of data requests suggest that the scattering of the data collected is wide across the different jurisdictions. In terms of the number of data requests, this means that the institutions, depending on the jurisdiction, could be subject to only a few to dozens of additional data requests in addition to the harmonised EU-wide reporting. In the Single Prudential Mechanism, around 470 requests to significant institutions are triggered by the ECB or the Joint Supervisory Teams, which is complemented by almost 170 requests triggered by the national prudential authorities. The differences among the jurisdictions seem significant even when taking into account the different sizes of the banking sectors and the fact that there could be several banks in a jurisdiction requiring elevated prudential focus.

Figure 2: Number of additional data requests to banks by country (quantitative and qualitative)[19][20].
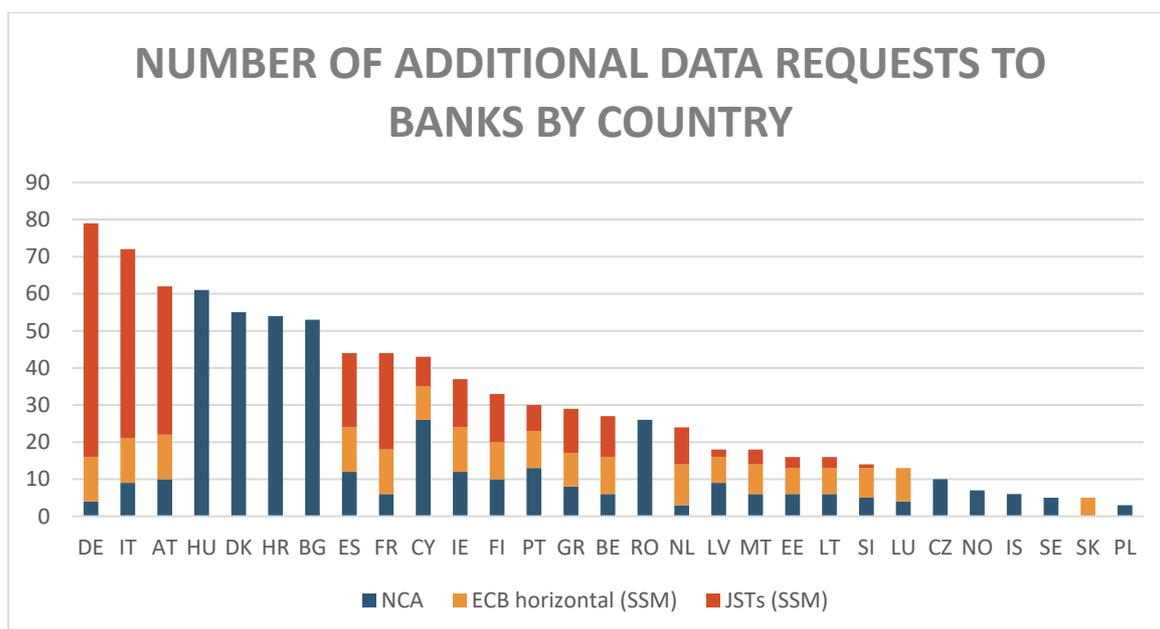
Source: Data inventory. Requests stemming from EU harmonised reporting framework or other authorities, such as the EBA, are excluded.

---

[18] There may be overlap, as some requests refer both to quantitative and qualitative information. In such cases, the request is counted both as quantitative and qualitative.

[19] The figure refers simply to the number of all ongoing additional requests in a jurisdiction, without differentiating between requests with a different scope. Some additional data requests thus refer only to a single entity, while others to a group of entities of all relevant institutions in the jurisdiction.

[20] It should be noted that the SSM's Less Significant Institutions specific requests are excluded from the overview. The numbers thus represent all ongoing additional data requests to Significant Institutions in a jurisdiction, raised either by the SSM or the relevant NCA.

2.2.2 Frequency of additional requests

48 As reported by the competent authorities, most requests, aligned with the harmonised prudential reporting, are on a quarterly basis, mostly for credit and counterparty credit risk, while a monthly frequency of reporting is also common. Most of the monthly reporting requests are for purposes of credit and counterparty credit risk, as well as liquidity risk. Liquidity risk monitoring is also the main reason for weekly and daily data requests.

Figure 3: Number of additional data requests by frequency (qualitative and quantitative).
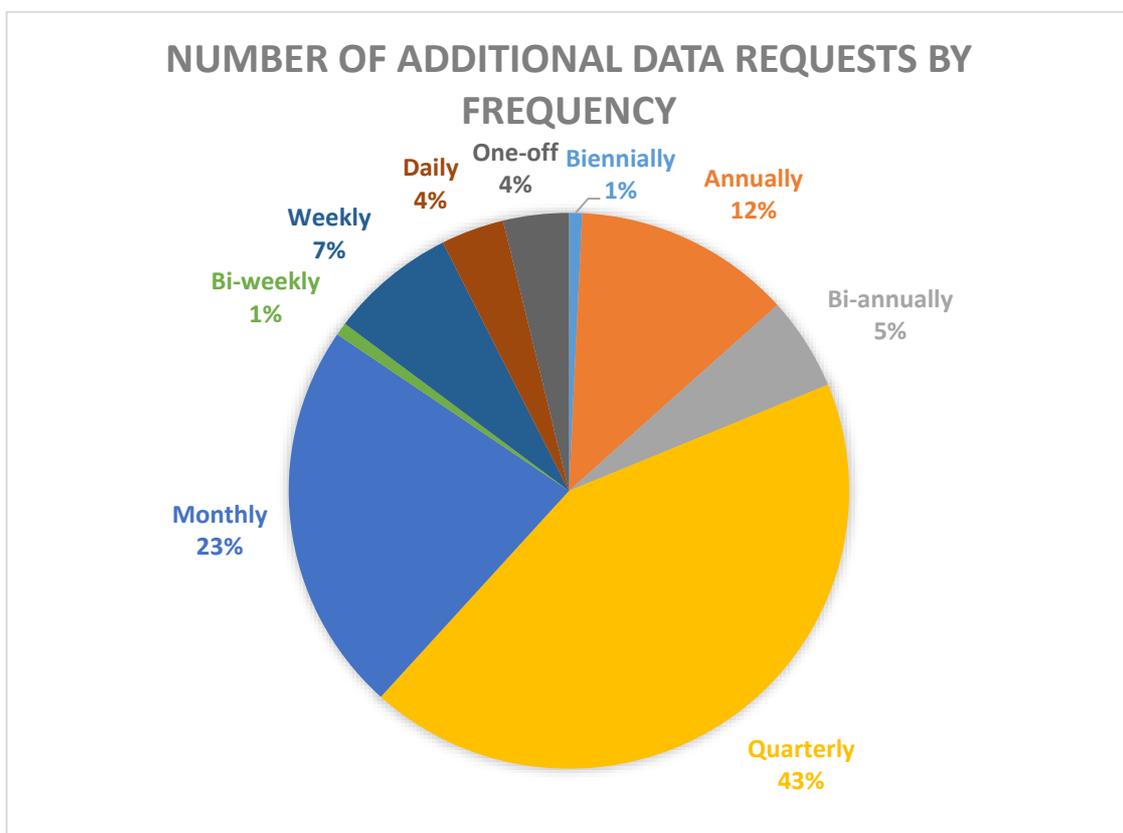
Source: Data inventory. Requests stemming from EU harmonised reporting framework or other authorities, such as the EBA, are excluded

.

It is understood that, for specific purposes or entities, monitoring the situation with a higher frequency than what is specified in the prudential reporting might be necessary, e.g. monitoring the liquidity situation of a troubled institution. However, such requests should be restricted to the troubled entities, and should also have a time limit. Thus, they are unlikely to make up a sizeable portion of the additional data requests.

2.2.3 Additional requests by risk area

49 As can be seen from **Error! Reference source not found.**, the bulk of the additional requests concern credit and counterparty credit risk, followed by liquidity monitoring, which together account for 55% of all data requests[21].

Figure 1: Number of additional data requests per main risk areas (qualitative and quantitative).[2223]
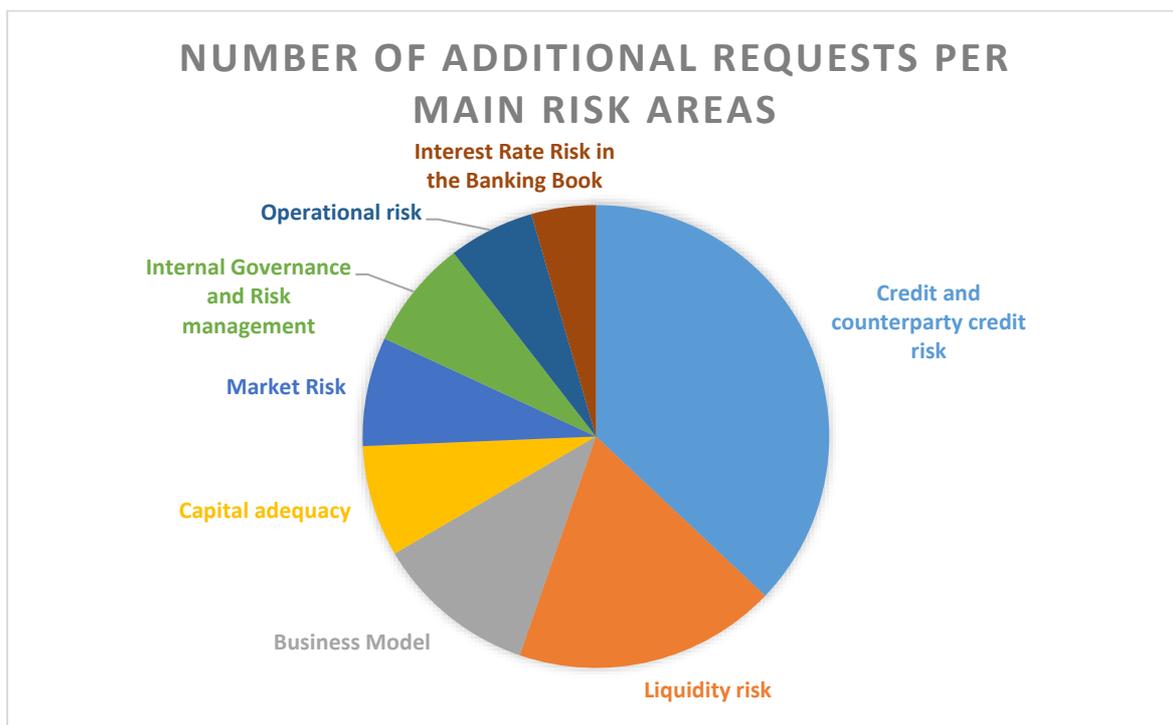
NUMBER OF ADDITIONAL REQUESTS PER MAIN RISK AREAS

Source: Data inventory, requests stemming from EU harmonised reporting framework or other authorities, such as the EBA, are excluded.

50 This picture seems to complement the harmonised EBA reporting framework, where credit risk monitoring comprises over a quarter of all existing data points[24]. The other two major risk areas in the harmonised prudential reporting relate to additional liquidity monitoring and financial information (FINREP framework), each accounting for another 16% of the defined data points.

---

[21] Only the main risk areas are taken into account.

[22] SSM requests concern only Significant Institutions, while requests made by SSM NCAs concern only Less Significant Institutions.

[23] The figure refers simply to the number of additional requests in a jurisdiction, without differentiating between requests with different scope. Some additional data requests refer only to a single entity, while others to a group of entities of all relevant institutions in the jurisdiction.

[24] Analysis is done on the defined DPM data points, meaning that data reported in open tables are counted only once.

2.3   Resolution data

51 Resolution reporting has only recently been included in harmonised frameworks across the EU with the EBA and SRB frameworks[25]. Just under 1,000 defined data points exist in the EBA framework, of which banks on average[26] report around 450 per reference period, while the SRB framework encompasses around 2,500 defined data points, some of which belong to open tables. However, and while more work is envisaged to align the reporting, the concepts used differ for some reporting items due to an underlying regulation. As the minimum harmonisation principle applies, the frameworks also differ in scope, and there may be further data collected by the national resolution authorities.

2.4   Statistical data

52 Institutions, in their capacity as deposit-takers, also provide data for various statistical purposes. To have a broad overview of these data requests, the EBA is leveraging on information made available by the ESCB on the harmonised statistical reporting for ESCB purposes. This amounts, on average, to over 8,000 data points or attributes[27] per institution being collected for statistical purposes to fulfil the mandates of the ESCB only. This reporting is harmonised across EU, however there exist additional requests on a national level, which likely broaden the scope of data collected and reduce the level of harmonisation.

53 The bulk of reporting focuses on balance sheet items, and payments and settlement systems statistics. Almost the entire set of data is provided with at least a monthly frequency[28.] Regarding the level of granularity, data collections comprising two thirds of the data points have both aggregated and granular aspects, while the rest are aggregated and only a very small percentage is reported only on a granular basis. However, key balance sheet items are reported on a granular basis, e.g. loan by loan. In terms of the number of institutions involved, roughly 86% of the data points are reported by more than 3,000 institutions[29], with mixed and aggregated granularities, and monthly frequency as a minimum.

54 While there are some links between the prudential and statistical frameworks, they vary due to significantly different purposes. However, in some cases, such as for statistical reporting and FINREP, many concepts and definitions are shared, although their scope is different[30.]

---

[25] The two frameworks differ in the requested data, however follow the same structure and format, aligned with supervisory reporting.

[26] Average for the sample of banks currently reporting to the EBA (Largest banks).

[27] For data collected on a granular basis, the number of attributes is provided instead of the number of data points.

[28] Reporting frequencies are indicated at data collection level, ranging from daily to annual. For simplification purposes, the EBA considered the highest minimum frequency as being applicable to all the data points in a data collection due to lack of more granular information.

[29] For perspective, around 4600 credit institutions are currently registered in the EEA. Statistical data is reported mostly by credit institutions, however it includes other financial institutions with monetary liabilities.

[30] For more details on the differences between the two frameworks, please see https://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/reporting/html/bridgingeba.en.html.

2.5 Evidence from the Cost of Compliance study

55 The cost of compliance study[31] was designed to measure historical reporting costs in relation to the EBA Implementing Technical Standards (ITS) on Prudential Reporting. Nonetheless, the industry responses to the questionnaire used for the purposes of the EBA study of the cost of compliance with supervisory reporting requirements provide interesting insights on the perceived costs of ad hoc and national requirements, and on costs originating from dealing with data requests from different sources (for example national and EU requirements).

56 The existence of data requests from different bodies was indicated as having a high or medium-high impact on reporting costs for the majority (78%) of the respondents, with an even higher percentage for large banks[32] (84%). Similar evidence is found when asking to evaluate the impact of complexity, clarity (or lack of clarity) of ad hoc reporting requests from prudential or resolution authorities. 68% of participating institutions identify overlaps between (EBA/standardised, regular) reporting requirements and reporting requirements of non-standardised/non-regular nature (ad hoc requests) as heavy contributors to the cost of reporting, providing evidence for the benefits of further integration.

57 When examining the impact of increasing granularity, one could expect a reduction in the number of ad hoc data requests. While acknowledging that the evolution of risks leads to new definitions and hence new data requests, it is also reasonable to assume that in a high granularity setting such data could be already available. Only one third of the respondents to the cost of compliance questionnaire support this vision, and agree with the statement 'Regularly requested reporting data in a more granular manner reduces the number of ad hoc requirements'. The percentage of large banks that agree is higher (46%).

58 There is no clear indication on how burdensome the interaction with the data recipient after submission and resubmissions is for banks, nor regarding the necessity to interact with multiple data recipients for one and the same or different reports. Both questions have split views with nearly half of the sample indicating the contribution as either high or low, across all size classes. This is somewhat in contrast with the widespread perception that interacting with competent authorities is demanding for reporting agents. However, 60% of the sample claims that the coexistence of different technical formats for different reporting requirements has a substantial impact on costs, from which we infer that harmonising the format for national/ad hoc requests and EU ones would be beneficial for those reporting agents.

2.6 Preliminary conclusions

59 The stocktake provided an overview of the current situation with respect to the prudential, resolution and statistical data requests. The data shows that on top of harmonised prudential data, reported by all institutions, the additional prudential data requests vary across different

---

[31] The cost of compliance questionnaire does not represent Lithuania, Slovenia, Sweden or Iceland due to non-participation in the exercise.

[32] Size classification is based on banks' own self-assessment.

jurisdictions. The same can be said for statistical data, where there are national extensions to harmonised reporting. Additionally, the minimum harmonisation principle applied to resolution reporting results in potential national discretions, leading to differences in reporting among jurisdictions.

60 It is understood that some of the non-harmonised prudential data requests cannot be fully integrated into the harmonised EBA reporting framework, as they are either specific to certain banks or certain situations. Some others are already envisaged to be included in the EBA reporting framework. Based on the observations it is considered that the following aspects need to be explored further in order to optimise the current system:

- Given the current underlying legal frameworks, explore having a common data model and dictionary as a single point of reference.

- Integration of data requests within the existing definitions and frameworks[33] where possible.

- Increased data sharing among competent authorities.

- The possibility of deriving the requested data from already existing more granular information.

- The possibility of requesting certain sets of reported data with higher frequency if necessary.

- Banks not being requested to report the same data multiple times.

61 While, in order to fulfil their tasks, prudential authorities should have ad hoc data requests at their disposal, this stocktake suggests that additional data requests should be considered as possible sources of request duplications and redundancies. This observation leads to the notion that alignment of ad hoc data requests with regular requests could be a source of efficiency gains.

Questions to consider

5) Do you confirm the findings presented in the stocktake? If you have additional information, please provide more specific details about the amount of data collected.

---

[33] In the process of creating an integrated reporting framework, the existing definitions and frameworks would have to be revised and consolidated as necessary.

3. The reporting process: criteria used for the integration assessment

3.1 Introduction to the reporting process chain

62 The mandate requests the EBA to investigate 'the feasibility regarding the development of a consistent and integrated system for collecting statistical data, resolution data and prudential data', which could include a common data dictionary and a central data collection point.

63 The EBA is taking a holistic approach and considering the impact of integration at each step in the reporting process chain (data definition, data collection, data transformation and data exploration) along the different levels of abstraction (from the high-level business concepts and logic - semantic level, through the formal and standardised formats -syntactic level, to the more tangible components of the technological architecture -infrastructure level), which enables the use of data by final users).

64 The reporting process is a sequence of concatenated processes which start with the definition of the data needs and end at the point where the data is made available for consumption (by a user or a different process). Figure 5 depicts the regulatory lifecycle where the data dictionary is at the core of every step of the process.

65 The data dictionary is a central piece all along the reporting process chain serving the needs of both producers and users of data. The assessment of feasibility of integration at different stages in the reporting process provides the basis for the analysis of feasibility of creating a common data dictionary (Section 4). In addition, the way in which the data collection is organised (options for design of the central data collection point are described in Section 5) is an essential piece in the reporting process chain, not only by the role of facilitating the data collection but also regarding its possible role in the other data processes (e.g. the central data collection point might also represent the point where data is transformed).

Figure 5 – Regulatory data lifecycle
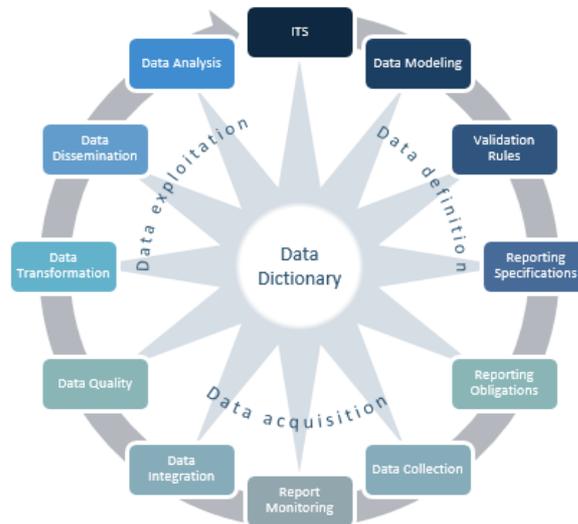
66 The regulatory data lifecycle can be summarised in the following reporting process chain (Figure 6):

- **data definition**: process in which the data requirements are defined;

- **data collection**: process used for exchanging data;

- **data transformation**: process in which the received data is transformed, for instance to create new data for analysis/disclosure or to calculate aggregates from more granular data;

- **data exploration**: process that allows to use, share and cross with other data, by final users or systems.

Figure 6 - Reporting process chain
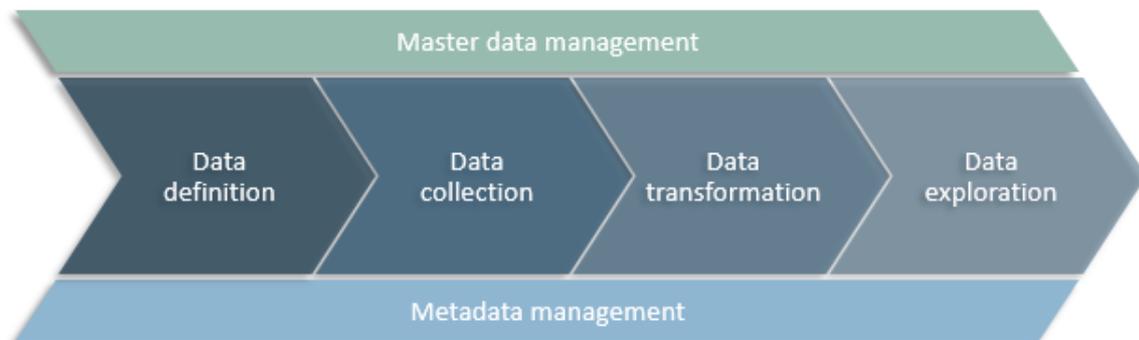
3.2   Criteria used for the integration assessment

67 The different parts of the reporting process chain can be represented at different levels of abstraction, from the high-level business concepts and logic (semantic level), to the formal and standardised formats (syntactic level), to the more tangible components of the technological architecture (infrastructure level) as shown in Figure 3. In the following, the costs and benefits of integration at different levels in the reporting process chain is provided and possible interdependencies between the different levels are identified. While direct interdependencies might represent a constraint to be accounted for in the integration decision, additional indirect interdependencies are important for the cost-effectiveness of the integration process. Costs and benefits might therefore be different should different components be integrated together vs independently. In addition, in order to prevent the 'waterbed-effect' (when a benefit at one point could determine further costs at another point), an overall cost and benefit analysis should be performed at a later stage of this process.

Figure 7 - Process Chain Levels of Abstractions

Level	Context	Focus
Semantic	Business Management	Meaning of business concepts and business rules
Syntactic	Information Management	Rules and principles used to exchange and process data (exchange formats, protocols and interfaces)
Infrastructure	System Management	Information systems and technologies needed to support the full data lifecycle

68 Figure 8 below provides a representation of the reporting chain considering the different levels of abstraction. A high-level analysis of the integration at each process step and abstraction level is carried out for the purpose of the feasibility study:

Figure 8 - Process chain and the three levels of abstraction

	Data definition	Data collection	Data transformation	Data exploration
Semantic level	*e.g. Glossary, data objects, validation rules*	*e.g. Reporting rules*	*e.g. Transformation rules*	*e.g. Analytical models*
Syntactic level	*e.g. DPM structure (meta-model)*	*e.g. Data exchange format*	*e.g. Validation and transformation language*	*e.g. Dimensional schemas*
Infrastructure level	*e.g. DPM database and tools*	*e.g. Data collection portal*	*e.g. Calculation engine*	*e.g. Data marts and BI tools*

3.2.1   Data definition: Integration at the Semantic level

	Data definition	Data collection	Data transformation	Data exploration
Semantic level				
Syntactic level				
Infrastructure level				

69 **Relevance:** Integrating the semantic level of data definition is required to ensure a unique understanding of the meaning of different kinds of concepts and the relationship between them.

70 **Objective**: The semantic integration of the data definition process aims to integrate the glossaries of all reporting frameworks, so that business concepts with the same meaning are uniquely named and defined. It should harmonise or clearly distinguish between the definitions of reporting requirements by different regulators and ease the understanding for all stakeholders. Integration at this level stands for harmonisation at the level of meaning by making sure everyone has the same understanding on what is being asked to be reported.

71 **Status quo**: In terms of first-level reporting (from institutions to national authorities), across jurisdictions, national authorities have different approaches in defining the semantic level of reporting requirements (the semantic data definition). Some of them have adopted the European definitions, others have further extended or mapped them to a national framework. Very few authorities already have in place at national level integrated data dictionaries covering

(partially) prudential, statistical and resolution data. In terms of second-level reporting (from national to European authorities), the EBA, the ECB and national frameworks remain isolated from each other due to the different purposes they are serving. The EBA has been publishing an integrated reporting framework for all CRR areas since 2012 and since then has also integrated the Bank recovery and resolution directive (BRRD) in the same data dictionary (DPM). Currently the ECB is working on the internal integration of its frameworks within the IReF project.

72 **Integration implementation:** Achieving semantic integration of the data definition process requires the highest level of coordination and cooperation among all authorities responsible for setting up reporting requirements. In practice, the end result would be a single data dictionary of concepts covering clear definitions of the prudential, statistical and resolution data requirements. A proper governance structure would be needed to coordinate the work in such a way that time and resource costs are kept under control.

73 **Expected** costs and benefits of integration:

	Costs	Benefits
Competent authorities	• High implementation costs at the outset, as it would require complete harmonisation of definitions and business concepts throughout the common semantic EU dictionary (**resources, time, financial**). • Identifying the relationships between concepts, as semantic mappings in the common EU dictionary have high costs (depending on its complexity and stakeholders involved in its analysis) (**time, financial, resources**). • Continuing analysis of all existing reporting frameworks would be needed going forward, which may undo some of the reduction in cost described to the right (**time, financial, resources**).	• Business concepts with the same meaning will be uniquely named and defined (**consistency**). • Facilitate the understanding and interconnection of reporting requirements defined by different regulators (**clarit**y). • Easier compilation of tailored datasets across different requirements for monitoring cross-sector/risk issues (**usability**). • Once established, the maintenance cost of a common dictionary would be lower as new concepts for example would just be appended after a first harmonisation stage (**simplification, efficiency, fewer resources, and financial cost reduction**).

> ➢ Benefits and costs may differ (on the competent authorities and institutions side) depending on the granularity of the common data dictionary, which is determined by the granularity of the collection layer.

Institutions	• Institutions might have already developed such integrated definitions in their systems (accounting for their specificity) and mapped them to the current reporting requirements (sunk costs) (**financial, time, resources**).	• Increased data quality due to the consistency in the definitions provided by European authorities (**data quality improvement**). • Facilitate data lineage and data stewardship internally (**clarity, efficiency, fewer resources**). • Avoids data duplication and data overlaps due to increase in consistency and harmonised definitions between different data frameworks **(consistency).**

74 **Feasibility:** Semantic integration is considered feasible, highly desirable and one of the main steps in order to achieve integration. Building a complete semantic data dictionary requires a close coordination and cooperation between different authorities in order to analyse the business concepts used across all reporting frameworks and using uniquely defined terms for the same concepts, or differentiating accordingly where concepts are different. Costs can be assessed in terms of time and coordination needed for harmonising existing frameworks, national extensions and integration of new data needs. The maintenance and future amendments to the dictionary also have to be part of a coordinated process. In addition, having different semantic dictionaries expressed using different standards (e.g. different metamodels) and different infrastructures will significantly add to the burden of analysing the different reporting requirements.

75 Based on the fact-finding workshops and seminars, the analysis concluded that some of the **aggregated concepts** compiled across national authorities might share similarities, but only a few are identical. Thus, harmonising these existing concepts, under the existing regulation, may be difficult to achieve due to marginal differences (often stemming from the underlying regulation) that lead to conceptually different definitions. Creating a dictionary of these concepts may require listing all the differences explicitly in the dictionary (and mapping concepts whenever the definitions align). While making the differences and commonalities explicit is desirable, and represents a first step towards integration, further improvements could be brought to the semantic integration by amendments to the regulation or going more granular in the concepts collected (more details in Section 4.6), in order to further **simplify** and **streamline the data definitions.** However, it is of utmost importance that prudential and resolution reporting requirements stay fully aligned with the underlying regulations and accounting standards. Any discrepancies would create additional compliance costs.

76 Dependencies **and constraints:** The integration of the data dictionary at semantic level (semantic data definition) could be done completely independently from the other processes and levels, as it aims to create a complete/common glossary of business concepts. However, syntactic and infrastructure integration might further facilitate the process of semantic integration. Having all the business concepts across different reporting frameworks depicted

3.2.2 Data definition: Integration at Syntactic level
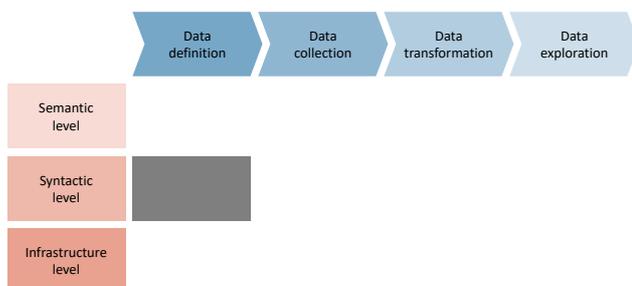
77 **Relevance:** The syntactic level of the data dictionary is referring to the standard and formalised model used to depict the business concepts. Choosing a unique standard to depict the reporting requirements (e.g. the business concepts defined at semantic level) for the data dictionary would achieve the syntactic integration, as opposed to using different standards. Such standard models are referred to by the name of metadata model (or metamodel which is a model to depict the metadata). The importance of such a model is crucial for the reporting process as it ensures a unique, unambiguous interpretation of the data requirements facilitating the automation and digital processing of regulatory data.

78 **Objective:** Integration of the data definition at the syntactic level is limited to defining a standard structure (metamodel) for all reporting dictionaries, so that reporting requirements are defined in the same way across all frameworks. The integration of the syntactic level eases preparation of regulatory reports required by different regulators. It enables the design of common reporting solutions for different reporting frameworks.

79 **Status quo:** In terms of first-level reporting, for prudential and resolution reporting, some authorities have adopted the EBA model (DPM) for the European harmonised frameworks while few others have integrated the requirements into their national data models. Regarding additional national requirements, few national authorities have extended the DPM model while others are using the national one. Regarding integrated models, very few authorities depict their statistical, resolution and prudential data requirements (at least in parts) using the same metamodel (e.g. matrix model, OeNB metamodel). In terms of second-level reporting the prudential and resolution European harmonised frameworks are depicted using one single metamodel (e.g. EBA DPM). The ECB is currently using different metamodels for the different frameworks of the statistical requirements but is planning to switch to a unique metamodel, part of the IRef project (that would be directly applicable to deposit-taking institutions). The Single Data Dictionary (SDD), the ECB internal data dictionary, describing different datasets used at the ECB is using a single metamodel, the SMCube Methodology[34].

---

[34] https://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/smcube/html/index.en.html

80 **Integration implementation**: Achieving syntactic integration of the data definition process requires the coordination of different authorities in charge with defining reporting requirements to agree on a standard metadata model to be used.

81 **Expected** costs and benefits of integration:

	Costs	Benefits
Competent authorities	• Potential high one-off costs for national authorities during the first onboarding phase in the form of time and resources. Nevertheless, these costs will be amortised over time as supporting multiple metamodels will no longer be needed (**financial, time, resources**).	• Standard structure (metamodel) for all reporting dictionaries, so that reporting requirements are defined the same way across all frameworks (**simplification, clarity**). • Possibility of an easier combination of different datasets on the user side, at least from the technical point of view (**interoperability**). • Use of the same standards for all reports improves the usability and reduces processing costs (**usability, reusability, efficiency / financial cost reduction, fewer resources**).
Institutions	• Moderate costs for institutions during the initial onboarding phase which could be amortised over time as supporting multiple metamodels will no longer be needed (**financial, time, resources**).	• Using a unique data structure to define reporting requirements would enhance the usability of data. Processing datasets coming from different metamodels is time consuming for institutions' internal engines. In the long-run, hence, a standard data structure for defining reporting requirements will imply cost reductions since institutions' internal systems would process (map) to a unique metamodel structure of the reporting requirements.(**financial cost reduction, efficiency, fewer resources, usability**) • The reporting efficiency would increase since institutions would no longer need to create different queries or structure the data in different ways to produce the reports (**efficiency, simplification, streamlining, and usability**). Hence, timeliness of institutions' reporting could be improved (**timeliness[35]**).

---

[35] Included in the table in the annex as 'Quality improvements'

82 **Feasibility:** There are multiple examples of effective and widely used syntactic standards that could be used to define reporting requirements, thus achieving integration at this level is thought feasible.

83 **Dependencies and constraints**: The integration of the data definition at the syntactic level could be achieved without integration at the semantic level, as no alignment of concepts from the business point of view is needed. Integration at this level can be achieved as one or multiple semantic data dictionaries (representing the metadata content / business content) could be depicted using the same structured format to store the metadata content. The data model for Solvency 2 reporting is an example of syntactic integration with CRR/CRD reporting since, even though semantically distinct, the frameworks published by the EBA and the European Insurance and Occupational Pensions Authority (EIOPA) are supported by the same metamodel, with ongoing efforts towards a full integration. The Single Data Dictionary (SDD), the ECB internal data dictionary, is also describing different datasets used at the ECB using a single model but capable of keeping the semantics of the different frameworks unintegrated.

84 While the integration of the syntactic level is not dependent on integration in the other blocks, integration at the syntactic level is crucial for the cost-effectiveness integration of the rest of the reporting process levels. In addition, syntactic level integration might facilitate semantic level integration.

3.2.3 Data definition: Integration at Infrastructure level
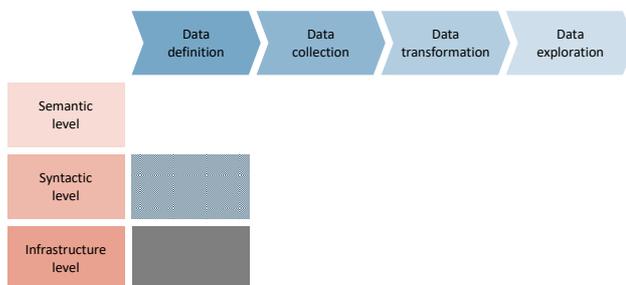
85 **Relevance**: Infrastructure integration of the dictionary used to manage the data definition process (including the semantic and syntactic representation process) is an essential instrument that would facilitate the collaboration between different stakeholders that is much needed when creating a unique data dictionary.

86 **Objective:** The aim of integration of the data definition at the infrastructure level is to build a common technical solution to support the development, management and evolution of all reporting frameworks (support the data definition process). In addition, the infrastructure should facilitate the development of interoperable reporting systems (the infrastructure of the data definition process could be accessed by other applications).

87 **Status quo:** For first-level reporting, national authorities have developed (some are in the process of doing so) in-house applications to facilitate the metadata management. Depending

on the level of integration achieved, the same infrastructure may be used to manage different reporting frameworks. While syntactic integration correlates to having infrastructure integration, it is not a precondition. Authorities also use additional tools to translate different metadata models into other structures (e.g. exporting of the DPM to the Multidimensional data model) in their applications. Similarly, having a single metamodel does not restrict from having different infrastructure to manage the metadata. At second-level reporting the EBA and EIOPA share a similar version of the DPM database (syntactic level), but the development and disseminating tools are different.

88 **Integration implementation:** as the infrastructure is a tool to facilitate the data definition process, the optimal design will depend on how the decisions on other aspects have been taken. Given that the dictionary would be used and shared among different stakeholders, for efficiency reasons, one application solution or alternatively different interoperable applications might be preferable. The ultimate solution should account in addition for proper governance (e.g. access rights, data privacy) given the integrated nature of the underlying dictionary it is supposed to store and manage.

89 **Expected** costs and benefits of integration:

	Costs	Benefits
Competent authorities	• In addition to the initial costs of onboarding, the main efforts required will be agreeing on the technical architecture of the solution(s). Depending on the infrastructure model adopted, the relative costs of defining standard interoperable interfaces and operational maintenance would vary accordingly (**financial, time, resources**). Data definitions will evolve and need to be constantly updated. In addition, checking transformation of new data may increase maintenance costs.	• Integration at this level would lead to a common technical solution to support the development, management and evolution of all reporting frameworks (**simplification, clarity and streamlining**). • The solution would enable the development of interoperable reporting systems (**interoperability**). • Using similar technology stacks would enable better use of resources and increased efficiency with similar design patterns and tools (**reusability / fewer resources, efficiency**).

Institutions	• Depending on the type of architecture (centralised/decentralised), costs may vary for institutions. It will depend on whether contributions are requested from institutions for the maintenance of the infrastructure (**financial, time, resources**). • Data definitions will evolve and need to be constantly updated. In addition, checking transformation of new data may increase maintenance costs (**time, resources**).	• Reporting to one single data infrastructure would be beneficial in terms of efficiency of data submissions to the database. A single data infrastructure makes it possible to trace the datasets that feed the database. Institutions would hence not need to report the same datasets again if already provided and contained in the database infrastructure (**efficiency, reusability, fewer resources**).

90 **Feasibility**: Depending on the architecture chosen, the complexity and the costs of the solution may vary.

91 **Dependencies and constraints:** Given the current technology advancements, for the integration of the data definition at infrastructure level, the integration at the syntactic and semantic level are not necessarily a precondition. The solution can implement one or multiple data dictionaries (semantic and syntactic), describing different reporting requirements, with the same technological support. However, for efficiency reasons, syntactic integration might be preferable.

3.2.4 Data collection: Integration at Semantic level
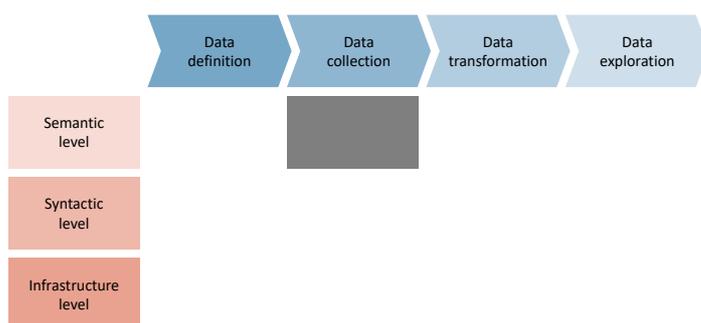
92 **Relevance**: To achieve integration, a standard set of reporting rules[36] is required. The reporting rules specify the reporting obligations for the institutions, (i.e. who must report, what and when: 'Module X should be reported by entities of type Y with frequency Z'). Once the type of collection

---

[36] Diverse rules from specification of the dataflows topologies to rules like filling indicators. This also covers the definitions of reporting obligations of who has to report what and when.

system has been agreed, the best way to define the reporting obligations should be investigated, also depending on the type of data collected.

93 **Objective**: The reporting rules are relevant for the data collection at the semantic level. They indicate the content (including waivers), format, timeline, and frequency. Reporting obligations are set out in the regulation and some depend on the characteristics of the institution (e.g. internal ratings based models, IRB templates only need to be reported by those institutions with such models in use). The purpose of integration at the semantic level is the centralisation of the reporting obligations, facilitating institutions' understanding of all the reporting requirements from every regulator.

94 **Status quo:** At European level, for the EBA harmonised reporting frameworks such rules have been harmonised. Such reporting rules have been harmonised also at national level where an integrated system has been set in place (e.g. in the case of France, the statistical and supervisory semantic requirements have been fully integrated into rules defined at national level) and additionally, in some cases, they have been integrated to some extent at national level even though an integration system might not have been set in place for the data definition.

95 **Integration implementation:** integration of reporting rules will depend on the level of granularity and the type of data of the reporting requirements.

96 **Expected costs and benefits of integration:**

	Costs	Benefits
Competent authorities	• To integrate the data collection and semantics, a standard process will need to be set up. The cost may vary depending on the governance structure defined but should not imply many changes (**financial, time, resources**).	• Integration of data collection at the semantic level would lead to the use of a common set of rules for the development of different reports and could imply a good starting point for defining a coherent and comprehensive set of data in the centralised data collection point (**simplification, consistency and streamlining**). • Depending on the type of solution used, due to the harmonisation of reporting, different authorities could also reuse the data collected for a specific purpose for other different analyses or prudential purposes (**reusability**).

Institutions	• No impact.	• Integrating the reporting obligations will facilitate the institutions' understanding of all the reporting requirements from each competent authority (**consistency, clarity, usability, simplification**).

97 **Feasibility:** Reporting requirements are aligned to the different types of data. For instance, frequency will differ depending on the type of data and its use. An agreement will be necessary within authorities to align the reporting requirements considering the singularities of each type of data to be integrated.

98 **Dependencies and constraints:** The integration of the reporting requirements could be done independently of all the other blocks**.**

3.2.5   Data collection: Integration at Syntactic level
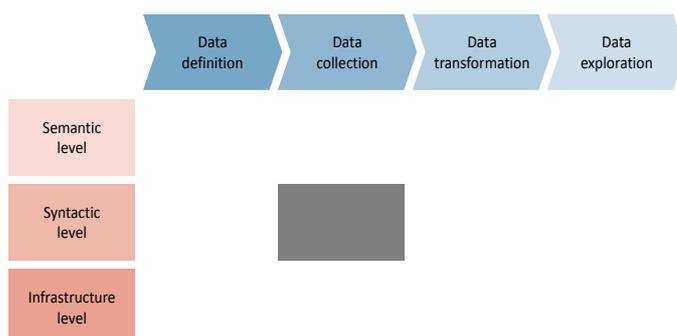
99 **Relevance:** The data collection could allow different data formats. However, with a central data collection point in place, it would make sense to have a common collection format as otherwise, the aim of simplification, streamlining and reusability would not be fulfilled. The use of the same data exchange format may have different impacts on the authorities and institutions, depending on how the central data collection point is defined.

100   **Objective**: Integration at this level requires defining or agreeing on a common data exchange format for all reporting. This will ease the reporting burden of institutions that would use only one format for all reporting to the authorities. An example of syntactic-level integration of the data collection process would be the use of the same **data exchange format**.

101   **Status quo**: For first-level reporting for the EBA ITS, many authorities collect the data using eXtensible Business Reporting Language (XBRL). For the rest of the data, there may be different data formats across different competent authorities (CAs) and also variations in formats for different data collections within CAs. In the case of second-level reporting of the EBA ITS, the XBRL is needed; the ECB requires Statistical Data and Metadata eXchange (SDMX) but may require other formats in addition; the SRB uses XBRL but may require also other formats.

Expected costs and benefits of integration:

	Costs	Benefits
Competent authorities	• It could lead to initial increased one-off costs to convert existing reporting schemes into the new common data exchange format (**time, resources, financial**). The costs would be partially off-set by reducing the maintenance cost as described above.	• It would provide authorities with the same data formats across all datasets, making it easier for them to read, check and (re)use the data (**simplification, streamlining, clarity, usability, reusability, interoperability**). • Different datasets reported in the same format further allows for much better exchange of the data in the system (**interoperability**). • Reduced maintenance cost as there would no longer be different reporting formats with different technologies and knowledge required (**fewer resources, efficiency, financial cost reduction**).
Institutions	• A significant change to the reporting format could lead to initial increased costs for banks since their internal reporting systems would require adjustment (**financial, time, resources**). In particular, the impact on smaller institutions will need to be reflected on.	• Reporting all data requirements in the same data format should hugely improve institutions' reporting process and efficiency (**fewer resources, efficiency, clarity**). • Using a standard format could lead to direct cost reductions since standardisation could make reporting a market commodity (**financial cost reductions, fewer resources**).

102   **Feasibility:** It should be possible, at least for all European reporting frameworks, to agree on a single data exchange format, but it would be desirable to standardise the use also for national reporting and the ad hoc reporting. In addition, proportionality aspects should be accounted for when deciding on the exchange format.

103   **Dependencies and constraints:** The use of a common data exchange is independent from the integration of the reporting rules (semantic integration of the data collection). However, integration in the syntactic data definition process is important, as the data exchange will consider the metadata structure.

3.2.6 Data collection: Integration at Infrastructure level
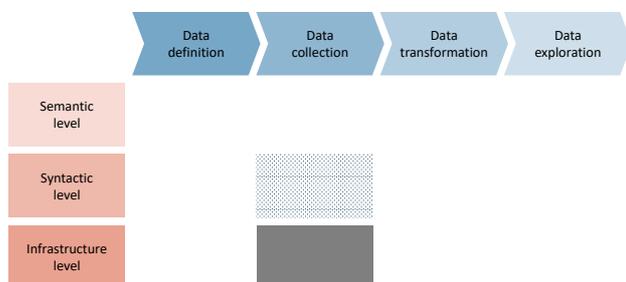
104 **Relevance:** the data collection integration at infrastructure level is the last step of the integration process. Having a unique data dictionary and a single reporting scheme are preconditions for the infrastructure integration. In addition, the system integration could improve the efficiency of existing processes and business functions that rely on different information systems.

105 **Objective**: The infrastructure level of the data collection process is the data collection system. Different system architectures (**blueprints**) that would allow integration could be envisaged (see Section 5 of this report).

106 **Status quo:** in the current reporting landscape many systems coexist for the collection of data at the different levels of reporting.

107 **Dependencies and constraints:** Although an integrated infrastructure does not necessarily need the integration at the data exchange format (i.e. the syntactic level), as it could allow the collection of all types of data formats, it would make sense to use the same data exchange format. The integrated infrastructure should be independent of the need for any other integration at the reporting chain.

108 The cost, benefits, feasibility and implementation will depend on the architecture selected and how it will be operated.

3.2.7 Data transformation: Integration at semantic level
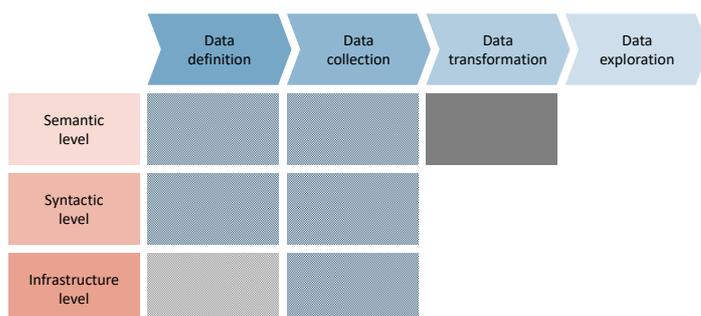
109   **Relevance:** The data transformation rules can be prepared to serve different requirements, both of regulators and of institutions. Semantic transformation rules refer to defining the business logic embedded in the transformation of the data. On the regulators' side, the sharing of semantic transformations can improve the data and its quality by i) enabling clear data validations; ii) obtaining new derived data; and iii) facilitating integration of different regulatory frameworks into a common level of reporting granularity and calculating the necessary aggregates.

110   **Objective:** Integrated semantics of the data transformation will support cooperation between different stakeholders in defining and sharing the transformations that should be applied to the data and can affect the processes from data collection trough data validation, data analysis and data disclosure

111   **Status quo**: Currently data is collected at the level of aggregation needed, however there is limited visibility on how the data is transformed by the institutions to arrive at the aggregated figures and how the data is further used by the authorities. Uncertainty of what rules should be used and how to best support them is one of the key challenges identified during the fact-finding workshops and seminars. Some national authorities have developed integrated systems where data is collected at a more granular level and transformed to obtain the output figures collected by the authorities. Other projects, such as BIRD[37], are looking to define common transformation rules for the data, starting from a common input layer, to obtain the collection layer requested by the authorities.

112   **Integration implementation:** From a semantic point of view, the goal is to define transformation rules with the same business meaning for all reporting institutions and data-collection authorities to allow for calculating final data (e.g. ratios or aggregates) from common, more granular data or to define validation rules. At the semantic level, such transformation rules can represent a description of steps needed to achieve a particular reporting. The transformation rules could be seen as part of the common data dictionary alongside the data definition.

113   **Expected costs and benefits of integration:**

Costs	Benefits

Competent authorities	• Depending on the scope of transformation rules to be defined and how granular data is collected, implementation costs could be high initially, but the maintenance costs would ultimately be lower (**financial, time, resources**).	• If data is collected at a more granular level and can be transformed, data already collected could be reused for future data requirements, transformed with new transformation rules, thereby reducing future data requests (**reusability, efficiency**). • Common rules are expected to increase data quality (**quality improvements**).
Institutions	• If the transformation rules are used to collect more granular data, the impact on institutions may differ depending on their current reporting process. For some institutions, reporting granular data may be more costly, for others it may simplify their processes and therefore, it may be less costly (**resources, financial, time**).	• Defining harmonised transformation rules would facilitate Institutions' effort in interpreting the regulatory requirements, at least what aggregated data should represent (**simplification, streamlining, usability, efficiency**). • Data could be collected at a more granular level, which would limit duplication of efforts and enforce a 'report once' policy (e.g. data for creating some items of the balance sheet statistics and FINREP) (**efficiency, fewer resources**). • Centralisation of the transformation rules and integration across all reporting requirements will limit the institution's need to construct or obtain the transformation framework from external providers (**financial cost reduction, efficiency, fewer resources**).

114 **Feasibility:** Achieving a higher degree of integration will require agreement on definition of the data concepts and on transformation rules used to create new data, derive aggregates or define validation rules. Integration of data transformation at the semantic level will depend on the degree of integration achieved in terms of data definition (what needs to be transformed). Moreover, defining harmonised and standardised integrated transformations will highly depend on the complexity of the operations required to achieve the needed output data.

115 **Dependencies and constraints:** Integration at the data definition level is needed to be able to understand unequivocally the data and how to transform it. In addition, any integration at the syntactic level will depend on the level of integration achieved at the semantic level.

3.2.8 Data transformation: Integration at syntactic level
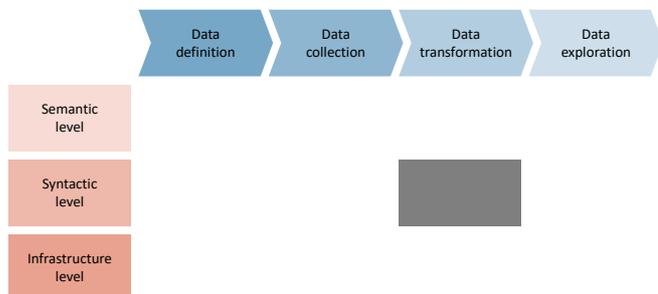
116   **Relevance:** The syntactic level of data transformations refers to the way in which the business logic of the transformations (the semantic level) is formally represented. Having all the transformations defined in the same format will ensure a common and unambiguous understanding to all stakeholders of the data requirements, avoid overlaps and facilitate harmonisation of standards.

117   **Objective:** From a syntactic perspective, the goal is to define a common language for the representation of the transformations and a standard structure (metamodel) to store and share the transformation rules.

118   **Status quo:** For the European harmonised reporting frameworks (prudential and resolution) the validation rules have been defined using a single structure. While at national level, authorities have developed national rules or extended the European ones, there is currently limited harmonisation in the way and the means by which this has been done across different frameworks or reports. For other type of transformations, there is currently limited visibility on how they are performed. National authorities that have developed integrated reporting systems have defined transformations and validation rules in a harmonised way. The extent to which the syntactic level integration has been formalised in a common language also varies: the ECB BIRD initiative has chosen a standard, formal language (VTL), for the definition of its transformations while in the case of Italian PUMA[38] these transformations are less formalised.

119   **Integration implementation:** From a syntactic point of view, the goal is to define transformation rules using the same format, requiring a high level of coordination between authorities in deciding on the characteristics of such a format that should satisfy different users' needs.

120   Expected costs and benefits of integration:

Costs	Benefits

---

[38] Unified Business Matrix Procedure (PUMA) is a voluntary cooperation initiative of the banking and financial system, promoted and coordinated by the Statistical Survey and Processing Service of the Bank of Italy

Competent authorities/regulators	• Harmonisation of data transformation will have an impact on how the transformations are defined and executed and who bears the responsibility for them (**financial, time, resources**).	• By sharing the same dictionary components, definitions and transformations, it would be possible to share and execute the same standard validation rules on reported data, share the calculations of new data and share how to transform granular into more aggregated data (**usability, clarity, consistency, streamlining**) • Regulators are interested in having a strong validation machine and they will save money by using common definitions and standards which can also enable the sharing of executable infrastructure and the reduction of operating costs (**usability, efficiency, financial cost reduction**).
Institutions	• Harmonisation of data transformation will have an impact on how the transformations are defined and executed and who bears the responsibility for them (**financial, time, resources**).	• Defining transformation rules using the same syntactic representation will ensure a common understanding of the regulatory requirements and facilitate communication (**usability, clarity, consistency, streamlining**). • Facilitate the use of technology in regulatory reporting by the use of standards and common formats for representing information facilitating machine readability of the requirements (transformations) (**usability, efficiency**).

121 **Feasibility**: Given the multitude of languages and standards for writing code/pseudocode/rules (ranging from high level – closer to business users to low level – closer to machine language), implementing an integrated syntactic level for the transformations is seen as feasible. Choosing the best standard and the best way to implement it should consider the needs of different users.

122 **Dependencies and constraints:** the syntactic level integration will depend on the semantic level integration of the transformations. In addition, integration at the level of the data definition (especially the metadata model of the data) is essential, as those are the structures to which the transformations are pointing (the data on which the transformations will apply).

3.2.9   Data transformation: Integration at infrastructure level
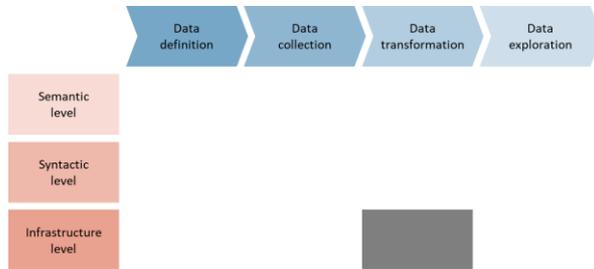
123   **Relevance:** Integration at infrastructure level refers to the application that would facilitate the creation, maintenance, testing, validation, sharing and running of the transformations. In addition, the infrastructure would also refer to the calculation engine that would apply the transformations to the data collected. An integrated infrastructure is essential to ensure a common process and platform where different users with different needs could coordinate their processes.

124   **Objective:** From an infrastructure point of view, the goal is to define an application that would support all the processes related to transformations, including possibly a calculation engine. The infrastructure related to defining transformations may be separated from the infrastructure related to running transformations. Such infrastructure (or at least part of the functionalities, e.g. visualisation of the metadata linkages) could be built individually or shared among various stakeholders.

125   **Status quo:** for some of the national authorities that have implemented integrated reporting systems, their infrastructure supporting validations and transformations has been only partially integrated (the IT implementation, including calculation engine, is left to be implemented by different software companies).

126   **Integration implementation:** from an infrastructural point of view, the goal is to agree on infrastructure to be implemented, requiring a high level of coordination between authorities (and possibly institutions) in deciding on the characteristics of such infrastructure that should satisfy different users' needs.

127   **Expected costs and benefits of integration:**

Costs	Benefits

Competent authorities/regulators	• IT and human resources cost needed in setting up the infrastructure and maintaining it (**time, financial, resources**). • Additional considerations for the governance process of the infrastructure, both in building and maintaining it (requiring significant coordination among different authorities and regulators, and between business users and developers) (**time, resources**).	• Significant benefits in terms of harmonisation and coordination of processes and information sharing(**efficiency, clarity, consistency, usability**).
Institutions	• Some of the costs on the authorities' side could be shared with institutions (depending on the governance of the infrastructure). • IT and human resources cost needed in setting up or buying similar infrastructure to serve the needs of the institution (**time, financial, resources**).	• Significant benefits in terms of harmonisation and clear communication of requirements from authorities to institutions (higher benefit for groups operating cross-border) (**clarity, consistency, efficiency, usability**).

128   **Feasibility**: this aspect is considered to be feasible. Depending on other decisions (e.g. the syntactic integration of the transformations) one could use (or adapt) solutions that might already be available on the market or may choose to build different infrastructure.

129   **Dependencies and constraints:** for efficiency reasons, an integrated infrastructure would benefit from integration in all aspects of data definition and in addition semantic and syntactic integration of the transformations. In addition, as the transformations are applied to the defined data it is important that the infrastructure used to manage the metadata (data definition and data model) allow interoperability with the infrastructure of the transformations.

3.2.10 Data exploration
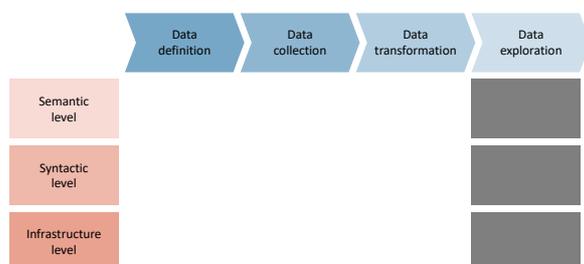
130   **Relevance:** One of the purposes of the integrated reporting is to be able to use existing data from different frameworks for purposes other than those originally requested; therefore, being able to explore the data and combine it is of relevance for the exploitation of an integrated reporting platform. However, as data exploration is the last function of the regulatory data

lifecycle, there will be de facto no degrees of freedom in the choice of integration level (semantics/syntax/infrastructure). The earlier integration decisions (i.e. on data definition and data acquisition) will constrain the integration level for the data exploration. Integration at the exploration level will have the benefit of avoiding duplications, the same information serving the needs of different users. In addition, data exploration may also need to be tailored to the specific needs of the statistical, prudential and resolution authorities. There are different possible outcome scenarios for the data exploration, depending on the choices on semantic/syntactic and infrastructure level integration made for the preceding reporting process chain elements. Each outcome scenario will have specific implications for the data exploration. The table below describes the different outcome scenarios of various integration levels achieved in the preceding elements of the reporting process chain described in the previous sections. The table also highlights the constraining conditions that may hinder data exploration activities, therefore limiting the potential benefits that could be further achieved on an integrated reporting platform.

Integration Level	Outcome Scenario
Semantic-only	• When exploring the data, the users will be able to understand the data definitions and **effortlessly** understand commonalities and differences across the data definitions for multiple data collections. N.B.: if the syntactic and infrastructure levels of integration are not achieved, the following constraining conditions may hold: • Syntactic-level integration not achieved --> The reconciliation of the definition will have to be performed **manually (i.e. by visual and mechanical inspection by the user)**, though there will be no ambiguity on the identification of the integrated concepts. • Infrastructure-level integration not achieved --> Users may need to implement or buy or develop multiple data exploration tools based on different technologies.
Syntactic-only	• Data users will be able to effortlessly explore the data definitions and data collections, but they will not be unambiguously able to find common concepts among the different collections. For example, if a data item A stands for the same concept of data item B, the user will have to fully rely on his/her subjective interpretation of the relationship between these two data items. N.B.: if the semantic and infrastructure levels of integration are not achieved, the following constraining conditions may hold: • Semantic-level integration not achieved --> Users will have to rely on their own judgment for the identification of the same concepts across

multiple data collections (limited reuse and cross-capitalisation of the collected data).

- Infrastructure-level integration not achieved --> Users may need to implement or buy multiple data exploration tools based on different technologies.

Infrastructure-only	• Data users will be able to use their own (as-is) technology stack to access the data systems and will be able to explore the data entirely manually as much as their technology allows. For example, they will be able to manually query the data system to receive the list of available tables, their definitions and their content. N.B.: if the semantic and syntactic levels of integration are not achieved, the following constraining conditions may hold: • Semantic-level integration not achieved --> Users will have to rely on their own judgment for the identification of the same concepts across multiple data collections (limited reuse and cross-capitalisation of the collected data). • Syntactic-level integration not achieved --> Using available technology, users may need to develop new tools or existing ones to achieve automation in the data explorations at the above levels.

131 **Objective:** If there are different, misaligned systems, they may have different capabilities and therefore integrated data exploration may be cumbersome, as exemplified above with three outcome scenarios. However, as mentioned above, the data exploration function offers less degrees of freedom in the choice of the integration level than the other elements in the reporting process chain, as it is indeed highly constrained by the integration choices for the data definition and data collection functions. Specifically, if the other elements do not achieve a level of integration, it may not be feasible to achieve that integration level in the data exploration function, as manual activities will be needed to overcome the constraining conditions identified in the above outcome scenarios.

132 Although the CRR article does not explicitly refer to data exploration, this functional group includes capabilities that could be directly relevant to support the article's objective. It could also leverage on the availability of a central data collection system to similarly support cost reduction in the execution of other business activities relying on data exploration. Specifically, the following two groups of capabilities[39] can be shown:

133 **Data exploration for data auditing purposes on the entire system or more briefly, standard auditing capabilities**: These capabilities are directly linkable to the article's goal. These

---

[39] https://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/reporting/html/index.en.html

capabilities can enable users to explore the business metadata and follow the data lineage for various purposes, including verification of the data transformations, auditing, data and concepts identification/exemplification, etc.

134 **Data exploration for analysis purposes or more briefly data analysis capabilities**: These capabilities are indirectly linkable to the article goal as they are further capitalisation of the integrated data for the establishment of a new platform for data analysis services. Similar to the integrated reporting platform, such a data analysis platform could further contribute to reducing data exploration costs (besides reporting costs). These capabilities can be built on a second platform (to be designed) that can host banking supervision services. These services will be defined and agreed on with the stakeholders, depending on business needs and priorities, primarily on the reduction of data requests to competent authorities and individual reporters. However, an example of first pilot services could be the supervision services internally in use at the EBA, such as the Self-Service Business Intelligence and various Dashboards (e.g. Funding Plans, etc.), the operation of which will be restricted to follow the policy of the input data.

135 **Expected costs and benefits of integration:**

	Costs	Benefits
Competent authorities	• High one-off costs to develop one system for the integration of the other elements in the reporting process chain (definition and data collection) (**financial, time, resources**), these will however be partially offset by lower data analysis services costs. • New costs arising from the governance activities (**resources, time**). • Potential sunk costs due to the dismissal of internal investments on the data exploration function (**financial, time, resources**).	• Seamless data exploration and analysis of multi-granular and multi-framework data (**usability, reusability, clarity, efficiency**). • Reduced total cost of ownership of data analysis services, deriving from economies of scale and of scope, therefore potentially increasing further the return on investment for integrated reporting (**financial cost reduction, fewer resources**). • Higher level of data consistency for cross-border activities (**consistency**). • Limited expertise required and lower capital barriers for the implementation of data exploration (including data analysis) capabilities (**fewer resources, financial**). • Lower system development and maintenance costs due to economies of scales and cost sharing among the national competent authorities (NCAs) (**financial, fewer resources, efficiency**).

| Institutions | • None expected. Subscription charges may be applied for the use of data exploration services (on own data), pending governance decisions. | • Limited expertise required and lower capital barriers for the implementation of data exploration (including data analysis) capabilities (**fewer resources, financial cost reduction**).
• Potential benefits from lower operational costs of data exploration services (**financial cost reduction**).
• Potential benefits (long term) arising from service standardisation (**simplification, clarity, efficiency, fewer resources**).

If users can reuse the data from other frameworks there may be less data requests. This is not only dependant on the developments performed by authorities on this matter, but also dependant on how each bank is managing their internal reporting systems. From the fact finding, we could infer that banks are also in a process to define their reporting systems and data traceability. |

136 **Feasibility**: For data exploration, feasibility considerations result directly from the feasibility assessments for the antecedent data functions (collection and transformation). Within this function, the possible outcomes (depending on the integration level) can be more evidently visualised, and therefore outcomes are to be considered in the feasibility assessments of the antecedent functions. N.B. If these functions do not achieve integration at one level, it becomes unfeasible to achieve integration at the same level in the data exploration function.

137 **Dependencies and constraints:** The data exploration function depends on the antecedent functions of data definition and data collection. As such, the deeper the degree of integration achieved in the antecedent functions, the easier and more feasible it becomes to achieve a deeper degree of data exploration integration (i.e. joint semantics, syntactic and infrastructure) which can in turn support the automation of non-knowledge intensive data exploration activities in data analysis processes and cases. Specifically, referring to the above outcome scenarios, the constraining conditions may be more difficult (if not impossible) to overcome with automation, i.e. they may only be overcome with manual/human intervention, thus limiting the realisation of the benefits of the integrated reporting for data.

Questions to consider

6) Do you agree on the holistic approach used and the assessment done for the integration assessment (different steps of the reporting process chain and different levels of integration? What solutions should the EBA investigate in these areas that could help to reduce reporting costs?

7) Please specify any further costs or benefits you envisage related to the different stages of the reporting process chain.

4. Data dictionary

4.1 Considerations

138   This section describes the characteristics and requirements of the common data dictionary that could support the envisaged European integrated reporting system.

139   For the purpose of data definition, the proposal is to have a unique data dictionary using one shared vocabulary to support all frameworks. Following the objectives of integration of data from different frameworks, the option of using one unique dictionary is expected to be more efficient than maintaining two or more dictionaries and correspondent duplication efforts of defining or mapping definitions and aligning rules of different dictionaries and methodologies.

140   This unique data dictionary should cover the characteristics and demands of the different frameworks: the different sources of regulation, the characteristics of the different types of data, the frequency of changes, the complexity of the calculations, the data reconciliation processes and the data quality requirements.

141   The common and unique data dictionary would cover the prudential, resolution and statistical reporting frameworks that the institutions have to report: the European-wide and the national frameworks, the regular and the ad hoc reporting for supervision, resolution and statistics. Their different demands should be incorporated into the same data dictionary taking into consideration their specificities.

142   Regarding national and ad hoc requirements, the data dictionary should enable the metadata and transparency on the request definition in order to reduce additional demands, but some leeway must be guaranteed to allow any collection that would respond to specific or urgent needs.

143   In addition to the perspective of the authorities, the analysis focuses on how the data dictionary can contribute to reducing the problems identified by reporting institutions in the Fitness check on prudential reporting at the EU, published by the European Commission.

144   The discussion on the data dictionary requirements took into account the previous work regarding data dictionaries for integrated data collections in the ECB and the EBA, and some national integration efforts – Austria, Italy and Spain – whose experience and lessons learned were very useful to check and confirm the design of the future common data dictionary. The existing experiences on data dictionaries and integrated reporting bring to the discussion important references that help to shape the future data dictionary and at the same time, enable identification the costs and benefits of the existing experiences.

145   The characteristics of the proposed data dictionary are aimed at achieving an effective end-to-end digital regulatory reporting chain that ensures the regulatory-specific purposes in a more efficient way, minimising the overall costs of the stakeholders involved, and in particular, reducing the burden on reporting institutions.

146   Requirements for a data dictionary are considering the complete lifecycle of the regulatory reporting, starting with the stage of data definition and creates the data dictionary, which supports the processes of collection, validation, transformation (including calculation of derivable information), analysis, regulatory disclosures and dissemination of data.

147   The envisaged data dictionary should address three different layers: the semantic requirements, the syntactic formal capabilities and the infrastructure means.

148   The semantic integration of data from different frameworks depends very much on their regulatory constraints, their underlying data concepts and convergence possibilities. The data dictionary should be able to include all the different semantic definitions of the frameworks it supports, and all the semantic integration obtained, independently of further future achievements.

149   The syntactic data dictionary is the model structure prepared to support the formal, standardised and consistent translation of all the data concepts of the different regulatory frameworks. A syntactic data dictionary for integration should enable the data comparability across different frameworks and be ready to support automation and digital processing.

150   The data dictionary should be a platform of common understanding of the data of different frameworks. It should be agnostic to any technology but available for sharing to all the stakeholders involved in regulatory reporting. The data dictionary infrastructure should use the syntactic data dictionary (structure) and the semantic definitions (content) to facilitate the system interoperability along the different processes of the digital regulatory reporting.

4.2   Frameworks – data differences

151   Data dictionaries include the metadata about the data elements required by the different regulatory frameworks, definitions of the data elements, as well as the rules for their use and application.

152   A preliminary overview of the profile of data of the different frameworks and the current integration status are relevant elements to explain the data dictionary requirements.

153   Data for prudential and resolution purposes is requested at different levels of granularity depending on the subject and type of data. While some of the collected data is granular, a significant part of the harmonised prudential and resolution data needs to be calculated by the reporting institutions and calculations cannot be performed outside their systems. Some simpler calculations like aggregations could be described in the data dictionary and may be calculated outside the reporting institutions systems.

154   When envisaging common transformation rules, from a legal standpoint regarding CRR requirements, banks are and should remain responsible with the calculation of their ratios. Most of the prudential data comes from internal models owned by the banks themselves.

155   The data for statistical purposes is requested at a more granular level and the statistical aggregates can be described in the data dictionary and may be calculated outside the reporting institutions systems.

156   All the three frameworks require a data dictionary able to define the more granular data and able to define the aggregation formulas that can be performed outside the reporting institutions' systems. The prudential and resolution frameworks require a data dictionary with special features to define consistently different levels of granularity and the more demanding set of complex calculated variables.

157   The prudential, resolution and statistics reporting have been changing at different rhythms. Prudential and resolution harmonised regulation provided under the EBA mandates are very much determined by European level 1 legislation resulting in a highly intensive calendar of changes implying often annual revision of each reporting framework and because of the different calendars it originates new reporting frameworks added more than two or three times in a year. It is of utmost importance that prudential and resolution reporting requirements stay fully aligned with the underlying regulations and accounting standards. Any discrepancies would create additional compliance costs. The statistical harmonised regulation is determined by the ECB and has been more stable, with the significant revisions made at medium/long term depending on the frameworks.

4.3   Frameworks – current state of integration

4.3.1   European harmonised reporting

158   The European regulatory reporting of prudential, resolution and statistical data have different data dictionaries defining the characteristics of the data required and the related elements that contextualise and support the data definition e.g. vocabulary, templates. The data dictionaries include also the data validation / transformation rules for data quality assurance.

159   On prudential and resolution reporting the EBA implemented the DPM data dictionary, which integrates all the data definitions included in the reporting regulations produced by the EBA and the reporting requirements defined by the SRB. The DPM integrates under the same common and unique data dictionary all the different frameworks produced since 2013. By applying the same methodological approach, the same data model and unique vocabulary, the integrated reporting achieved the non-redundancy and data comparability of each and all of the data concepts requested of institutions. The DPM is available as a free public service and published in different formats to target different kinds of stakeholders' needs. The DPM is the central element of the XBRL taxonomies used in all EU countries by all NCAs in the second level of reporting, and by an increasing number of institutions at the first level of reporting.

160   The DPM has around 70,000 different and integrated data concepts and support a number of reported values, which are circulating from institutions to national authorities and in the second level of reporting to the ECB and the EBA. All prudential and resolution data is stored in an integrated and comparable way in the European Centralised Infrastructure for Supervisory Data (EUCLID) the EBA European data hub.

161   On statistical reporting the current situation shows high heterogeneity across the frameworks. The development of the frameworks were exclusively focused on the initial purposes of each particular reporting statistical area. This approach lead to different dictionaries, creating compartmentalised frameworks, complementary to each but not integrated.

162   The IReF initiative of the Eurosystem aims to consolidate the ESCB statistical requirements and stems from the main objective of reducing the reporting burden. The intention is for the ECB to issue the statistical data requirements for deposit-taking corporations in a dedicated ECB regulation rather than in separate legal acts, and for the requirements to be directly applicable to euro area deposit-taking corporations, without any translation into national collection frameworks. In order to effectively integrate the existing requirements, the IReF Regulation will encompass a set of requirements with different levels of granularity that will consolidate the existing reporting lines in a unique framework and avoid any duplication of the requirements. The reporting scheme will be covered in the legal act on the basis of a standardised data model and dictionary, thus ensuring standardisation of the definitions and methodological alignment with statistical standards.

163   Work on IReF has progressed significantly in a joint effort of the ESCB. The implementation date is envisioned for 2024-2027. Currently there is an ongoing IReF questionnaire asking for input from banks (deposit-taking institutions actually) on a series of aspects that will form the basis of the design of the statistical reporting.

4.3.2   National regular reporting

164   On prudential data, the national authorities collect from reporting institutions the exact harmonised data defined at European level by the EBA. Under the maximum harmonisation principle, national authorities have the power to request additional data from reporting institutions only outside the scope of the harmonised data. Some national authorities extend the DPM data dictionary to incorporate their national regulatory requirements together with the EBA European harmonised data into the same unique data dictionary.

165   On resolution data, the minimum harmonisation principle enables national authorities to request specific national data in addition to the minimum requirements defined at the EBA European harmonised requests.

166   On statistical data, national central banks are allowed to collect the statistical information necessary to fulfil the ECB's statistical requirements as part of the statistical reporting

framework they have established under their own responsibility. This solution dates back to the establishment of the European Monetary Union and was well justified at the time as it meant that statistical reporting could be founded on well-established national reporting approaches. Over time, it has become inefficient and hence costly for the banking industry (especially in the case of cross-border banks).

167   This way the IReF aims to integrate not only the different statistical frameworks required at European level, but also to reformulate the national statistical reporting frameworks to reduce costs on reporting institutions. Currently there is an ongoing IReF questionnaire asking for input from banks (deposit-taking institutions actually) on a series of aspects that will form the basis of the design of statistical reporting.

4.3.3   Non-harmonised reporting

168   In addition to the regular data, there are a number of additional ad hoc or recurring data requests from supervisory, resolution and statistical authorities, of which the exact volume and characteristics are unknown, as they are not registered. An overview of these requests is provided in Section 2, Stocktake of current data requests.

4.3.4   Integration gaps

169   The European-wide regulatory reporting of prudential, resolution and statistics are at different stages of semantic integration and syntactic implementation.

170   On prudential and resolution frameworks, all the different frameworks are semantically integrated and all the data definitions have being formally translated and stored in the DPM syntactic data dictionary by a data standardisation process in place since 2013 and performed by the EBA with the collaboration of national authorities. The data of resolution reporting of the SRB is also integrated with prudential and resolution data defined by the EBA and stored in the same data dictionary.

171   The data standardisation process uses a consistent and formal approach and a common data vocabulary to define all the new framework data concepts and their data validation rules. The data dictionary definitions are publically available in a database together with other EBA infrastructure outcomes, like the standard taxonomies for digital data exchange.

172   The DPM data dictionary is a metadata repository which is related to all Level 2 reported data and used by the EBA and some NCAs to structure the storage of data in their databases (EUCLID in EBA).

173   On statistical, all the different European-wide frameworks are not integrated and have different data dictionaries. The ESCB has started the IReF initiative and plans to start a project in 2024-2027 to integrate the different independent statistical systems. The approach will be also to have a unique data dictionary for all statistical frameworks. The single data dictionary

(SDD) is still pure metadata repository not bridging metadata with data received. This gap is important in terms of operationalising the systems at implementation stage

174   At national level, some authorities are using the DPM and have extended it to integrate their national requests on supervision, resolution and statistics. Others have their own data dictionaries to integrate national discretions and others do not integrate the national requests with the harmonised European data.

175   The different approaches followed by authorities doesn't enable having a complete picture of the dimension and scope of national reporting requests and makes it even more difficult to know the same on ad hoc reporting.

176   Each authority has its own infrastructure and the only sharing are the XBRL taxonomies produced by the EBA for data exchange and the XDMX taxonomies produced by the ECB for data exchange.

177   This segmented approach is responsible for a less efficient model with missing opportunities of semantic integration and duplication efforts on redundant data dictionary efforts.

178   The possible semantic overlapping can exist in between the European-wide reporting and the national reporting at prudential, resolution and statistical frameworks.

4.4   Requirements for integrated reporting

179   The data dictionary should address different types of data: different levels of granularity, harmonised or non- harmonised data, regular or ad hoc data, quantitative and qualitative data originated in different regulatory frameworks and regulators.

180   The Data Dictionary should include as essential components:

- ▪ dictionary of vocabulary and data concepts;

- ▪ dictionary of data validations and transformations.

181   The data dictionary for data integration should be a formal and standardised data dictionary (syntactic data dictionary) with all the elements to enable automation and digital processing of regulatory data. The syntactic data dictionary facilitates any effort of semantic integration. The setup of this syntactic data dictionary is feasible and should be in place as the central piece of the integrated reporting system.

182   Integrated reporting requires the setting up of a central common and unique data dictionary, with the following characteristics:

- **comprehensive:** all different data scopes and granularities; data out of dictionary is data impossible to integrate;

- **incrementally implemented:** it should be updated with each new extension of the contents of the data dictionary, reusing the existing elements or adding new ones if necessary;

- **complete:** all formal and standardisation elements that enable the digital processing of reported data along the different processes of the data chain;

- **centrally manage:** to ensure the technical standardisation and create the necessary data definition consistency and technical quality; each new semantic integration should be included consistently in the existing data dictionary. On national extensions, ensure NCAs can share the tasks accordingly with their respective fields of competence;

- **centred on a common and unique vocabulary**: unique single vocabulary covering all data to support not only the data collection;

- **focus on data comparability:** the principal objective of a data dictionary for integration is to define data consistently and achieve comparable data definitions and relationships between distinct data granularities, alignment of data frequencies and reference dates, clear identification of consolidation and clear identification of the characteristics of target reporting institutions;

- **common platform of understanding** for all stakeholders involved;

- **ready for digital processing:** the data dictionary is the central piece of data-driven and effective system integration. Data dictionary should be technology-agnostic and compatible with any data exchange standards;

- **ready for human interface:** easy to use and understand;

- **serving all regulatory data chain processes:** from data collection, data validation and transformation, data analysis and disclosure to support system interoperability, the different processes and stakeholders from data collection to data validation and transformation, data analysis and data dissemination.

183   Annex 2 describes the necessary characteristics in more detail, separately from authorities' and institutions' perspective.

Questions to consider

8) Do you use one or more data dictionaries in your compliance and reporting processes?

9) What are the characteristics you think a data dictionary should have? Do you agree with the references in this document? Do you think any characteristic is missing or should not be included?

10) What is the role you think the data dictionary can have in regulatory compliance and reporting?

11) How would a standard data dictionary help institutions to improve the processes of:
    a. understanding reporting regulation;
    b. extracting data from an internal system;
    c. processing data (including data reconciliation before reporting);
    d. exchanging and monitoring regulatory data.

12) How important is it for institutions to have a unique and standard data dictionary for all regulatory data with the aim of ensuring consistent use across supervisory, resolution and statistical reporting?

13) How much would it cost to move to a unique regulatory data dictionary?

14) How much cost reduction is expected by integrating the national regulatory reporting with the harmonised reporting regulation into a unique data dictionary?

15) How much cost reduction is expected by integrating ad hoc regulatory reporting with harmonised regulation into a unique data dictionary?

4.5 Data dictionary – Costs and benefits

184   From a data dictionary perspective, the implementation of a common and unique syntactic data dictionary will be essential for the effective implementation of an integrated reporting system. The data dictionary will affect all organisations and people involved in the reporting, including the general public and other less active stakeholders in the regulatory chain. The following table includes the costs and benefits from the perspective of regulators and reporting institutions.

Costs	Benefits

| Medium to long term | Competent authorities | Additional responsibilities for supervisors to decide and coordinate semantic integration initiatives that involve frameworks from different regulators and different legal restrictions (**time, resources**). Evaluation and decision of feasible semantic integration (**time, resources**). Evaluation and decision of feasible common granular reporting (**time, resources**). Semantic integration of frameworks from different regulators (**time, resources, financial**). The cost of the design of the common syntactic data dictionary can very much be reduced by relying on the experience of integration already existing to different degrees across regulatory products and the dictionaries already used. Significant costs involved in the implementation of the syntactic data dictionary to support the central data collection point and the different integration initiatives (**time, financial, resources**). Learning curve effects on adopting a new regulatory data dictionary (more costs and less visible benefits in the initial phases) (**time, resources**). Change of existing systems to adapt to a new syntactic data dictionary (**time, financial, resources**). Semantic data definition, data standardisation and creation of infrastructural tools and other outcomes supporting reporting phases (collection, transformation and exploration) (**time, financial, resources**). Coordinated management of communication on regulatory outcomes related to data dictionary (**time, resources**). | Creation of comprehensive information of all regulatory frameworks, their value and scope of application (permanent stocktake availability) (**usability, clarity, efficiency, simplification**). Increase the scope of data comparability and uniform mechanisms to data exploration and dissemination (**consistency, clarity, usability**). Creating a level playing field for analysis by implementing and sharing the enrichment data processes that support different types of analysis: comparative analysis, time series, ad hoc analysis, predictive analysis (**usability, efficiency, consistency**). Increases data sharing between authorities and enables the collection of data only once (multi use of data), thus reducing redundancies (**efficiency, reusability**). Additional possibilities for data users to process and analyse data across different frameworks and formulate their needs in terms of data ('Which data do I really need?'; 'What will be my focus?'; 'What questions should the data address at a minimum?') (**usability**). Learning curve effects on adopting a new and common regulatory data dictionary (increasing efficiency benefits) (**efficiency**). |

| Institutions | • Costs of introducing the new data dictionary in their reporting processes (**time, financial, resources**).
• Learning curve effects on adopting a new regulatory data dictionary (more costs and less visible benefits in the initial phases) (**time, resources**). | • Contributes to producing clearer and more structured reporting rules (**clarity, usability**).
• Contributes to a common understanding amongst people with different roles, backgrounds and skills involved in the reporting process (**usability, quality improvement**).
• Creates better data access and eases the preparation of valuable data reports with interest to institutions and other stakeholders, returning value to reporting institutions (**usability, efficiency, simplification, clarity**).
• Creates a level playing field in the application of the requirements, by ensuring everybody is sharing the same common understanding on the data requested and transformations performed (**consistency, clarity**).
• High level of integration of regulatory requirements available to be efficiently applied in institutions' digital solutions (**efficiency, usability**).
• Enables interoperability within and across institutions (**interoperability**).
• Enables standardisation and integration with other national and international standards (**consistency, clarity, usability**).
• Enables following the same approach in a clear and transparent manner across different regulatory obligations (**clarity, consistency, streamlining, efficiency**).
• Learning curve effects on adopting a new and common regulatory data dictionary (increasing efficiency benefits) (**efficiency**). |

Questions to consider

16) Do you agree with the costs and benefits highlighted in the chapter? Do you see other benefits and costs when implementing a standard data dictionary?

4.6 Granularity

185   This section investigates the feasibility of increasing the granularity of the data to be collected with the purpose of: i) meeting regulators' objectives and facilitating institutions' compliance with reporting requirements (increasing the efficiency of the reporting process); and ii) facilitating the work on integrating the prudential, statistical and resolution data.

186   Increases in the granularity of the data collected for statistical purposes is already under discussion, as part of the IReF[40] project. For this reason, this section does not focus on the granularity of the statistical data, although at some point the degree of granularity between data frameworks may need to be coordinated or reconciled.

187   This section focuses on i) an overview on the topic of granularity; ii) costs and benefits of collecting data at a more granular level; iii) options regarding the granularity of a possible integrated reporting system; iv) transformations; v) additional challenges to be considered.

188   Going further, the decisions on integration with respect to granularity can be considered gradually in the integration process, moving from the current level of granularity in the data collected to a higher granularity– as further analysis and experience will show the extent to which this will be possible and cost effective.

4.6.1 Definition

189   Different authorities need information on business concepts/phenomena that can be quantified[41] (measured or defined) and therefore further analysed. 'Loans', for example, can be understood as such a business concept/phenomenon.

190   Granularity is the level of detail at which the business concept is represented/defined. Full granularity would mean describing the concept in all its measurable or defined constituent parts (to what level of measurable or defined detail the loan can be represented). Aggregated data would be anything that is not fully granular. In between fully granular and fully aggregated there are different levels of aggregation (levels of granularity) at which the data can be represented.

---

[40] https://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/reporting/html/index.en.html

[41] E.g. information on the amount due on a loan can be easily quantified as opposed to more abstract concepts such as the willingness of a debtor to repay the loans (for such concepts more quantifiable proxies can be used)

191   A higher level of aggregation of the data is obtained from less aggregated data (i.e. more granular data) by performing some aggregation operations or more complex calculations across one or multiple dimensions in which the less aggregated data (more granular data) is represented (e.g. nominal value of total loans in Europe, representing more aggregated data, is obtained by summing nominal value of total loans in each European country, representing less aggregated data). The type of aggregations that could be defined can range from very simple operators (e.g. additions) to highly complex transformations.

192   Increasing the granularity of the data would mean going in the inverse direction which involves much more than just reversing the process of aggregation described above. This would entail describing the granular and the more aggregated data explicitly using the same constituent parts (e.g. describe the value of total loans, representing the more aggregated data, by identifying all the relevant characteristics of each individual loan that are used in the aggregation or calculation of the aggregated data).

193   Reporting requirements have been defined at different levels of granularity reflecting the underlying regulation purpose and responding to different policy needs. Devising ordered levels of granularity (equivalently orders of aggregation) in a granular (aggregated) perspective is the result of the business expert knowledge to divide the data based on economic/geographical/business and other domain-specific setup and according to their needs. This effort depends very much on the nature of the business concept, but also on the regulatory and financial markets evolution and for this reason it is never possible to ensure they will be stable.

4.6.2   State of play

194   **Prudential reporting** covers mostly the collection of aggregated data with various, limited breakdowns across different categories (sectors, currency, geography, exposure class). There are some exceptions of granular data reporting but they are very limited in scope. The concepts defined are usually complex and banks may require a complex set of transformations to produce the regulatory figures. The reporting institutions are responsible for the calculated values, which need extremely high accuracy as prudential decisions and actions are based on the reported data. Therefore, a reliable data quality process is very important. Moreover, supervisors are interested in assessing the ability of reporting agents to calculate and report accurate data, part of the assessment of internal governance within the SREP (which also includes checking institutions'compliance with the Basel Committee on Banking Supervision's standard number 239 (BCBS 239)[42]). Supervision is focused on the supervision of a consolidated group rather than a solo institution therefore the ability to infer the data values in the right context (e.g. consolidated values) is of utmost importance (see more in Section 4.6.8).

---

[42] https://www.bis.org/publ/bcbs239.htm

Questions to consider

17) What would be the implication of granular data reporting on the institutions' compliance with BCBS 239 (also in the context of the options presented in Section 4.6.5)?

195   **Resolution reporting** is similar to prudential reporting in terms of the complexity of the concepts covered and the need for accurate reporting. It covers aggregate data with various, limited breakdowns across different categories and it has a larger collection of granular data reporting compared to prudential reporting. Similarly with supervisory activities, resolution activities are focused on resolution groups and depend on the resolution strategy, therefore the data values should be available at the needed consolidation (see more in Section 4.6.8).

196   **Statistical reporting** aims to calculate time series and growth rates of broad economic concepts such as monetary aggregates and lending to the economy. The data collection has recently shifted towards more granular reporting (loan-by-loan, security-by-security data) although it has retained a significant amount of data collected at a more aggregated level (e.g. loans at country level). The more granular data collection from the perspective of a statistical regulator still requires from institutions a certain level of aggregation efforts. Statistical data is mostly used to create various aggregates across many dimensions, users being mainly interested in cross-sectoral or cross time trends and developments as opposed to information on a precise data point. For these reasons, information on changes in stocks due to reclassifications and price changes is required for statistical purposes to calculate growth rates for each sector, while this is not the focus in prudential and resolution reporting. Concepts defined for statistical reporting have more straightforward definitions (less complex concepts compared to prudential and resolution concepts) and are harmonised by international and EU standards (e.g. System of National Accounts (SNA) and European System of Accounts (ESA) 2010). The transformations applied to obtain aggregated figures are conducted uniformly across various reporting agents although the heterogeneity across them requires some adjustments (as opposed to supervision and resolution where the data is aimed at precisely depicting the specificity of the reporting institution and cannot be adjusted by authorities).

4.6.3   Granularity considerations in the reporting process chain

197   The level of granularity of data raises multiple requirements through the reporting process.

Data definition of the collected data:

- The concepts defined in the data dictionary (semantic level) should account for the granularity of the data it depicts. In addition, the same concept could be represented at different granularity levels. A more granular dataset might imply the need for common vocabulary standards across different domains to ensure correct identification of concepts.

- The structures in which the concepts are represented (syntactic level) should store all the metadata referring to different concepts at different aggregation levels. The system should identify where the same concept is defined at different aggregation levels (e.g. trade receivable loans that are defined as an aggregate as opposed to trade receivable loans that are defined at instrument-by-instrument level). In addition, it should identify uniquely the same concepts and the similarities between concepts irrespective of the granularity at which they are defined (e.g. the concept of loan is related to all data calculated with loans).

- The tools to deal with the metadata (data definition at semantic and syntactic levels) should account for the different needs that granular data and/or aggregated data might need in terms of the capacity and performance of the system (e.g. metadata of more granular concepts might imply different visualisation needs, metadata lineage between granular and aggregated data).

- The metadata of the data dictionary should be defined in a digital-friendly way supporting the efficiency of the data collection processes.

Data collection:

- Appropriate reporting rules (see Section 3.2.4) should be defined depending on the granularity (e.g. timeliness and frequency of granular data may be different than for aggregated reports).

- Data exchange formats should be best suited for dealing with more granular data and accommodate aggregated data at the same time. Possible validation rules for different levels of aggregation (e.g. totals) are required.

- IT infrastructure should account for the volume of data implied in the more granular collections.

Data transformation:

- Transformations (semantic level) will have to be defined to link the data across different aggregation levels (going from granular to more aggregated) or to derive new concepts.

- Decisions on the proper representation of the transformation rules will have to be considered (syntactic level). The model would also have to link the transformations to the underlying data structures they refer to (granular input and more aggregated output).

- The necessary infrastructure that would facilitate the process of developing transformations and enable running them and performing the required checks will have to be in place (infrastructure level).

Costs and benefits of increasing data granularity

198   For statistical data, increasing the granularity of the data collected is already a topic of discussion in the IReF project. A decision on the level of granularity of the data collected for statistical purposes will be decided following the cost and benefit questionnaire currently ongoing. For this reason, the costs and benefits below will mainly focus on increasing the granularity of the prudential and resolution reporting although many of them are also valid for statistical collections.

199   Any decision on collecting more granular data (in comparison to the current level of granularity) is closely linked with the ability to define transformations to ensure the more aggregated data needed by authorities can be obtained (see Section 4.6.7).

200   In addition, for supervisory and resolution purposes, banks have to remain responsible for all the data. Therefore, for some 'important' data, both aggregated and granular data would have to still be reported while for some other 'less important' data it is assumed authorities would not need to check the reconciliation at the more aggregate level. This can be ensured by defining feedback loops and anchor values to be reported (see Section 4.6.7)

201   Increasing the granularity of the collected data for prudential and resolution purposes might provide for the following costs and benefits.

Costs	Benefits

Medium to long term

Competent authorities

- (Significant) duplication of some data transformations (including validation rules): the calculation of some regulatory data needs to be done by both regulators and institutions, as the latter would be required to ensure regulatory compliance (**time, resources, financial**).
- The quality assurance process might be more costly. This might be due to: i) the reconciliation process in the aggregated figures by running in parallel transformations on the supervisor's side and transformations in the institutions and the identification of discrepancies (feedback loops / anchor values); ii) the need to implement quality criteria on the granular data and in addition all the current validation rules on the supervision and resolution frameworks; and iii) possibly increased number of resubmission (**time, resources, financial**).
- (Significant) coordination effort among authorities to maintain and enhance a common granular collection layer to be used for various needs (**time, resources**).
- (Significant) costs for the coordination of maintaining and enhancing a common set of transformation rules (i.e. aggregation rules and requirements mapping depending on the governance setup with respect to transformations) (**time, resources, financial**).
- Maintenance costs might be significant as the common regular reporting and the set of calculations are evolving each time as new reporting requirements are added or data is redefined (**time, resources, financial**).
- Possible risk transfer from institutions to authorities on detecting and acting upon breaches (e.g. supervisors might need to act upon additional insights from the granular data and might have reputational and legal implications) (**time, resources**).
- Authorities might run the risk of being disputed by institutions (e.g. legal risks), if granular data is collected and aggregated by authorities (the setup of the transformation process and who bears the responsibility for it is under discussion (see Section 4.6.7) for the respective purpose without an efficient and well-functioning feedback loop. This is due to the fact that without a sufficient control / double-check opportunity institutions potentially cannot be held responsible for the (aggregated) data used by regulatory authorities (**financial**).

- More flexibility in the usage of data through the possibility to transform it to respond to new policy needs or support additional analysis (possibilities of data users to process and analyse vast quantities of data, reducing additional requests) (**reusability, efficiency, streamlining**).
- Higher comparability of the data received, uniform implementation of requirements and transparency in the aggregation process (**clarity, consistency**).
- Enables the collection of data only once (multi use of data), thus reducing redundancies (**reusability, efficiency**).

Institutions

- Costs related to the collection/compilation of the data at a more granular level from various systems in the institution in a timely manner (**time, financial, resources**).
- Granular data reporting might not exempt banks from reporting (and therefore also calculating internally) additional aggregated data to ensure compliance with the regulatory ratios and other risk indicators (e.g. feedback loops and anchor values (see also Section 4.6.7) (**possibly some cost reduction).**
- (Significant) costs for the maintenance and enhancements of a common set of transformation rules (depending on the governance setup with respect to transformations) (**time, financial, resources**).

- More stable or reduced new reporting requirements over time as various ad hoc requests to fill information gaps may become less frequent, given that data may already be collected as part of an increased granular collection layer (**efficiency, streamlining, fewer resources**).
- Further cross-country harmonisation and standardisation of national reports with particular benefits to large groups that operate across borders within the EU. Although the European-wide data collection already follows the maximum harmonisation principle for prudential data, going more granular in the data collection might further incorporate part of the current additional national requirements (national extensions) making them no longer necessary (as they will be part of the harmonised granular collection). This aspect might be more relevant in the case of resolution data that is based on a Directive (which allows more room for national extensions)(**efficiency, consistency, streamlining, clarity**).
- Creates a level playing field in the application of the requirements by ensuring everybody is following the same approach: clearly and transparently defined transformations, which nevertheless might still require some verification and compliance process

conducted by the responsible authorities) (**clarity, consistency, usability**).

- Data reported might become closer to the data stored in the banks' systems, implying a potential simplification of the internal processes of reporting preparation (**fewer resources, simplification**).
- Reduced duplication in the data reported as some of the concepts defined at granular level might be used to calculate information for more than one template (**efficiency, streamlining, reusability**).

202   Initial financial and human resource costs on the side of both competent authorities and banks would be needed to ensure a smooth transition to more granular reporting. Challenges for banks are mainly related to data quality (new standard in terms of completeness, accuracy, and timeliness), data availability and infrastructure. Challenges for authorities are largely related to coordination and governance aspects in setting up the data dictionary, classifying granular data down to the minutest detail, infrastructure costs as well as possible legal aspects that need to be overcome to allow the collection of data that is more granular. In addition, designing and implementing the transformation rules will require a significant amount of effort and coordination between authorities and institutions. In addition, as will be clearer later in the report, a large part of prudential and resolution reports might not be possible to be generated in a cost-effective way from granular reporting (e.g. instrument level reporting).

Questions to consider

18) For which reporting areas (prudential, statistical and resolution or modules/parts of these areas) may the use of granular data present a solution?

4.6.4   Evidence from cost of compliance study

203   Evidence from the Cost of compliance study[43] shows that a large share of responding institutions are rather sceptical towards the potential benefits of moving to a granular data

---

[43] Cost of compliance with supervisory reporting | European Banking Authority (europa.eu)

collection – be it a potential reduction in the number of ad hoc requirements, simpler reporting or better data quality. However, as the evidence from the Cost of compliance study did not detail the meaning of 'granularity', institutions' final verdict seems to be dependent also on the way the granular collection would be set up and used. Some respondents see the following as a precondition for the success of granular reporting:

- Users need to commit to exploiting the available granular data without requesting additional attributes on a continuous basis (use, stability).

- Granular data has to refer as much as possible to data in its raw state that is available and can be extracted from the institution's data warehouse without noteworthy intervention or transformation (reporting the highest granularity level). In their view, reports produced with a high (but not the highest) level of granularity (e.g. numerous low-level aggregates by country, products, sectors) would be too exhaustive, the maintenance would be high (not scalable solutions, still some transformations needed) and the granular information might not be reused or reusable by the institution for its internal purposes.

- Granular data only needs to be provided once (i.e. not different sets of granular data on the same transaction or position for different recipients).

204   Some comments received saw the benefits of moving to granular data mainly from the perspective of moving to standardisation and harmonisation of concepts (which are entailed in the discussions of moving more granular but not only) rather than the granularity aspect per se. While large institutions might be more inclined to agree that the use of a common dictionary for all of the current reporting is easier if done at a more granular level, the medium and non-complex institutions would largely disagree.

205   Obtaining the current FINREP aggregated reports from more granular data using standard formulae and without manual adjustments was seen by many respondents as being unfeasible. Some respondents referred to the fact that consolidation adjustments (e.g. restatements) performed at holding level are not depictable at single transaction level. In addition, a complete elimination of the need of manual adjustments would possibly mean that all adjustments can and should be done in the data production process in a manner that would not affect transformations.

206   The majority of reporting institutions would prefer the current granularity defined for FINREP/COREP as opposed to going more granular.

207   The way in which the more granular collection process could be designed (accounting for certain constraints such as the need for feedback loops and/or anchor values) will ultimately infer/reveal the benefits to both banks and users to going more granular.

4.6.5   Options for the possible granularity level of the future integrated system (including statistical, prudential and resolution data collections)

208 **Setup:** In the following, the collection layer refers to the data defined and collected by the authorities. Banks derive the collection layer by applying a series of transformations to the data stored in their internal systems (the input layer of a bank). Each bank is responsible for the transformations it applies to the input data to ensure compliance with the regulations. Each bank has its own input layer that may be similar or very different to other banks' input layer. In general, the input layer of a bank is likely to have a much higher granularity level than the collection layer requested by the authorities.

209 In the current situation, each authority defines the data that is collected at its needed granularity level, in accordance with the applicable legislation. All reporting frameworks request data at different levels of aggregation regarding the collected data (e.g. certain loans are collected on a loan-by-loan basis; at the same time, loans are also collected on an aggregated basis)[44].

210 The current collection layer for the statistical data on average can be considered to have a higher granularity of the data it collects and in this way is closer to banks' internal systems. Traditional monetary statistics represent mainly aggregated data collections while the more recently issued regulations were focused on granular data collections (such as AnaCredit and Securities Holdings Statistics). The IReF project, aiming at integrating the various statistical reporting requirements, is also exploring possible increases to the granularity of its datasets with the purpose of: i) deriving the aggregated figures and releasing the banks from the requirement to report them; and ii) exploring links between different statistical reports with the purpose of eliminating data duplications and deriving concepts where possible. Where the compilation of statistical aggregates will be done by authorities, a single data point will be reported only once and could be used for several statistical purposes. It should be recalled that for statistical purposes it is possible for banks to delegate the data aggregation to authorities, while according to the CRR and BCBS 239 principles, banks should remain responsible for the aggregated data reported to authorities.

211 The prudential and resolution data collections have on average a higher level of aggregation and therefore the collection layers deviate more from the data as defined in the bank's internal systems.

212 Three possible options have been identified for how the future integrated system for data collection for statistical, prudential and resolution could look like in terms of the granularity of the datasets it collects. The options differ in the degree of granularity of the collected data and the implication this choice has for the setup of the reporting process and governance. The different levels of granularity envisioned by the options are driven by the objective of: i) meeting users requirements more easily (granularity allows a more flexible use of data, thus enabling authorities to better fill the data gaps that emerge in the presence of shocks) and increasing the

---

[44] Loans collected on an aggregate basis are not necessarily coinciding with the loan-by-loan collection, although there are overlaps

efficiency of the reporting process; and ii) facilitating the work on integrating the prudential, statistical and resolution data (integrating the granular data collection with the same data requested at a higher level of aggregation with the aim of reducing data duplication) and achieving the desired level of integration.

213 The options, as presented below, do not specify what is the desired / optimum granularity level that the collection layers should contain (how much more granular authorities should request the data). Given the two objectives above, granularity can be increased to the extent that objective ii) is achieved or granularity could go beyond the minimum needed for integration, up to the point of requesting a granularity level close to the bank's internal system (data should be requested only once at a certain level of granularity as opposed to data should be requested only once at the highest possible level of granularity). As institutions have different internal organisation of data, when defining a common granular reporting method it will be necessary to achieve the more adequate level of granularity that does not increase institutions' costs nor does it shift the costs to authorities.

214 The choice of the exact level of granularity should be determined by the benefits of granular reporting as opposed to the costs and constraints (as defined in the report).

215 In **all three options, we may consider that a common data dictionary should be defined.** The design of the common data dictionary might differ depending on the options chosen.

Option 1

216 This option considers that the current level of granularity in the prudential and resolution output layer could be maintained in the future integrated system. For the statistical data collection, the granularity will be further increased, as decided following the cost and benefit assessment launched in the IReF project. The difference in the integrated system case to the current situation (or the situation without an integrated system) is that the collection layer of each of the three reporting frameworks will be contained in a common integrated data dictionary. If the right incentives are in place, semantic integration may be achieved by closely analysing the concepts defined in each reporting framework and uniquely defining them in the dictionary where the definitions are in fact the same, irrespective of the granularity of the underlying data (details in the dictionary section).

217 **Granularity**: both aggregated and granular data will be collected, as is currently the case.

218 **Implementation**: from an implementation perspective, having both granular and aggregate data has implications for the setup of the common data dictionary as its design should allow for the identification and representation of both granular and aggregated data, in the most efficient way.

219 **Transformations:** no transformations need to be defined in addition to the current situation as the data collected is at the level required currently by the authorities. In the case of statistical data collections, transformations will be defined as decided in the IRef process.

220 **Responsibility**: as is currently the case, each authority will be responsible for defining the data collected (defining the collection layer). Each reporting institution would be responsible for reporting the collection layer in accordance with the regulations.

221 This option does not consider the costs and benefits (Section 4.6.4) of moving to a more granular collection. A few costs and benefits of this option are highlighted below:

		Costs	Benefits
Option 1	**Competent authorities**	• Possibly a more complex dictionary for the collection layer of the integrated system to accommodate both aggregate data and granular data (e.g. loans defined at different levels of aggregations) (**time, resources**).	• Easy to implement while no further changes to the scope of the data collected will be made (**clarity, usability**). • Facilitates the work for further integration (e.g. facilitates the subsequent implementation of Option 2) (**usability**).
	Institutions	• May still imply possible data duplications in the data collected, as is currently the case (**time, resources**).	• No additional data points to be reported (**efficiency**).

Option 2

222 This option considers an increase in the granularity of the data to be requested in the collection layer especially in prudential and resolution reporting (in addition, for statistical purposes, the increase in granularity is already under consideration in the IReF project). The main purpose would be to integrate the more granular data collected with the data collected at more aggregated level such that the latter could potentially be dropped from the collection layer. The level of increase in granularity will be driven by the feasibility to achieve this integration (from granular to aggregate). In this respect, this option considers that not all required data is possible to be collected at a granular level and additional aggregated reporting will be needed (e.g. where transformations cannot be defined for legal or technical reasons). This option builds on Option 1 adding that a number of aggregated figures currently requested will no longer be part of the collection layer, but they will be derived from a more granular new collection layer.

223 **Granularity:** both aggregated and granular data, however, the scope of the former will be reduced in favour of the latter (the scope for the data collection at aggregate level will be reduced compared to the current situation, while the previously aggregated figures will be obtained by means of transformations starting from the more granular new collection layer).

224 This option considers the benefits for institutions, prudential, statistical and resolution authorities of requesting data at a more granular level, but also retaining the possibility for some

data to be requested also at a certain level of aggregation either for i) legal/compliance reasons; ii) data availability; or iii) the impossibility of deriving aggregate figures in an efficient, transparent and automated way from more granular data.[45]

225 Granular data reported for statistical reasons might be reused to derive some of the concepts in the prudential and resolution figures. However, certain considerations (detailed later in this report) indicate that additional reporting requirements might be needed for the statistical data to be reused for the purpose of the other frameworks, including in some cases even going more granular.

226 **Implementation:** Option 2 builds on Option 1 in the sense that there will be an integrated common dictionary for the collection layer representing both granular and aggregated data. The difference with Option 1 is that a part of the collection layer will be common for all three reporting frameworks and there will not be a straightforward way to distinguish the granular collection layer for the statistics with the one for prudential and for resolution[46]. This is because the data belonging to the same collection layer will possibly be used for deriving more than one reporting framework. In this setup, the previously defined collection layer (from Option 1) will be significantly reduced, as a large part of the data will be derived from the more granular collection. The collection layer in Option 2 of the integrated reporting will be composed of the common granular collection layer and in addition, the data collected at a higher aggregation level as needed for each reporting framework.

227 **Transformations**: will have to be defined where the aggregated reporting has been eliminated in favour of the granular reporting. Transformations will apply to the new more granular reporting (collection layer) in order to derive the aggregated concepts needed for both compliance and other prudential / resolution activities / statistical aggregates. A series of transformations might be jointly defined for more than one framework while other transformations will be specific for a certain reporting framework, depending on the chosen implementation.

228 **Responsibility**: each authority would be responsible for defining the data collected where data cannot be derived from more granular collected data. Where more aggregate data is requested across different reporting frameworks, the authorities are responsible for coordinating and avoiding duplications. Authorities should collaborate together in defining the common granular data that will be further used through transformations to derive the needed more aggregated data. A standard process for accessing and defining data should be set in place together with the proper governance.

229 Institutions would be responsible for sending the requested collection layer (the granular and more aggregated parts). In addition, institutions will be responsible for reporting additional aggregated figures that are not part of the collection layer but for which they are responsible to

---

[45] Further details can be found in the Sections 4.6.7 and 4.6.8

[46] This depends on the actual design of the solution.

be compliant with. To ensure that banks remain responsible also for the latter from a legal perspective, the requirement to implement feedback loops and anchor values needs to be considered (see further Section 4.6.7).

230 This options takes into account the costs and benefits (see Section 4.6.4) of moving to a more granular collection. In addition, few costs and benefits of this option are highlighted below:

		Costs	Benefits
Option 2	**Competent authorities**	• Human and IT resources needed for setting up, testing and maintaining the transformation (requiring a lot of coordinated action among different authorities, involvement of business experts from different areas, new IT systems). In addition, depending on the decided governance, also possibly being responsible, or sharing responsibility for the transformations (**time, resources, financial**). • Highly coordinated action among different authorities needed for defining the common granular layer collection (**time, resources**). • Human and IT resources needed in aggregating (including e.g. quality assurance / feedback loops) the necessary information from data which is not available as aggregated anchor values (resources, financial).	• The availability of clear and transparent transformations is likely to lead to less Q&As on what data should be reported, at least at the more aggregated levels (**efficienc**y). • The availability of clear and transparent transformations would lead to a better understanding for everybody of what kind of data is reported and increase the comparability of the data reported (**consistency, clarity, usability**).

Institutions	• Additional granular data and attributes will have to be collected (possibly huge amounts of data) on top of aggregated figures (albeit a reduced amount), adding to the reporting burden (proportionality aspects-see Section 4.6.8) (**time, resources, financial**). • Increased human resources and IT costs in the reconciliation process (aggregated figures on the CA side with the ones coming from banks), requiring manual processes and where possible further investments in automated systems to smooth the process) (**resources, financial**). • Participation in the form of human resources needed for setting up, testing and maintaining the transformations (and possibly remaining responsible) (**resources**).	• Improves the quality of bank data (as possible discrepancies/errors identified at more aggregated levels will lead to further corrections also at the granular level of the data) (**quality improvements**). • Clear and transparent transformations will ensure a level playing field for the application of regulatory requirements (**consistency, clarity, usability**). • Creating and following clear and transparent transformations and standards, would facilitate the compliance of banks with BCBS 239 principles (**usability, clarity, simplification**). • Less benefits for moving to granular data collection as banks would be required to compute part of the aggregated figures and report them on top of the granular data collected (some possible cost reductions).

Option 3

231 This option considers an increase in the granularity of the entire data currently collected taking as a hypothesis that all aggregated figures could be derived from the granular data at its highest possible level. A common granular collection layer for prudential, statistics and resolution reporting requirements would result in the full data integration of the reporting requirements under the principle of define once and collect once ('once only'). This option particularly distinguishes itself from the others by allowing all aggregated figures to be obtained through applying transformations on the more granular data while no further aggregated data may be required to be reported (except for the data under the scope of feedback loops or anchor values to ensure compliance with some aggregate figures (see further Section 4.6.7).

232 The hypothesis taken by this Option, correlated to evidence from the cost of compliance study and additional challenges highlighted in this report (see further Sections 4.6.7 and 4.6.8) point to the fact that Option 3 might not be feasible to be implemented.

233 **Granularity**: the data requested will have just one (high) level of granularity, at the level of the input layer.

234 **Implementation:** In Option 3 the collection layer will be common for all three reporting frameworks, therefore a high level of coordination between the authorities that define the data requirements, will be needed so that the granular data may serve the needs of all users (as no additional aggregated data is envisioned to be collected, unlike in Option 2 where we have a common granular collection layer and in addition more aggregated collection layers for each of the reporting frameworks

235 **Transformations:** will have to be defined for all additional levels of aggregated data that users might need. Similarly, with Option 2, a series of transformations might be jointly defined for more than one framework while other transformations will be specific for a certain reporting framework, depending on the chosen implementation.

236 **Responsibility:** the authorities that define the data requirements will be jointly responsible for defining the collection layer, while banks will be responsible for reporting the collection layer. As previously explained, for resolution and prudential purposes, institutions are responsible for their data (otherwise reliable prudential and resolution measures and analyses are not possible). Therefore, even under Option 3, feedback loops with the authorities and anchor values are necessary, which indeed require institutions to maintain internal processes for data aggregations.

237 This option fully takes into account the costs and benefits of moving to a more granular collection. Most of the costs and **benefits** defined for Option 2 also apply for Option 3. Few differences between the two options are highlighted in the table below:

		Costs	Benefits
Option 3	**Competent authorities**	• Possibly more complex process of creating transformations and more complex transformations (as all aggregated figures will be derived) (**time, resources**). • More coordination needed for creating the common collection layer (**time, resources**).	• Less complex collection layer (**simplification, usability**).
	Institutions	• Possibly more complex process of creating transformations and more complex transformations (as all aggregated figures will be derived) (**time, resources**).	• One unique common (highly) granular collection layer (**simplification, usability, efficiency, streamlining**).

Figure 9: Overview of the 3 options of going more granular

Common Data Dictionary		
Option 1	**Option 2**	**Option 3**
Current aggregated level of data collected (granular and/or aggregated data)	*Increase the collection of granular data and reduce the collection of aggregated data*	*Collect just granular data*
Separate collection layers	Common granular collection layer + separate (more aggregated) collection layer	Common (highly) granular collection layer
Data duplications	Possible data duplications	No data duplications
No additional transformations	Transformations need to be defined	Transformations need to be defined
	Feedback loops and anchor values	Feedback loops and anchor values

Questions to consider

19) Which of the options regarding the granularity of the possible future integrated reporting do you think is feasible (given the challenges and constraints highlighted in the discussion paper and possibly others) and preferable for you? Which are the main challenges and possible solutions to consider. Please rank potential challenges in ascending order (i.e. starting with the most challenging item in your view).

20) In case of Option 2, please specify how the granular collection layer should be designed to your best advantage (and benefit of reporting more granularly)?

21) What are the main benefits and costs of implementing the option considered feasible and preferable by you from Question 19?

22) What possible aspects related to the design of the option (Question 19) would make the costs for this option higher than the benefits and therefore not worth implementing.

4.6.6   Transformations

238 **Transitioning** to a more granular data collection, transformations will have to be defined to obtain the aggregated figures needed for the different policy objectives. The topic of transformations focuses on the context of Option 2. While Option 1 does not require transformations, the elements highlighted regarding the feasibility of creating transformations may indicate the infeasibility of Option 3. As institutions remain responsible for the more aggregated figures of reporting regulation it is necessary to take into account their needs on designing the data transformation solution to ensure that from a cost perspective, they are better off than in the current situation.

239 **Responsibility for aggregated figures**: as highlighted in the options, banks are responsible for all data collected, granular and aggregated figures. However, additional considerations should be made when it comes to aggregated figures derived through transformations[47] from a more granular collection. Figures aggregated for statistical purposes could be thought to be fully within the responsibility of the relevant authorities. In supervision and resolution, it is required that banks retain the responsibility for aggregated figures (and across different levels of aggregation) from a legal perspective, e.g. if individual banks shall be held accountable for failures such as too low capital ratios. The exact data points obtained by the aggregation process, might in some cases be of utmost importance as prudential and resolution decisions and actions are taken based on them. Hence, collecting granular data and 'simply' aggregating it (as required by relevant authorities) will not be enough from a legal perspective for all the data collected. Institutions need to be held accountable by some kind of 'feedback loop' and 'anchor values', whereby the institutions confirm their responsibility with the (aggregated) data.

240 What the responsibility for aggregated figures ('important aggregated figures') could look like:

- **Feedback loops:** the bank runs transformations on its side and compares the aggregated data with that obtained on the regulator side in a back and forth process.

- **Anchor values**: authorities could define a set of 'anchor values' that should be reported alongside the granular collection, i.e. aggregates designed mainly for data quality checks.

  The universe of aggregated values for which banks will have to be responsible (data in the scope of the feedback loops / anchor values (the 'important aggregates') will have to be decided. The governance process of setting and checking feedback loops and anchor values has to be set up. In addition, from a governance point of view, one could distinguish between responsibility for those values, aggregated according to transformation rules, and those which are derived in another way (e.g. for ad hoc analyses).

---

[47] Here transformations should be thought of in a broad sense (both transformations defined together with the authorities) but also any other analysis – any data point may be used for analysis purposes and supervisory actions.

241 In addition, it needs to be legally clarified whether anchor values or feedback loops are sufficient to ensure responsibility of institutions for all their data reported or whether a mix of both should be used.

242 **Responsibility for defining and conducting transformations**: ultimate responsibility for the transformations could be attributed to i) the authorities; ii) jointly to authorities and reporting institutions; iii) reporting institutions; or iv) the transformations would not be binding but the result of a cooperative effort and no responsibility attribution would be required. However, depending on the way responsibility is decided for the transformations, the decisions on the possible implementation and design of the options to move to a more granular collection will be inferred. The governance of the transformation rules should be decided together with the governance for the process of creating and maintaining the transformations.

Questions to consider

23) If transformations are to be defined (as depicted in Option 2 or Option 3), who should be responsible for their definition (e.g. who takes responsibility for their accuracy) and their execution?
24) If transformations are defined under different scenarios with respect to responsibilities, what are the major implications for the possibility of defining a more granular collection layer from a cost and benefit perspective also considering some of the challenges depicted in the paper (technical and legal, e.g. institutions need to remain responsible for all the data).

243 **Joint effort of all the stakeholders involved** (e.g. CA, regulators, reporting institution) in defining transformations: although the ultimate responsibility for the transformations is yet to be decided, a joint effort might be needed from all stakeholders' sides for the definition of these transformations mainly due to information asymmetry. Different experts from different areas would have to contribute to the reconciliation of data definitions and development of transformation rules. Different legal frameworks might require different solutions with regard to industry participation (e.g. FINREP vs COREP). To avoid duplicated transformations and for data lineage, (e.g. a more aggregated concept derived from granular data might be shared between two different regulatory reports) some transformations could be common for all three reporting frameworks while some others could be only needed for specific reporting frameworks. Depending on the setup, the governance of these common transformations together with the stakeholders' involvement should be decided.

244 Characteristics of the transformations: in order to be able to define transformations that would be uniformly applied to all reporting institutions' data collected, these transformations should be designed having a series of characteristics in mind:

- congruence with their purpose – achieve the required figure in line with the regulation;

- clearly defined;

- feasible and sufficient to be implemented in the collected data.

245 **Feasibility of defining the transformations**: Although the feasibility of creating transformations can only be considered on a case by case basis there are general considerations that should be accounted for:

- Regulation that relies on **principle-based approaches being taken** (e.g. in the case of CRR provisions or in the case of accounting standards; IFRS-based FINREP, FINREP nGAAP are constrained by principle-based regulations). In general terms, this would mean that in certain cases regulators refrain from defining detailed, prescriptive rules and rely more on high-level, broadly stated rules or principles to set the standards by which regulated firms must compute the regulatory figures. Principle-based regulation would be an impediment to creating common transformation rules and define the regulatory reporting at a granular level. At the same time, since principle-based regulation potentially cannot be avoided completely, it needs to be ensured that supervisors are able, if needed, to understand and double check the private solutions followed in these cases.

- The accounting frameworks currently underlying the reporting requirements should be accounted for when designing transformations so as to avoid the risk of decoupling requirements for annual accounts of institutions and the concept definitions and derivations in the new integrated system.

- Some institutions alter their granular data available internally by 'manual adjustments'. An integrated and granular reporting system needs to be able to cope with such kinds of individual adjustments, otherwise, it will be hardly achievable to ensure a comparable data basis, which would be derived from the granular data for the purposes of individual institutions. Alternatively, institutions could be required to align their 'manual' adjustments with the new environment. However, that granular input data should not be 'adjusted' in a way that it might not be comparable between different banks, since in this case the input data would not be a consistent basis for the collection layer. In any case, ultimately, the institutions are responsible for meeting the data requirements raised by regulators.

- There is a certain limit to the ability to transform the regulations into standard formulas, setting a boundary on how granular data can be collected by regulators. Many of the concepts referred to in the regulations are the result **of internal models developed by banks** (e.g. probability of default). In some cases, the derived concepts can be attributed to a granular-level element while in other cases the allocation of the value would involve an allocation mechanism difficult to uniformly represent. In this respect, the possibility to move to granular data for aggregated reports will most likely be easier for statistical data and FINREP and it will be more complex for COREP and resolution data.

- The **reconciliation of the data definitions** among different reporting frameworks might be difficult to achieve from a technical and legal perspective. In particular, for COREP and resolution, data transformation rules and reconciliation of data definitions are less straightforward and in addition, involve interpretation (e.g. no uniform SME definition, no definition of 'guarantee' for CRM purposes). If different regulatory data with different purposes cannot be reconciled, then it has to be considered as different requested data, limiting the scope of writing transformations and moving to data that are more granular. Alternatively, if there are no constraints (e.g. legal ones) the need to create transformations might create the opportunity to standardise definitions.

- Transforming the regulation from CRR/CRD/BRRD into standard formulas, where feasible, is further challenging due to the **possibility of national derogations/enhancements** making a uniform application of the rules demanding, e.g. while the CRR is generally applicable to all CRR institutions within the EU, the CRD is transformed into national law and usually allows for some kind of discretion. Hence, a comprehensive 'coding' approach for the CRD (or 'Directives' in general) should in addition consider that all relevant national laws are accounted for in the coding framework.

- **Different valuation required for the same concept:** different reporting frameworks might be aligned to different views with regard to the same underlying business concept given the different purpose of the use of data across the statistical, prudential and resolution sectors. The granular reporting will therefore have to consider if accounting[48], prudential, statistical and resolution values (possibly all different) would have to (and could) be required for each instrument/loan/contract individually or if the specific concepts may be derived across the frameworks.

- Other aspects to be considered: The governance structure for the entire process of designing and maintaining transformations should be clearly defined based on individual and shared responsibilities as per the points discussed above. The infrastructure needed to create, maintain, visualise, run and test the transformations should be considered together with the language to be used in defining and designing them. Additionally, transformations should be considered as an integral part of the common data dictionary and possibly the central data collection point.

- Moving to a more granular data collection and creating transformations should be a carefully considered decision, balancing the costs and benefits for all the stakeholders. On the one side, detailed rules provide certainty, a clear standard of behaviour, are easier to apply consistently; however, there is also the risk that detailed transformations may lead to gaps, inconsistencies, rigidity and are prone to the need for constant

---

[48] In addition, different accounting systems may also require different kinds of valuation. This may also vary between solo and consolidated figures (nGAAP and IFRS valuations may both be needed).

adjustment to new situations. Reaping the benefits of such a decision is very much dependant on the ultimate design of the solution and its implementation.

4.6.7 Further challenges to consider

246 Exploring the possibility of obtaining the aggregated data required in a certain reporting framework from the more granular data reported in other reporting frameworks means that a variety of dimensions defined (explicitly or implicitly) across the reporting frameworks for the same business concept have to be identified, compared and possibly controlled in the aggregation process, (e.g. one has to make sure that the 'loan' concept defined in statistics is actually the same as the 'loan' concept that is defined in FINREP, as identified by a series of dimensions that are deemed relevant). Differences across dimensions considered relevant (e.g. frequency) including differences in data flow processes would have to be controlled (additional dimensions/details on the data is available to be reported at more granular levels) in order to be able to reconcile data and move to a more granular collection. Choosing among the three options above requires a deep analysis of their feasibility, not only as a balance between costs and benefits, but also the possibility to overcome strict boundaries imposed by legal acts and technical feasibility.

Legal considerations

247 The legal boundaries regarding the level of granularity in data collection from existing law (e.g. with regard to confidentiality and data privacy) should be closely investigated. It is important to be able to assess who is able to access which kind of collected data, at which level of granularity and for which kind of purpose. Further details are provided in Section 7 on Governance. In addition, decisions to move to a more granular data collection and the mechanisms of feedback loops and anchor values would require its amending or enhancing the current legal framework for reporting data. Such a legal framework is needed at a minimum to define the data ownership rights, institutions' responsibility, enforcement.

Individual reporting and scope of consolidation

248 Statistical, prudential and resolution reporting frameworks require data to be collected at different levels of consolidation (individual and/or consolidated). The consolidation under the different frameworks might differ in terms of scope (e.g. entities that report under prudential consolidation might be different from those that are for the scope of resolution consolidation). In addition, the individual reporting (for supervision/resolution) would not be the same as the observed agent defined for some statistical frameworks (e.g. in the AnaCredit regulation we have legal entity excluding foreign branches, and branches in other Member States). While one possibility would be to align all scopes of consolidation, this would not eliminate the difference between concepts used for individual and consolidated reporting.

249 While certain attributes related to a business concept do not change depending on the level of consolidation at which it is reported or the entity that reports it, some others do (e.g. due to the

fact that intragroup transactions are eliminated in the consolidated reporting, or to differences in the applicable accounting standards). Therefore, requesting information at a more granular level that could be used to build more aggregated reports would require additional information to pinpoint exactly the information needed in different contexts (e.g. such additional information would mean to request at the level of the individual instrument both the standalone and the consolidated value). The general feasibility of such an implementation should be considered for each attribute as well as the data availability and the legal framework that would allow for collecting and identifying this data.

Frequency

250 The frequency with which data is requested to be reported might not be aligned between the aggregated data (which might for example be requested at lower frequency) and the same data requested at more granular level (which potentially might be requested at a higher frequency). One of the reasons why such a case might arise is the different purposes for which the data was initially requested (e.g. to compute quarterly statistical aggregates versus monitoring liquidity which requires a more frequent surveillance). Therefore, integration of the data implies aligning the frequencies of the aggregated reports with the granular data needed by either:

- requesting the aggregated data with a lower frequency;

- requesting the granular data with a higher frequency;

- requesting only a subset of the granular data (in case this subset, specific attribute and entities from the whole granular dataset that are specifically need as input in the computation of the aggregated report can be identified) to be reported with a higher frequency .

251 One important aspect to consider is whether the more granular data is available or possible to be submitted in time for the aggregated figures to be computed. This aspect might be relevant also when considering the topic on timeliness.

Timeliness

252 Timeliness refers to the time at which the data is sent to authorities, irrespective of the frequency (e.g. quarterly data for prudential purposes should be sent almost one and a half months after the end of the quarter to which it refers). Different reports might have different deadlines for when the data needs to be received allowing also for varying periods of time for resubmissions.

253 In the process of obtaining aggregated reports that might need to be computed based on granular data, the timeliness aspect is important. The timeliness in the case of aggregated reports should account for:

- the timeliness for the underlining granular data, in accordance with the data users' needs;

- the time it takes to run the transformations;

- possible timeliness with respect to the reconciliation of the aggregated figures obtained by the authority when running the transformations and those provided by banks as well as the time it takes to run required feedback loops.

254 In addition, different aggregated reports are computed/needed at different points in time based on the underlying more granular data. Choices will have to be made with respect to whether the same data for the same period should be sent several times if there are several reports with different deadlines and whether all versions for the same data and period should be stored. In addition different versions of the same data are likely to be used to produce different reports which might lead to spurious differences in the aggregated figures.

255 For efficiency reasons the timeliness for reports might be aligned as the data may be computed all in one time. On the other hand computing all reports in the same period might burden the IT system.

Data Quality

256 As a general remark, since the granular data requested by the collection layer is assumed to be closer to the banks' internal systems, the data collected might be of a higher quality compared to the current reporting as the number of iterations/transformations/systems the granular data went through is lower. A precondition for this is that the collection layer is well defined, ensuring that the granular data is derived in a consistent and unambiguous manner.

257 If data will be collected at granular level, the current data quality process is likely to require changes (e.g. automation, different types of checks) The quality assurance process of the granular data might have an impact on the timeliness of aggregated data and more weak/loose validations rules might be considered (some granular data might be allowed with a higher discrepancy difference as the impact on aggregate level might be very low). In addition, for timeliness reasons, asking for granular data with a tighter schedule might have repercussions on its quality.

258 Regarding data quality, aggregated data checks could remain in place, since they serve as a second line of defence in case there are qualitative issues with the input data.

Proportionality considerations

259 Proportionality, which is currently an important parameter of reporting requirements and of underlying legal frameworks, needs to be redefined in the context of granular reporting.

260 The more granular future reporting requirements are defined, the more difficult it might become to consider proportionality in reporting as we understand it now, i.e. a reduced scope / number of data points in a template-driven reporting. Proportionality in such an environment might for example be reached by dropping reporting requirements for sub-sections of granular data points for banks. This would come with the cost of losing information about a sub-section with regard to the affected institution, while at the same time the relief for the institution in question could be easily reached and would presumably be effective. On the other hand, in a multidimensional matrix, proportionality could be achieved by dropping certain attributes for the less relevant reporting agents. The decisions on the future common data dictionary should consider for a possibly more granular data collection. Therefore, proportionality aspects should be considered in the design of the dictionary.

261 An important aspect to consider is that it might potentially not be feasible to dispense on (certain) granular data, since such granular data is required to develop/calculate central regulatory figures.

262 Furthermore, when discussing the need to implement proportionality, not only the relative easements for smaller and non-complex institutions compared to bigger institutions under a potentially new granular framework should be considered, but also the cost / proportionality measures established in such a framework relative to the current framework. A major challenge in this regard will be to allow for proportionality which does not fall short of at least the current set of proportionality measures for smaller institutions (in terms of reporting costs, complexity to implement the reporting, data volume to be reported, etc.). A more granular framework comes with the risk of requiring more data, which at the same time might be more burdensome and difficult to compile also for small institutions. Hence, a granular framework needs to account for such new aspects, which need to be considered when developing a new approach of proportionally. Most importantly, a granular framework should not contradict the current work and the ultimate results of the Cost of Compliance study, which will very likely lead to additional easements for smaller institutions. A granular reporting should not eliminate the efforts for a targeted more proportionate reporting.

Questions to consider

25) How should the transformations be in terms of formalisation and readiness for digital processes?
26) How could some of the challenges highlighted for defining transformations be overcome?
27) What kind of data should be part of the feedback loops?

4.6.8   Data integration in terms of granularity – initial insights

263 Going forward there might be two possible directions to achieve data integration in terms of granularity:

▪ Reconciliation between the current regulations (the currently more granular data collection could be used to derive the more aggregated data collections). Building an

integrated reporting system in this sense would require gradually developing the statistical regulations hand in hand with prudential and resolution needs and vice versa. The integration and reconciliation of data definitions across the legal frameworks of each area of regulation (statistics, supervision, resolution), particularly requires considering the collaboration of the different legislators. The lowest possible level of granularity that would allow integration will be determined by the level of reconciliation of data definitions between the different legal frameworks.

- An alternative to reconciliation would be the creation of a new granular data collection, designed following a methodology that would allow the computation of statistical data, prudential and resolution data, avoiding the need to align the regulations of the currently collected data. Similar concepts in the currently collected data, even though different in methodology, might show that there would be some common foundation for building such a common granular layer.

264 In both cases it would be necessary to closely analyse the commonalities and differences in the data and the purpose for which it is collected under each regulation as well as the current process for collecting it to the extent possible. This stocktake investigation should cover all data requirements including optionality, discretions and enhancements at national level as well as any known ad hoc data requests in statistics, prudential and resolution frameworks.

265 The European harmonised prudential and resolution frameworks already share the same dictionary and concepts and have been already integrated. In statistics, the IReF project has performed a similar analysis as the one requested above for the different reporting frameworks under its remit and has managed to form an integrated view of the data collected for statistical purposes that is now under discussion. The outcome of the consultations that are now ongoing in the IReF project will be relevant for the further decisions on integration with the statistical and resolution data.

266 The idea of integrating across regulatory frameworks has been explored already to some extent in the past. The Joint Expert Group on Reconciliation (JEGR) of credit institutions' statistical and prudential reporting requirements was set up by the ESCB's Statistics and Financial Stability Committees and the EBA in June 2008. Its mandate has been renewed twice (in 2010 and 2012). The aims of the JEGR were to identify and promote common elements in the statistical and prudential reporting frameworks relating to credit institutions (e.g. definitions, concepts, valuation rules, reporting templates) and, where possible, to reconcile them. The JEGR activity was finalised in 2014 with the publication of the bridging manual for their respective reporting frameworks also including a common modelling of the reporting frameworks using the EBA DPM model.

267 The experience gained under the JEGR (even if limited to the aim of semantic and syntactic integration) and further ongoing work aiming to reconcile data (e.g. working groups at euro area and national levels looking into reconciling AnaCredit and FINREP solo) might be representative in showing that under the current situation, integration, from a reconciliation perspective, is

limited in scope. Identified similarities however will indicate to a possible starting point where a common granular data collection might be possible to be designed. Going further, the decisions on integration with respect to granularity can be considered gradually in the integration process, moving from Option 1 to Option 3, as further analysis and experience will show the extent to which this will be possible from a cost-effective and legal perspective.

Questions to consider

28) What other areas should the feasibility study investigate in terms of granularity and transformation rules?

5. Central data collection point

5.1 Considerations

5.1.1 Purpose and approach

268 This chapter focuses on the configuration of the reporting network – whose nodes are composed of Institutions, national aAuthorities and European Authorities – and aims to identify possible scenarios and to analyse them against the requirements of Article 430c.

269 The identification of alternative integration architectures is inspired by the IT network topologies most commonly used in integrating enterprise systems, and transposed to the larger scale of interconnected organisations.

270 The discussion of each scenario seeks to draw conclusions about its technical feasibility, and its relative advantages and disadvantages, especially in relation to compliance with the requirements of Article 430c, and its ability to support a CDCP.

5.1.2 Interpretation of the mandate

271 Article 430c of the CRR2 reads as follows:

> *'Following Article 430c of the CRR2, the EBA shall prepare a report on feasibility regarding the development of a **consistent and integrated system for collecting statistical data, resolution data and prudential data** […] including as a minimum:*
>
> a) *[…]*
>
> b) *the establishment of a **standard data dictionary** of the data to be collected [by the competent Authorities], in order to increase the **convergence of reporting requirements** as regards regular reporting obligations, and to **avoid unnecessary queries**;*
>
> c) *[…]*
>
> d) *the feasibility and possible design of a **central data collection point** for the integrated reporting system, including requirements to ensure strict confidentiality of the data collected, strong authentication and management of access rights to the system and cybersecurity, which:*
>
> > i. *contains a **central data register** with all **statistical data, resolution data and prudential data** in the necessary granularity and frequency for the particular institution and is updated at necessary intervals;*
> >
> > ii. *serves as a **point of contact** for the competent Authorities, where they **receive, process, and pool all data queries**, where queries can be matched*

*with existing collected reported data and which allows the competent Authorities **quick access** to the requested information;*

    *iii.   provides additional support to the competent Authorities for the transmission of **data queries to the institutions** and enters the requested data into the **central data register;***

    *iv.   holds a **coordinating role for the exchange of information** and data between competent Authorities; and*

    *v.   takes into account the proceedings and **processes of competent Authorities** and transfers them into a **standardised system.'***

272 The EBA interpretation of the central data collection point described in the mandate is as follows:

- *Design of a **central data collection point.*** The system should include (but not be limited to) some form of central data storage. From the perspective of the reporting institutions, the system should offer consistent interfaces for data collection across all jurisdictions and all types of reports, with the same protocols and formats for data exchange between the institutions and the Authorities. In addition, the central data collection point shall require the definition of common roles and access control rules, quality assurance and quality controls.

- *Contains a **central data register.*** The register can be understood as a central data catalogue that inventories and organises all data collected via the central data collection point and stored in its central data hub; it should use metadata to help its users to discover, understand, trust, and manage data, for governance or exploration purposes. The data catalogue should provide a comprehensive view of data available in the integrated system, contextual business information, the identification of the data custodians, and search functions that enable users to find data easily.

- *Serves as a **point of contact for the competent Authorities.*** The system should facilitate data sharing between the Authorities. The existence of a central data hub and a central data catalogue, together with clear data governance rules and data access policies, should enable the design of efficient mechanisms to provide authorised users with quick access to the requested data, with little or no need for intermediation.

- *Provides **additional support to the competent Authorities for the transmission of data queries.*** The system should help Authorities define ad hoc data requests in a manner consistent with regular data collections. These data requests would then be routed to the target reporting entities, and registered in the central data catalogue.

- *Holds a **coordinating role for the exchange of information** and data between competent Authorities.* The system should support a data governance process where

each Authority is responsible for managing the quality and accessibility of data in its custody. The system should support the management of data access rules and rights and maintain auditable data access logs.

- *Considers the **proceedings and processes of competent Authorities** and transfers them into a **standardised system***. The integrated reporting system's design must be consistent with the general good practices of the competent Authorities, so that all of them can be provided with, and benefit from, all the services required from a full-fledged reporting system.

5.1.3   Current regulatory reporting approach

273 The current setup of European regulatory reporting is based on the sequential approach, which is schematically represented in the following figure:

Figure 10: Current regulatory reporting approach
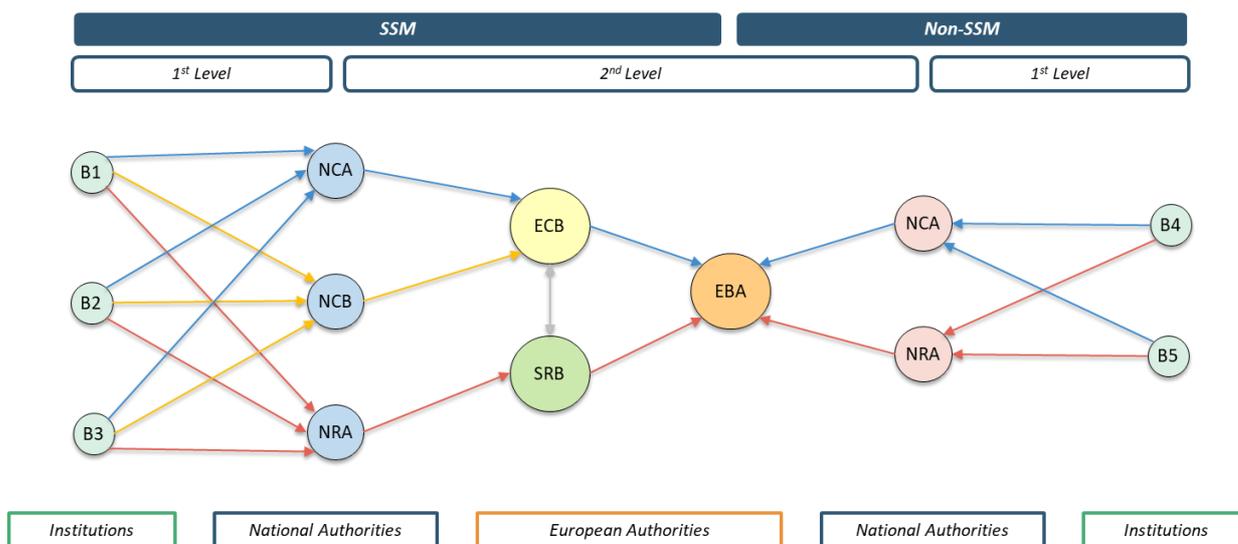
274 In this arrangement with a hierarchy of levels, data flows must pass through several nodes without transformation, but data are redundantly stored, and repeatedly monitored and validated at each node[49].

5.1.4   System integration topologies

---

[49] There are exceptions at first-level reporting, where countries have implemented a centralised collection of statistical, prudential and resolution data at national level.

275 Systems taking part in a network can be integrated according to different topologies, or arrangements in which they are connected to each other. The most popular integration topologies in the IT world are Point-to-Point, Service Bus and Hub-and-Spoke.

Figure 11: Point-to-Point topology

Point-to-Point
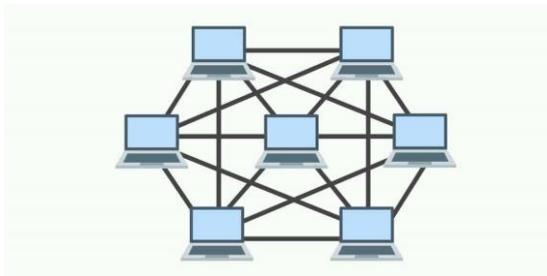
276 In this architecture, any two nodes are connected directly without a mediator. It is the simplest and easiest of all three topologies to implement, and the one with a higher maintenance cost.

277 Point-to-point topologies are reliable, with no single point of failure.

278 The main disadvantages are the exponential growth of complexity as new systems are added and problems related to data availability.

Figure 12: Service-Bus topology

Enterprise Service Bus
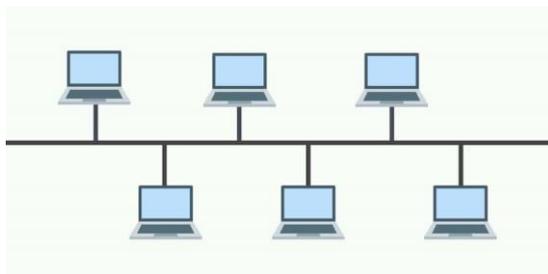
279 This integration approach relies on a specialised principal component that keeps track of what each system can do and stores.

280 This model enables the communication between the different nodes of the integrated system by standardised messages or by calling specialised services. The enterprise service bus (ESB) can be used to translate and map existing interfaces, enabling the transparent replacement of a legacy system with another.

281 It decouples systems from each other, allowing them to communicate without dependency on or knowledge of other systems on the bus. The ESB concept was born out of the need to move away from point-to-point integration, which becomes brittle and hard to manage over time.

282 A key issue of this model is the overreliance on a principal component that can make the entire system unavailable (single point of failure). Additionally, high network traffic decreases network performance, and transmission speeds decrease with the number of network nodes.

Figure 13: Hub-and-Spoke topology

Hub-and-Spoke
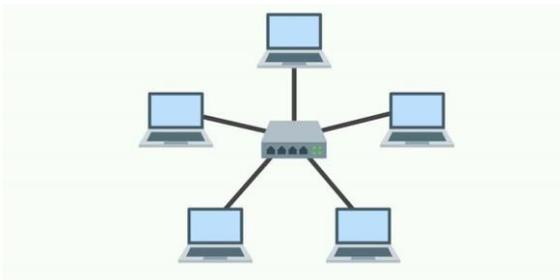
283 In this architecture, every node in the network is connected to one central hub, and indirectly connected to every other node.

284 The relationship between these elements is that the central network hub is a server, and other devices are treated as clients. The central hub is responsible for managing data transmissions across the whole network and acts as a repeater.

285 Hub-and-spoke topologies are most commonly used because the entire network can be managed from the central hub, and if a peripheral node goes down the network will remain up. The simplicity of the overall network design makes it easier for administrators to run troubleshooting when dealing with network performance faults.

286 The critical issue with this architecture is that if the central hub fails, then the entire network goes down. The performance of the network is also tied to the central node's configurations and performance.

5.2   Integration architecture scenarios

287 This section discusses six alternative scenarios as potential architectures of an Integrated Reporting System, connecting national authorities and reporting entities. Each scenario explores a particular topology and how it affects first and second-level reporting.

288 Reporting Entities are represented as small green circles, national competent authorities as blue circles, and national resolution authorities as pink circles. The three larger circles are identified as European authorities. To simplify graphical representation, only a small number of reporting entities and national authorities are represented, and each reporting entity is connected to a single national authority (i.e. merging NSAs, NRAs and NCBs in a single node, which is most often not the case).

289 The first scenario corresponds to the current sequential reporting topology, where reports flow first from reporting entities to national authorities (first-level reporting), and from those to the European authorities (second-level reporting); additionally, European authorities exchange data among themselves.

290 The next three scenarios are inspired on point-to-point, enterprise service bus and hub-and-spoke topologies, respectively, applied at second-level reporting. The assumption in either scenario is that national authorities would continue to use their local reporting systems, changing connectivity for second-level reporting, without necessarily altering existing connections with reporting entities at first-level reporting. Therefore, from the reporting entities' point of view, there would be no significant differences from the current scenario.

291 The last two scenarios refer to the alternative of replacing the diverse national reporting systems with a new single European reporting system, either in the form of a fully centralised system or as a distributed system of identical nodes. In one case, the reporting entities would have only one single data collection point, regardless of the jurisdiction or reporting framework. In the other case, despite the system's physical distribution with different nodes per national authorities, the connectivity protocols would always be the same, and therefore all data collection points would appear virtually identical to the reporting entities.

5.2.6 Sequential integration

Figure 14: Sequential integration
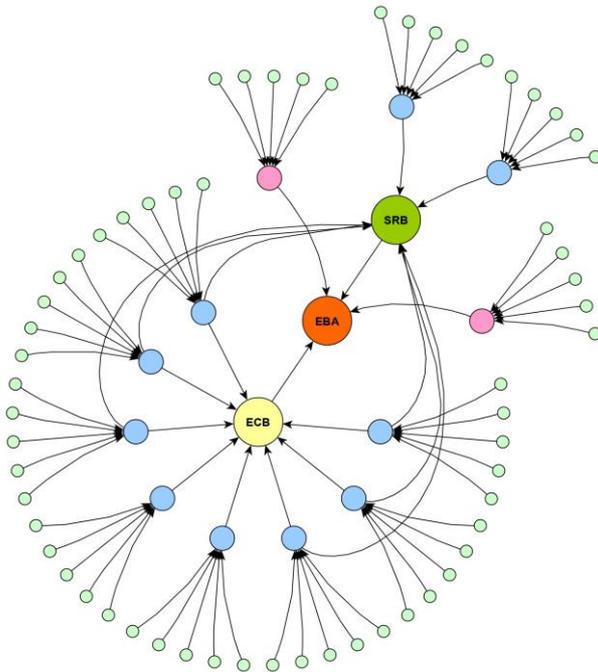
292 The image represents the current configuration of European reporting of supervision, resolution and statistical data, with unidirectional data flows from banks to national authorities, and to European authorities.

293 Processes, systems and connectivity protocols are specific for each Authority.

294 Each European authority manages an independent data hub whose data partially overlaps with the other two.

295 From the institutions' perspective, there is no central data collection point, as each national authority can define different data exchange formats, and uses various data exchange channels.

296 From the authorities' perspective, there is no central data access point, and the exchange of information between national authorities is limited.

297 Each national authority is currently running and maintaining its local reporting system. As a result, there are many different solutions for the same or very similar reporting requirements, and much redundancy of data processing and storage, aggravated by the sequential second-level reporting.

298 This architecture does not promote any global standards, and it is up to each authority to define the protocols for receiving data, and follow the other authorities' protocols for sending data.

299 Nevertheless, semantic and syntactical integration has been achieved for supervision and resolution reporting: both frameworks share a common data dictionary (EBA DPM) and the same data exchange format for second-level reporting (XBRL), which in most cases has also been voluntarily adopted by national authorities for first-level reporting.

		Costs/disadvantages	Benefits/advantages

Sequential	**Reporting Institutions**	• No central data collection point[50]
		• No harmonisation of reporting requirement specifications:
		  - Different data dictionaries by country and/or for prudential, resolution, and statistical data
		  - Different data exchange protocols, standards or formats of CAs in the same jurisdiction or in different jurisdictions
			• No change to existing solutions[51]

	National Authorities	• Each Authority develops, maintains, and runs a different reporting solution for what is essentially the same problem • No access to a central data register with all statistical data, resolution data and prudential data	• Autonomy at defining first-level reporting specifications[52] • Independent setup of national non-harmonised and ad hoc reporting • Possibility of reusing the same solution for other European reporting frameworks (e.g. Solvency II) • Preserving the status quo; no change to existing solutions
	European Authorities	• (Same as national authorities) • The complicated process of reconciling and maintaining the alignment of rules, metadata and master data in different systems	• Simple governance of own systems

---

[50] Central data collection points exist, as an exception, at national level, but they do not address the problem of cross-border banks reporting to both home and host supervisors.

[51] Changing or preserving the status quo can have different cost/benefit balances, depending on the situation of each particular institution.

[52] Most Authorities follow the Regulator's specifications while others have implemented their own specifications.

		Costs/disadvantages	Benefits/advantages
	Global Viewpoint	• Reported data is repeatedly pushed over the network through a sequence of nodes • Redundant data processing and data storage being executed at different nodes • Limited data sharing and exploitation; the value of collected data remains largely locked • Lack of global standardisation of data, processes, and communication protocols • Multiple solutions for the same problem; economies of scale are not explored	• Currently working solution

5.2.7 Point-to-point integration

Figure 15: Point-to-point integration
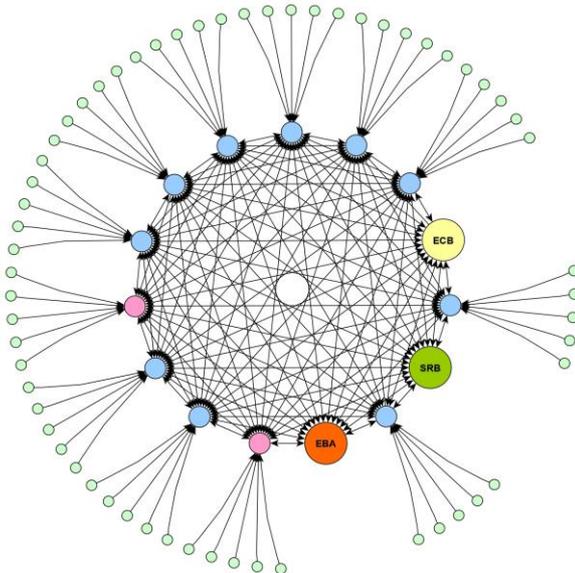
300 In the point-to-point network, each pair of connected systems would establish a bidirectional connection between the two nodes.

301 It is the most basic form of integration of legacy systems, with connections between each pair of systems, and no central hub.

302 The high degree of flexibility of this architecture may be favourable for a very limited number of nodes, but it becomes quickly impractical as the number of interconnected systems grows, with an exponential rise of integration costs.

303 From the reporting entities' perspective, this scenario would look the same as the current one, i.e. no central data collection point, as entities would only connect directly with their respective national authorities.

304 From the authorities' perspective, there would be no central data access point, but many possible shareable data sources.

305 This architecture is presented as a purely theoretical possibility, but not viable in practice due to the number of authorities involved in the reporting system.

5.2.8 Service-bus integration

Figure 16: Service-bus integration
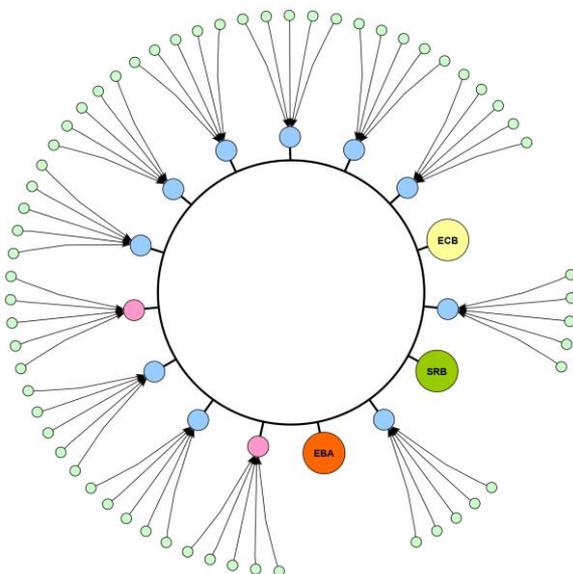
306 The service-bus architecture is one of the main topologies used for enterprise systems integration, where a specialised middleware system is dedicated to set up communication between other systems.

307 In this scenario, where all authorities are connected via the bus, each system that sends messages must follow the protocols used for exchanging messages in the network, and, similarly, each system that receives messages must be able to understand (syntactically, although not necessarily semantically) the message types.

308 Connections in this topology are reduced to only one adapter per system that connects directly to the bus; therefore the complexity is much lower when compared to the point-to-point topology (however, the service bus still has a high configuration and maintenance complexity, and limitations on network traffic performance, being therefore not suitable for large networks).

309 From the reporting entities' perspective, this scenario would still look the same as the current one, i.e. no central data collection point, as entities would only connect directly with their respective national authorities.

310 From the authorities' perspective, and although there is no central data hub, virtualised central access to distributed data stores could theoretically be implemented. In practice, however, there is a barrier to efficient data access across multiple distributed databases with mismatching data models and the need for continuous translation.

311 For the European authorities, who need to access and aggregate data from all countries, this architecture would fail to support their basic requirements, and is therefore not considered a viable alternative to the current scenario.

5.2.9   Hub-and-spoke integration

Figure 17: Hub-and-spoke integration
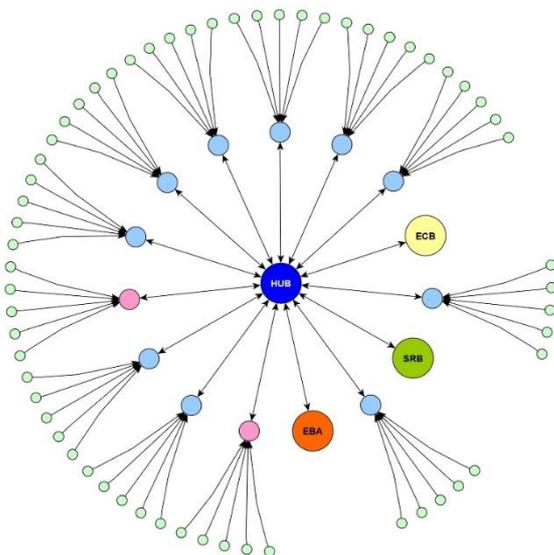
312 The hub-and-spoke architecture is another main topology for system integration. The communication is not done directly between pairs of systems, but between each system and the central hub.

313 In this scenario, the authorities connect directly to the central hub, which would not be just middleware with message broker functionality (with practical problems like the service bus), but

a complete reporting system with centralised storage of all reported data, and a common interface between the central hub and satellite systems.

314 From the reporting entities' perspective, this scenario would still look the same as the current one, i.e. no central data collection point, as entities would only connect directly with their respective national authorities.

315 From the authorities' perspective, there would be a central access point to all reported data, a hub hosting a complete data register, and a single data storage model. It would allow the implementation of efficient services to match queries and request access to existing data, or directly access the data hub when authorisation has been granted.

316 From the point of view of the efficiency of the global reporting system, there may still be total redundancy of data processing and storage, but some authorities could also choose to discard their local systems and rely only on the services of a central system that offers standardised support for all core regulatory reporting processes.

|		Costs/disadvantages	Benefits/advantages
Hub-and-spoke	**Reporting Institutions**	• No central data collection point • No harmonisation of reporting requirement specifications: - Different data dictionaries by country and/or for prudential, resolution and statistical data - Different data exchange protocols, standards or formats of CAs in the same jurisdiction or in different jurisdictions	• No changes to current solutions[53]

---

[53] Changing or preserving the status quo can have different cost/benefit balances, depending on the situation of each particular institution.

		Costs/disadvantages	Benefits/advantages
	National Authorities	• Each authority develops, maintains, and runs a different reporting solution for what is essentially the same problem	• Autonomy at defining first-level reporting for the harmonised reporting • Independent setup of national non-harmonised and ad hoc reporting • Possibility of reusing the same solution for other European reporting frameworks (e.g. Solvency II) • Minor changes only to existing solutions[54] • Access to a central data register with all statistical data, resolution data and prudential data (depending on access rights)
	European Authorities	• The transition from the current partially overlapping data hubs (ECB, EBA, SRB) to the single central data hub • Joint governance of a central hub is potentially a more complex task	• Single management of system's rules, metadata and master data (no reconciliation effort) • No duplicated efforts of monitoring the timeliness, completeness and quality of reported data • Access to a central register of all data (depending on access rights)
	Global Viewpoint	• Lack of global standardisation • Economies of scale are not fully explored (increase in the services covered with less input costs)	• Reduction of movement of data over the network • Reduction of redundant data processing and data storage on different nodes • The full potential of data sharing and exploitation by different stakeholders, thus unlocking the value of collected data

5.2.10 Centralised system

---

[54] Changes limited to complying with new specifications for second-level reporting.

Figure 18: Centralised system
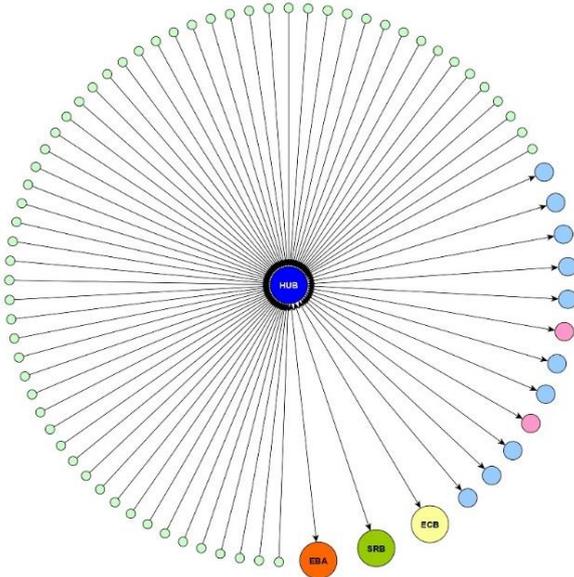
317 In this architecture, there is one single reporting system (with integrated design) to which all reporting entities and authorities are directly connected.

318 It can be seen as an evolution of the hub-and-spoke architecture by removing the authorities' local systems that intermediate the reporting entities and the central hub. The central system becomes then the single point of contact for both data collection and sharing.

319 From the reporting entities' perspective, there is a unique central data collection point, where the same connectivity protocols apply to all entities regardless of jurisdiction, and a single data exchange format may potentially be used for all types of reports.

320 From the authorities' perspective, the central access point to data would have the same type of availability and efficiency as in the hub-and-spoke architecture, and data processing would migrate from local implementations to central standard solutions (even though data ownership and flow control may still be decentralised).

321 The drastic reduction of data flows, when compared to the hub-and-spoke architecture, and eliminating redundant data processing and storage could mean large gains of efficiency and economy of resources.

322 Possible downsides of a single standardised system are:

- The lesser flexibility to support specific national requirements might prevent some authorities from switching entirely to the centralised system.
- A more complex change management process, due to the number of primary stakeholders.

323 The mitigation of the operational risk of single-point-of-failure requires the system to be fault-tolerant, i.e. able to maintain an acceptable level of functionality in the event of a component failure, instead of a complete breakdown. Fault-tolerant design techniques, like redundancy, will increase the solution's cost compared to a less robust alternative.

		Costs/disadvantages	Benefits/advantages
Centralised	**Reporting Institutions**	• Need for changes to current solutions[55]	• A central data collection point for all types of reports and across all jurisdictions • Fully harmonised reporting requirement specifications: - Single European data dictionary for prudential, resolution, and statistical data - Same data exchange protocols, standards, and formats in all jurisdictions
	National Authorities	• The transition from current local systems to the new centralised system • Setup of national non-harmonised and ad hoc reports on the centralised system is subject to the observance of strict syntactic rules[56] • No possibility of reusing the centralised solution for other European reporting frameworks (e.g. Solvency II)[57]	• Reliance on the centralised system to support all data collections; no need to maintain and run a full-fledged reporting solution • Maintenance of decentralised roles and responsibilities of reporting follow-up and data quality management • Access to a central data register with all statistical data, resolution data and prudential data (depending on access rights)
	European Authorities	• The transition from the current partially overlapping data hubs (ECB, EBA, SRB) to the single central data hub • More complex governance and operation of a large central data collection point • More demanding data management processes, including data access and data security	• Centralised management of system's rules, metadata, and master data • Possible reliance on CAs for most of the data quality management processes (which don't need to be repeated on the same platform) • Access to a central register of all data (depending on access rights)

---

[55] Changing or preserving the status quo can have different cost/benefit balances, depending on the situation of each particular institution.

[56] The capability of a metadata-driven centralised system to support national reporting is mainly dependent on the features of the data dictionary, which should be designed and managed to allow for national extensions.

[57] This restriction (affecting only some authorities) is due to the scope of Article 430c, but there is no functional or technical reason why the centralised system could not be designed for a broader scope.

	Costs/disadvantages	Benefits/advantages
Global Viewpoint	• Potentially increased operational risk (single point of failure) • More expensive fault-tolerance design	• Global standardisation • Economies of scale fully explored[58] • Data is collected only once, and not repeatedly at first and second levels; No redundant data processing on different nodes • Full potential of data sharing and exploitation by different stakeholders, thus unlocking the value of collected data

5.2.11 Distributed system

Figure 19: Distributed system
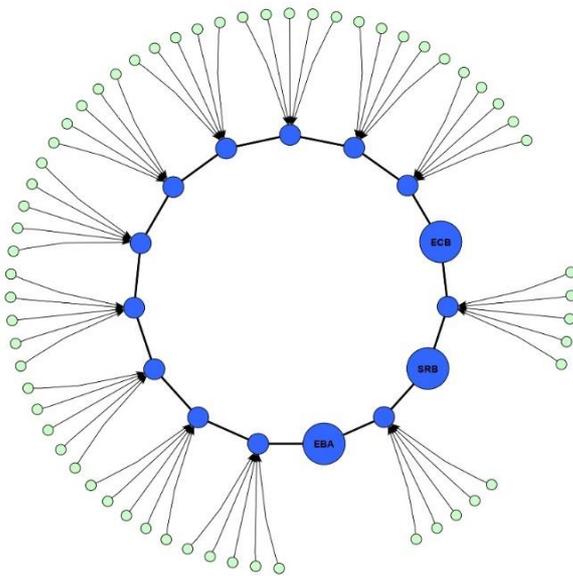
324 This architecture is a variant of the centralised system where the central hub at the centre of the star is replaced by a network of multiple identical nodes deployed at each national authority. Data processing and storage are physically distributed by country, but the processes and the data model are always the same.

325 Although the topology looks similar to the service bus, there are main differences: all legacy national systems have been replaced by an instance of the new system, and there is no need for adaptors or message translation, since there is a single data model replicated in all nodes.

---

[58] To the extent where the authorities can let go of their local data collection systems, which might not always be possible.

326 From the reporting entities' perspective, there may be a virtualised central data collection point, providing the same experience of a centralised system.

327 More flexibility could be allowed for supporting national data collections, possibly at the expense of some additional complexity of change and configuration management.

328 From the authorities' perspective, there would be a central access point to data partitioned across a distributed database. The distribution of data by country would favour in general the data access performance for the national authorities however the benefits to the institutions acting and reporting cross-border are limited.

329 This distributed architecture shares other advantages and disadvantages of the pure centralised system, but it appears to be less cost-efficient and complex to manage.

		Costs/disadvantages	Benefits/advantages
Centralised	**Reporting Institutions**	• Need for changes from current solutions to a single harmonised one[59]	• A virtualised central data collection point for all types of reports and across all jurisdictions • Fully harmonised reporting requirement specifications: - Single European data dictionary for prudential, resolution and statistical data - Same data exchange protocols, standards and formats in all jurisdictions
	National Authorities	• The transition from current local systems to a node of the new decentralised system • Setup of national non-harmonised and ad hoc reports on the centralised system is subject to the observance of harmonised syntactic rules[60]	• Reliance on the decentralised system to support all data collections; no need to maintain and run a full-fledged reporting solution • Reinforced roles and responsibilities of reporting follow-up and data quality management • Optimal access to a local data register with all national data • Access to a distributed data register with all statistical data, resolution data and prudential data (depending on access rights)

---

[59] Changing or preserving the status quo can have different cost/benefit balances, depending on the situation of each particular institution.

[60] The capability of a metadata-driven centralised system to support national reporting is mainly dependent on the features of the data dictionary, which should be designed and managed to allow for national extensions.

		Costs/disadvantages	Benefits/advantages
	European Authorities	• The transition from the current systems to a new system	• Centralised management of system's rules, and metadata • Reliance on NCAs for most data quality management processes • Access to a distributed data register with all statistical data, resolution data and prudential data (depending on access rights)
	Global Viewpoint	• Complex governance and operation of a large decentralised system • Complex data management processes, including data access and data security • More expensive distributed design	• Global standardisation • Data is collected only once and not repeatedly at first and second levels; No redundant data processing on different nodes • Full potential of data sharing and exploitation by different stakeholders, thus unlocking the value of collected data

5.3   Preliminary conclusions

330 The implementation of a reporting solution that would offer to the institutions a CDCP for all prudential, resolution and statistical data would imply a more or less adaptationof the national authorities' systems or their replacement by a new Integrated Reporting system.

331 At minimum, this CDCP would require full syntactic harmonisation of first-level reporting, including a common data dictionary, a common data model and a common data exchange formats and protocols, across all jurisdictions. For higher efficiency, the data quality management (DQM) and feedback should also follow common standards. With regard to the harmonisation of national requirements and of the ad hoc data collections, specific considerations are reported in Section 8.3.

332 The CDCP could be implemented either with a physically centralised system, or with a system virtualized over the network. Other possible hybrid architectures could also be implemented, allowing for more flexibility in a transition process or to address specific situations.

333 For instance, moving from the hub-and-spoke architecture to the centralised system can be accomplished in a staged approach, evolving the central hub and allowing each national authority to change at different times.

334 Another mixed scenario could envisage a more flexible architecture (e.g. a semi-decentralised system), where some authorities would depend exclusively on the central hub, while others would replace their legacy system with a local instance of the new system.

335 Any of the identified viable alternatives to the current sequential approach would support the required features of a central data register, providing secure access to authorities and other stakeholders to the prudential, resolution or statistical data.

Questions to consider

29) What other areas should the feasibility study investigate in terms of granularity and transformation rules?

30) Is your institution reporting to different authorities in your home country?

31) Is your institution reporting to other authorities in host countries? What problems arise from reporting to different authorities?

32) Are you using one or more data dictionaries for reporting? How?

33) Are you using the same or different formats for prudential/resolution reporting and for statistical reporting?

34) How important would it be for your institution to have access to a CDCP for all prudential, resolution and statistical reports? Why?

35) What would you think could be the challenges, costs and benefits of changing to a CDCP?

36) What solutions could the EBA investigate that would reduce costs?

6. Overview of costs and benefits

Level of integration / issue and applicability of costs and benefits			Benefits	Consistency	Simplification	Streamlining	Clarity	Efficiency	Usability	Reusability	Interoperability	Quality improvements	Fin. cost red.	Fewer resources	Financial	Time	Resources	Costs
LEVEL OF INTEGRATION (Section 3.2)																		
DATA DEFINITION	**Semantic**	CA	Unique naming and definition, facilitate understanding and interconnection of different reporting req.'s, compilation of tailored datasets, lower maintenance costs	x	x		x	x	x				x	x	x	x	x	Complete harmonisation at the outset, identifying the relationships between concepts, continuing analysis of all existing reporting FWs
		Inst.	Increased data quality, internal data stewardship facilitated, data duplication and overlaps avoided	x			x	x				x		x	x	x	x	Sunk costs: institutions may have already developed integrated definitions
	Syntactic	CA	Standard structure (metamodel), easier combination of datasets, same standards for all reports		x		x	x	x	x	x		x	x	x	x	x	High one-off onboarding costs (amortised over time)
		Inst.	Processing a unique metamodel structure, queries aligned	x	x		x	x			x	x	x	x	x	x	x	Moderate initial costs during onboarding (could be amortised over time)
	Infrastructure	CA	Common technical solution, allows for interoperable reporting systems, better use of resources, improved efficiency (similar design patterns)	x	x	x	x		x	x		x		x	x	x	x	Initial cost of onboarding, agreeing on the technical architecture of the solutions
		Inst.	No need to report the same/similar data twice				x		x			x		x	x	x	x	Possible contributions to and increase in maintenance costs of the infrastructure; costs may vary depending on the architecture

Level of integration / issue and applicability of costs and benefits			Benefits	Consistency	Simplification	Streamlining	Clarity	Efficiency	Usability	Reusability	Interoperability	Quality improvements	Fin. cost red.	Fewer resources	Financial	Time	Resources	Costs
DATA COLLECTION	Semantic	CA	Use of a common set of business rules for different reports, data can be reused for different purposes (depending on the type of solution used)	x	x	x				x					x	x	x	Standard process needs to be set up (costs will depend on governance structure, not much change expected)
		Inst.	Facilitate institutions' understanding of all the reporting business requirements	x	x		x		x									None.
	Syntactic	CA	Same data formats received by authorities across all datasets (easier to read, check and reuse data, exchange data), lower maintenance costs	x	x	x	x	x	x	x			x	x	x	x	x	Initial increased one-off cost to convert existing reporting schemes
		Inst.	Improvement in reporting process and efficiency, direct cost reductions as reporting could become a market commodity				x	x					x	x	x	x	x	Initial one-off costs: adjustment of internal reporting systems
	Infrastructure	CA	Cost, benefits, feasibility and implementation will depend on the architecture selected and how it will be operated.															
		Inst.	Cost, benefits, feasibility and implementation will depend on the architecture selected and how it will be operated.															
DATA TRANSFORMATIO	Semantic	CA	Data already collected can be transformed and reused, reducing future data requests, common rules would improve data quality					x		x		x			x	x	x	Potentially high initial implementation costs (depending on the scope of transformation rules to be defined and how granular data is collected), ultimately lower maintenance costs
		Inst.	Facilitated interpretation of various regulations on data aggregation, data collection at more granular level would reduce duplication of reporting, limit institutions' need to construct/obtain transformation frameworks	x	x		x	x					x	x	x	x	x	If it leads to reporting more granular data, costs will depend on the inst (could increase or decrease). Impact and feasibility of collecting more granular data under discussion
	Synt.	CA	Possibility to share validation, calculation and transformation rules; allows for a solid validation machine for regulators and reduced operating costs (common definitions and standards)	x		x	x	x				x			x	x	x	Changes to transformation execution

Level of integration / issue and applicability of costs and benefits			Benefits	Consistency	Simplification	Streamlining	Clarity	Efficiency	Usability	Reusability	Interoperability	Quality improvements	Fin. cost red.	Fewer resources	Financial	Time	Resources	Costs
		Inst.	Common understanding of the reporting requirements and facilitated communication, facilitated use of technology	x		x	x	x	x					x	x	x	x	Changes to transformation execution,
	Infrastructur	CA	Harmonisation and coordination of processes, information sharing	x			x	x	x						x	x	x	IT and human resources for setting up/maintenance of infrastructure, setting up associated governance process
	Infrastructur	Inst.	Harmonisation and clearer communication of requirements (more benefits for groups operating cross-border)	x			x	x	x						x	x	x	IT and human resources for setting up/buying the infrastructure
DATA EXPLORATION		CA	Seamless data exploration and analysis, reduced cost of ownership of data analysis services, data consistency, limited expertise and capital requirements for data exploration, lower development and maintenance cost	x			x	x	x	x			x	x	x	x	x	High on-off costs for developing one system, new costs from new governance requirements, potential sunk costs (dismissal of existing investment)
DATA EXPLORATION		Inst.	Limited expertise and capital requirements for data exploration, lower operational costs, service standardisation (long-term)		x		x	x					x	x	x			None expected (subscription charges may apply).
DATA DICTIONARY (Section 4)																		
		CA	Comprehensive info on all reporting frameworks, improved comparability, uniform mechanisms for data dissemination, level playing field for various analyses, improved ability to share data across NCAs, enhanced possibilities to use and process data across different frameworks, ability to formulate needs better, learning curve effects (benefits increasing over time)	x	x		x	x	x	x					x	x	x	Additional responsibilities for NCAs, evaluation and decisions on feasibility of various aspects, semantic integration of frameworks from different regulators, design cost of a common syntactic dictionary, costs related to supporting the central data collection point, change of existing systems required to adapt, learning curve effect, coordinated communication

Level of integration / issue and applicability of costs and benefits		Benefits	Consistency	Simplification	Streamlining	Clarity	Efficiency	Usability	Reusability	Interoperability	Quality improvements	Fin. cost red.	Fewer resources	Financial	Time	Resources	Costs
	Inst.	Clearer and more structured reporting rules, common understanding amongst all staff involved in the reporting process within institutions, facilitates the preparation of reports, common understanding across inst's (level playing field), efficient application of requirements in digital solutions, integration and standardisation of different standards, interoperability, same approach can be applied across different reporting obligations, learning curve effects (benefits increasing over time)	x	x	x	x	x	x		x	x			x	x	x	Implementation of the new dictionary in the reporting process, learning curve effect

GRANULARITY (Section 5)

Level of integration / issue and applicability of costs and benefits		Benefits	Consistency	Simplification	Streamlining	Clarity	Efficiency	Usability	Reusability	Interoperability	Quality improvements	Fin. cost red.	Fewer resources	Financial	Time	Resources	Costs
Medium/long-term	CA	More flexibility (transformation and responding to new policy needs), more comparability and transparency in the transformation process, uniform implementation, collection only once	x		x	x	x				x			x	x	x	Duplication (calculations may have to be performed by regulators and institutions to verify), potentially more costly quality assurance process, coordination efforts needed, maintenance costs (evolving rules and requirements), possible risk transfer from inst.'s to authorities (detecting/acting upon breaches), legal risk
Medium/long-term	Inst.	More stable/reduced reporting requirements over time, harmonisation of national reports, especially for cross-border groups (national extension might be incorporated), level playing field, simplification of internal processes (reported data closer to data stored), reduced reporting duplication	x	x	x	x	x	x	x				x	x	x	x	Collection of data from various systems, calculation of aggregate data possibly still necessary, maintenance and enhancement of transformation rules
Option 1	CA	Easy to implement, no further changes to the scope of the data collected, facilitates implementation of Option 2				x		x									Possibly more complex collection layer (accommodating both granular and aggregate data)
Option 1	Inst.	No additional data points to be reported						x						x	x		Still some duplications possible
Option 2	CA	Likely fewer Q&As on data to be reported, better understanding of reported data, consistency and improved quality of reported data	x	x			x				x			x	x	x	Human and IT resources (setup, aggregation, etc.), depending on the governance chosen: sharing transformation responsibilities, highly coordinated action needed

Level of integration / issue and applicability of costs and benefits		Benefits	Consistency	Simplification	Streamlining	Clarity	Efficiency	Usability	Reusability	Interoperability	Quality improvements	Fin. cost red.	Fewer resources	Financial	Time	Resources	Costs
	Inst.	Improved quality of banks' data, level playing field, facilitated compliance (through clarity and transparency)	x	x		x		x			x			x	x	x	Increased reporting burden, human resources and IT costs (setup, maintenance, reconciliation of aggregate figures)
Option 3	CA	Less complex collection layer		x				x						x	x		Possibly more complex transformation processes, need for more coordination
	Inst.	One common, unique collection layer		x	x		x	x						x	x		Possibly more complex transformation processes

Questions to consider

37) Would the industry be prepared to bear the costs of integrated reporting?

38) Where (within the different elements discussed in the paper) do institutions and CAs see institutions' involvement and cost contribution the most valuable in the development of an integrated reporting framework?

39) On a best effort basis, please include any monetary cost estimate you may be able to provide (% of operational costs) related to the implementation of an integrated reporting system for your institution.

7. Push versus pull architecture/model for data collection

336 Authorities collect data from institutions by requiring them to deliver certain information. Authorities give very specific instructions regarding what data should be delivered, in what format and by when it should be delivered. Reporting entities gather the data from their internal systems, transform it to obtain the output-required figures and deliver it to the authorities at the communicated date according to the instructions. This setup of the data flow is known as a 'push model', as the institutions are the ones that push the data to the competent authority.

337 The alternative to a 'push mode'l is the 'pull model'. In such an approach, institutions would no longer be required to deliver (to push) the data to authorities at a pre-specified date. Instead, the authorities would be able to access the institution's data repositories and directly retrieve the data (pull the data) from their systems. Given the heterogeneity across institutions in the way they organise their internal systems, the 'pull method' would only be effective if a priori institutions organise their data repository following an agreed data model. Without the data model in place, authorities would not know what data to ask for and what is available. The data model should, for efficiency reasons, be the same for all institutions otherwise it would be hard for one authority to manage all (it would be unfeasible for the authority to deal with a different 'model' for each institution). Once a data standard is set in place, authorities could pull the data by using agreed interfaces like APIs or WebServices.

338 Both methods would require a certain level of standardisation to work efficiently and technical and legal limitations might make them very similar in terms of scope of the data collected Standardisation will also ensure harmonisation and alignment with underlying regulations and accounting principles. However, the two methods require very different setups in terms of governance and technology and therefore their suitability should be considered in relation to the type and purpose of the data collected.

Transformations and accountability

339 The pull and push model may serve the same kind of data, at different degrees of granularity. However it might be the case that larger volumes of data including more granular data might be better served through a pull approach.

340 In a push model, institutions are the ones that are fully in charge with the transformations they apply to the data and able to justify their choices, and are in a better position to explain what data they are sending to the authorities. Reporting institutions are also held accountable for the data they send. Authorities on the other hand have very little knowledge on the internal

processes run by an institution to arrive at the regulatory figures. In the push model, authorities consider that it is better to ask institutions to perform the calculations, since they know the data and on the other hand, it would be very costly for authorities to perform complex transformations and aggregations from very granular data, which may very well be dependent on (many) specific internal models.

341 In a pull model authorities might have easier access to more detailed data from the institution's data repository that could better serve their analytical needs and for checking institution's compliance. In this setup, institutions would be accountable for the data they position in the repository and in some cases for aggregated figures. Transformations applied by the authorities should be made transparent to the institution to justify their policy actions and align with the aggregated figures produced by banks where the case may arise. The type of transformations that the authorities could apply might be limited in scope to only those that could be easily explained and applied in a standardised way across the institutions. This further dictates the kind of data authorities need to have access to from institutions' internal repository and the degree to which the data sitting in the repository is already transformed (data that is a product of institutions' specific internal model (e.g. PD) can only be accessed by the authority after the internal process has been run).

Timeliness

342 The time at which the authority gets possession of the data needed is one of the key differences between the two models. In the push model, institutions push the data to the competent authority, in general, in the context of a well-established cycle (reporting obligations calendar) or at dates communicated in advance to the institutions (e.g. in the context of ad hoc requests). The actual transfer of information is performed in a very limited number of rounds, always at a date agreed beforehand. Thus, the competent authority needs to wait for the specific date to get possession of the requested data.

343 In the pull model, the reporting system should be available (online) all the time; authorities may access the institution's repository and query the data at any moment and in this way might have access to institutions's data in near real time. Other process setups might also be envisaged, as the pull system could access the data from the transactional systems or other systems that might be updated with a lower frequency.

Data integrity

344 In a push model, any changes in the reported figures can be easily identified by the authority by comparing two different submissions. The reporting institution can check the quality of the data and if it is fit for purpose only after it was 'pushed', which can be especially challenging in the case of large volumes of aggregated reports (e.g. data coming via reporting templates) needed to be calculated and resubmitted.

345 In the pull approach, since the requests are on-demand, the authorities have limited ways to check if a particular value was not changed post factum ('how can you prove that at a certain time the fact value was Y'). To have some assurance, the authority would need to ask for all the data at regular intervals (which would make the model lose some of its benefits) or use some other source of trust (e.g. a distributed ledger or a trusted notary). The institution could be required to generate a timestamp in the data accessible for pulling for allowing authorities to check changes in values.

346 In a pull model the responsibility of data quality shifts, to a certain extent, to the authorities (as they are deciding which data and version of data they use as well as on how the data has been extracted), whereas in the current push model the obligation for correct data lies entirely with the institutions.

347 In addition, to avoid the risk of inconsistencies in the supervisory assessments and decisions and ensure a level playing field (at institution level or between countries) the participants would need to cooperate and agree in terms of timing, disclosures and usage with respect to the data that is pulled.

Standardisation

348 In the push model, authorities do not interfere with the institution's way of organising data in their internal systems. However, institutions are in charge of mapping the data from their system to the standard format required by the authority when delivering the data.

349 In the case of the pull approach, the reporting entities would have to reorganise their internal systems in accordance with the standard data model required by the authority. This could lead to increased inefficiencies, as the standard models may not be the best fit for the institution's business model. In addition, standardisation at this level might reduce flexibility in responding to change (authorities requesting new data and institutions' need to accommodate system constraints when considering changes in their internal processes and available technologies).

Access – transfer channel

350 The push approach establishes a persistent transfer channel between authorities and institutions. The transfer conditions (set up by the competent authorities) are well known in advance (reporting templates, calendar and other reporting instructions). Institutions are the ones responsible for establishing the connection to the central system and for the transfer of data to respond to each competent authority's instructions while competent authorities are passive receivers of the data.

351 In a pull model, the institution would need to manage a list of entities that competent authorities make pull requests of its data and set up the transfer conditions while the competent authorities are the ones that bear the cost of connecting to the institutions' system. A common challenge with pulling is establishing a queuing or request prioritisation and management system (e.g. questions such as what type of requests should be processed first (prudential vs

statistic) and issued by whom (for e.g. EU vs national request) need to be addressed. In addition, technology limitations such as memory availability or processing time should be carefully considered as the pulling of data may generate a lot of traffic (e.g. in the case the competent authorities are using the same fact value for multiple operations, the data point could potentially be queried multiple times). In this setup, institutions are the passive entities, waiting for the competent authorities to initiate their requests and pull the data. From the institutions' perspective, reporting exceptions and derogations can be more conveniently managed, since the institutions will no longer be required to setup exemptions.

Purpose

352 In the push approach, the competent authority knows very well in advance what data they need to cover their mandate while the pull approach is more suitable when the competent authority is looking to answer 'why' and 'how' questions at very granular level. A degree of this granularity can be built into the push model, but the ability to go and find data on any level in a timelier manner, still from within the institutions' repository, provides an invaluable means to dig much deeper and broader into the data and obtain meaningful insights from which to take action. However, it might be that these kind of granular questions do not have regulatory implications or relevance and would be more the concern of institutions' management. In this case, the competent authority runs the risk of having access to information and being blamed for not acting upon it, shifting the responsibility from institutions for managing their own business.

Group level data and institution's individual data

353 Similarly with the push approach, where data can be delivered by certain (e.g. parent) institutions (also on behalf of other institutions), in the pull approach, authorities need to know what data they should retrieve, from which sources, for what kind of needs. Supervision and resolution are focused on the supervision/resolution of a consolidated group/resolution group, rather than on solo institutions; therefore, the pull rules, as well as any microdata collection, should not affect the need for a consolidated view of the data.

Mix model

354 A possible alternative to using either a push or a pull model for data reporting would be the implementation of a mix model where some data would be reported by institutions through a push mechanism while other data would be pulled. This kind of mix approach would be better suited if we were to consider that from a cost benefit perspective, certain types of data might be better reported in one approach than another (e.g. granular data might be better suited for a pull approach; some critical sub-sets of data that could be available in almost real-time could be retrieved through a pull approach while the rest of data calculated could still be pushed with a regular frequency).

355 In addition, a mixed approach would be a solution in the cases where there are certain impediments (technical or legal) to get access to certain data (some microdata may not be

available because individual institutions/entities in the group simply do not belong to the integrated system or there are other legal concerns to obtaining this data).

356 A mixed approach could also be considered in the context of latent data (data that is not necessarily required by the authorities) that might be needed and required at short notice (e.g. such situations might be employed for crisis situations). In Spain, institutions aggregate and push data on deposits, and, in addition, keep in their internal databases a granular repository, at authorities' disposal, following a standardised data model established by authorities regarding those deposits.

357 A mixed model could also consider for proportionality aspects as smaller institutions might not have the necessary resources to align for the technical infrastructure and architecture required for a well-functioning pull mechanism.

Summary

		Disadvantages	Advantages
Push model	**Competent authorities**	• Little knowledge on the internal transformations applied by the institutions to the data • Access to the data can only take place according to the reporting obligations calendar	• Passive receiver of the data • Easier to detect changes to the data • Facilitates consistency in prudential assessments and decisions
	Institutions	• Responsible for setting up the connection to the competent authority infrastructure • Responsible for the mapping of the data from their internal systems to the required output format • Quality of the data can only be checked after the data has been submitted	

Pull model	Competent authorities	• Responsibility for the usage of a broader scope of data (legal risks of not acting) • Harder to identify changes to the data • Technology investments for pulling, analysing and storing potentially huge amounts of data • May imply a risk of inconsistent assessment of the data if standards and coordination mechanisms (e.g. in terms of frequency, timeliness) are not properly set in place	• Possibility to access the latest data in almost real time • Full transparency on the transformations applied to the data
	Institutions	• Efficiency costs of implementing a standardised model for the data repository • Responsibility for assigning different access rights to different users of the data (e.g. different authorities collecting data) • Might need to store both aggregated and granular data	• Passive receiver of the data requests

Questions to consider

40) On which model/architecture would you prefer the future integrated reporting system to be based?

41) What would be the main advantages and disadvantages to consider and/or what would be the main challenges to consider and what would be the possible design options (from both a technical and process perspective)?

8. Governance

358 The objective of this section is to highlight the relevant governance aspects to be taken into consideration for the feasibility of the integrated reporting system and what would likely need to be defined and agreed to ensure that any future integrated reporting system could be implemented and managed. Additional work on the legal considerations and appropriate allocation of responsibilities surrounding governance and operational issues would need to be undertaken at a later stage once a specific path is defined for data dictionary and central data collection point system. Detailed governance considerations should be provided before undertaking any more formal development.

8.1 General governance considerations

359 At this early stage of analysis of the different options, with many elements of the scope and scale still undefined, it is not possible to provide a specific governance set up. Instead, this discussion paper includes a broad analysis of the main governance considerations that would need to be taken into account to achieve an integrated reporting system. A more detailed assessment on these considerations would need to be performed once the different options are analysed on data dictionary and central data collection point. The general consideration could focus on the following main topics: submission of data, access to data, data sharing and the general governance of the possible future integrated reporting system.

8.1.1 Data submission

360 Under the current situation, many of the reporting frameworks follow the sequential approach where the data is reported to national authorities and then on to the European authorities. According to the fitness check study, a large majority of national authorities consider this model to be particularly effective since it was noted that the increased knowledge of and proximity to the reporting entities tends to have positive impacts and more timely access to data allows national authorities to react more promptly to any adverse developments.

361 Taking the assumption of a centralised data collection point as suggested by the mandate in Article 430c CRR, it would imply that the data could be submitted to a central platform and could be provided to the relevant authorities when needed. Some considerations would have to be taken into account in this case:

- Definition of agreements and possible legal and regulatory changes would need to be investigated to mandate submissions of data to the central platform.

- Strict confidentiality rules applying to the central platform would have to be defined.

- Definition of harmonised principles for data queries.

Figure 20: Data submission
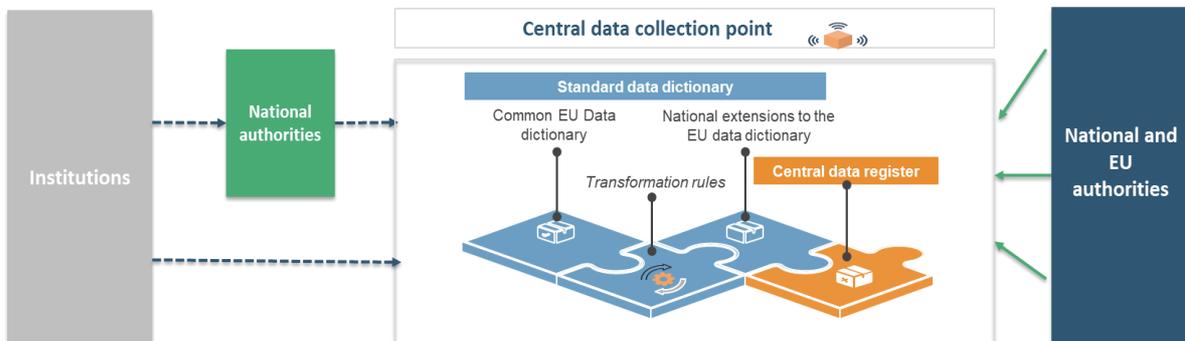
8.1.2   Data access and data sharing

362 Improved data-access and data-sharing arrangements across the relevant authorities at national and EU level would further streamline the reporting process, by reducing the duplication of data collected by multiple authorities to achieve the objective of 'report once'. In some cases, it has been noted that there are some difficulties in sharing data among authorities due to legal barriers, and in other cases, to avoid data duplication, some MoUs have been agreed. There is still room for improvement with regard to data sharing and cooperation among authorities. Therefore the mandate calls the EBA to investigate a CDCP which should serve as a point of contact for the competent authorities, where they could receive, process and pool all data and in addition hold the coordination role for the exchange of information and data between authorities.

363 All authorities could access data directly from the central data collection point. Therefore, from a governance perspective some principles for data access and a flexible approach to avoid data silos would need to be defined. These principles should specify:

- the data ownership of the different types of datasets;

- confidentiality settings and data protection;

- the list of participating authorities and the possible reciprocal access through data agreements by specifying:

- scope of data access: the central platform would provide authorities with a level of access in line with their data needs, providing further detail and specificity as to what data each competent authority is entitled to receive;

- permitted use of data: the permitted uses and other relevant mandates would need to be defined, as would protection of confidential information that is accessed and consequences of any breach;

▪ data responsible in case of any breach.

8.1.3 Information security considerations

364 The financial sector is a highly regulated environment; thus, information security and compliance will be important drivers during the choice of the integration model to be used. This will require that any solution should follow 'security by design' principles. More, as recent events have shown, data protection and privacy concerns should be addressed early on ('privacy by design'), so that any risks related to unlawful processing of personal data are mitigated.

365 Any integrated model should consider and define, at the minimum, the following:

▪ A RACI model, specifying the responsible, accountable, consulted and informed people in the process. All security efforts should be governed.

▪ The integrated system should have a unified information security policy that establishes the protection goals for the confidentiality, integrity and availability of the system data.

▪ A clear access management policy, coherent security markings for data assets (classification) and protection requirements for each marking category.

▪ An overarching risk management framework that constantly assesses the system and keeps track of the risks identified in a centralised risk register, which should hold at the minimum, the risk description, likelihood and impact, risk owner and the mitigation actions foreseen to lower the risk.

▪ The system design should follow well-established design and architecture patterns and use standard cryptographic primitives and protocols.

▪ The system should assign resources for maintenance activities (operating, patching, updating, upgrading, etc.).

▪ The system should be constantly tested and audited.

8.1.4 Legal obstacles and challenges

366 Any potential boundaries of the applicable legal frameworks regarding the interactions and derived governance structures between participants of the integration process (legislators, participating authorities, industry) will need to be assessed. However, in order to clearly define these considerations, further progress is needed in the different areas of integration.

367 In addition, there might also be some other legal barriers with regard to data sharing among authorities and the reuse of data for different purposes.

368 Some authorities could be reliant on the CDCP to do some checks on their behalf. This outsourcing of some tasks might not be allowed in some jurisdictions without a change in law or regulation. This particular issue, would have to be addressed in later stages of the analysis once a clearer picture of the integrated system functioning is set up.

369 The legal boundaries regarding the level of granularity in data collection from existing law (e.g. with regard to confidentiality and data privacy) should be closely investigated. In this context, the possibility to amend the relevant law, where appropriate and possible, or to find alternatives (e.g. anonymisation) should also be considered.

370 Legal constraints in national or European Union law aiming to protect central public good might impose further restrictions on the collection of granular data. Consequently, persistent legal requirements are natural boundaries for the level of granularity. On the other hand, there might also be legal constraints, which do not arise from national or European law, but from third-country law. These legal frameworks form a second natural boundary for the level of granularity (for at least certain parts of the dataset) which cannot be influenced at all by the statistical, prudential and resolution authorities, respectively.

371 The collection and use of personal data should comply with The General Data Protection Regulation (GDPR) in the EU. According to the GDPR, collecting personal data for statistical purposes does not represent an unsurmountable constraint for the data collection, provided that aggregated data no longer permit the identification of data subjects[61].

372 If personal data were collected, it might require anonymisation or pseudonymisation. Anonymisation requires that all information which would allow an inference on the actual natural person to whom the data belongs, be removed. When using pseudonymisation data is processed in a way that an inference on the actual person the data belong to is not possible anymore.

373 Furthermore, while reporting institutions remain responsible for the content of the data, data receiving authorities become responsible for compliance with data protection requirements as soon as they receive the data. Considering this, any integrated reporting approach should be based on the assessment of the strictest national (if any) or European law. Especially with regard to the 'need-to-know' principle, it is important to be able to assess in any case who is able to access what kind of collected data, at which level of granularity and for what kind of purpose.

---

[61] Recital 162 and Art. 89 GDPR.

Questions to consider

---

42) Could you please specify any legal obstacles and possible solutions you would see related to the following? (Feasibility of the central data collection point, collection and access to granular data by authorities, the responsibility of the reporting institutions for the reported data, if the transformation of granular data is conducted by authorities (i.e. not the reporting institution), other.)

8.1.5 Governance: centralised/decentralised approaches

374 For governance of an integrated reporting system, a model of centralised and coordinated decentralised approaches could be considered.

375 Different parts of the reporting process chain could benefit from different approaches and the following analysis provides an overview of possible impacts on the level of harmonisation and efficiency point of view.

Data definition

376 At the level of the definition of concepts, there is some consistency on definitions mainly through the DPM dictionary, which comprehends – for the moment – the definitions of the prudential and resolution reporting frameworks. The BIRD dictionary also includes statistical definitions and reuses the definitions of the DPM for the prudential data at input-layer level. Some harmonisation has been achieved during the past years in order to create some consistency across the different reporting frameworks. The consistency of definitions is key in order to create an integrated reporting system. Each new data element has to be included following the same consistent formal approach and selection criteria of the common vocabulary. Similar concepts need to be analysed and data standardisation has to be taken accordingly with the regulatory definitions.

377 In order to achieve a greater level of harmonisation necessary for a common data dictionary, it is desirable for the semantic level of data dictionary to follow a centrally managed approach to ensure the technical standardisation that creates an effective data consistency and data comparability between the different data concepts. A centralised governance of dictionary concepts would be challenging but may be necessary for achieving a complete harmonisation necessary for a common dictionary.

Data Transformation

378 Governance for data transformation should be managed centrally for the common harmonised transformation, regulatory concepts and validations.

379 Other transformations and additional validations could be managed in a decentralised manner.

Data collection

380 The precise governance model will need to be defined following the architectural design but a hybrid model of centralised and coordinated decentralised model could be achieved.

381 A more centralised approach could fit more centralised architectures and result in full data integration of the reporting requirements under the principle of 'define once and collect once'. A centralised approach would need to be considered in further detail due to the implications it will have on competent authorities and the investments already made by them during the last years. A more decentralised approach could be an option in order to benefit from the investments already made, but it would only be feasible if there were enough interoperability between frameworks to reduce the reconciliation efforts. In addition, it will also rely on a higher level of cooperation between authorities that receive the data, in order to receive data once for multiple reporting purposes and share it amongst authorities in an efficient manner.

382 Data quality is an important part of the reporting process and directly affects the accuracy of the data reported and could also be managed with a hybrid approach. A coordinated and decentralised approach for data quality offers benefits such as the prompt reaction from authorities at national level and better knowledge of each reporting entity in order to assess their data quality. National authorities could have priority access to the data and quality of the data reported would be checked on reception at the integrated system. However, the process would also benefit from an automated and centralised approach, where similar and harmonised systems and formats of validations could improve the data quality process and streamline parallel and manual processes.

Data exploration

383 The data governance of data exploration inherits what is decided in the earlier phases of the reporting process chain, as it is mainly a data consumer rather than a data producer part of the process.

384 There could be different options. If it is fully centralised, new services will have to be approved by a central authority before being released to all the users. In the event it is fully decentralised, each authority could release new services without any authorisation or quality check.

385 From this perspective, it seems that a hybrid approach could be more suitable, i.e. ensuring the reviewed and approved services are well identifiable and defined according to quality standards, while also allowing non-certified service to be shared. From those non-certified, the community

may elect or propose some of them for official endorsement, and perhaps they may or should already be developed according to some standardisation principles.

386 The hybrid approach could offer some degree of flexibility to steer the direction in the future.

8.2   Coordination of data requests

387 Although most of the regulatory reporting is harmonised, currently reporting entities provide different types of datasets to many different authorities, in some cases the data is requested through the harmonised reporting frameworks but in other cases, some further data needs are requested through ad hoc requests and national reporting. A detailed proposal of classification of reporting datasets according to their characteristics can be found in Annex I of this discussion paper.

388 In order to promote a better coordination between authorities with regard to data requests, this discussion paper proposes an agile coordination mechanism for ad hoc and national data requests.

8.3   Agile coordination mechanism of data requests

389 The Fitness check on prudential reporting provided evidence on ad hoc requests being particularly challenging and resource intensive and hence hampering efficiency of EU prudential reporting. Authorities deem ad hoc requests crucial, especially during crisis times, but also agree that there is room for improvement on how they are managed and used. Given the importance of these data requests in the overall reporting landscape, it is necessary to find an efficient way to manage these requests and at the same time ensure authorities have the necessary data in their use.

390 In order to better delineate that what is effectively required from authorities and regulators in order to carry out their tasks effectively without imposing excessive burdens on financial institutions, the EBA analysed some options in line with the concept of CDCP described in Article 430c CRR.

391 According to Article 430c CRR, the CDCP will serve as a point of contact for the competent authorities, where they receive, process and pool all data queries, where queries can be matched with existing collected reported data and which allows the competent authorities quick access to the requested information.

392 Taking into account the mandate, the CDCP could serve as a platform to facilitate the coordination of data requests in a structured and flexible manner and provide solutions to avoid the data duplication and overlaps when collecting data from institutions. This would not imply that the CDCP would have a role in setting up the reporting requirements, only for coordinating requests. The legal powers of authorities to request data would still reside with the authorities,

with no change foreseen, as supervisors must be in a position to collect all information necessary in case of need.

393 The CDCP and its agile coordination mechanism would aim to increase discipline on data requests by reusing and sharing data between authorities and enhancing transparency by including data definitions into the EU common data dictionary. The agile coordination mechanism is expected to be a simpler process if it is based on a centralised system, but the process could also be built on a network basis, where all authorities and their systems would have contact points and interoperability.

394 By increasing coordination efforts, the objective is to increase discipline of requests by increasing transparency of data requests, to promote reusing and sharing of data, improve alignment of definitions and convergence of requirements. At the same time, it is of utmost importance that all authorities have the data they need to accomplish their tasks in a timely manner without unnecessary administrative burdens.

Figure 21: Coordination mechanism for data requests
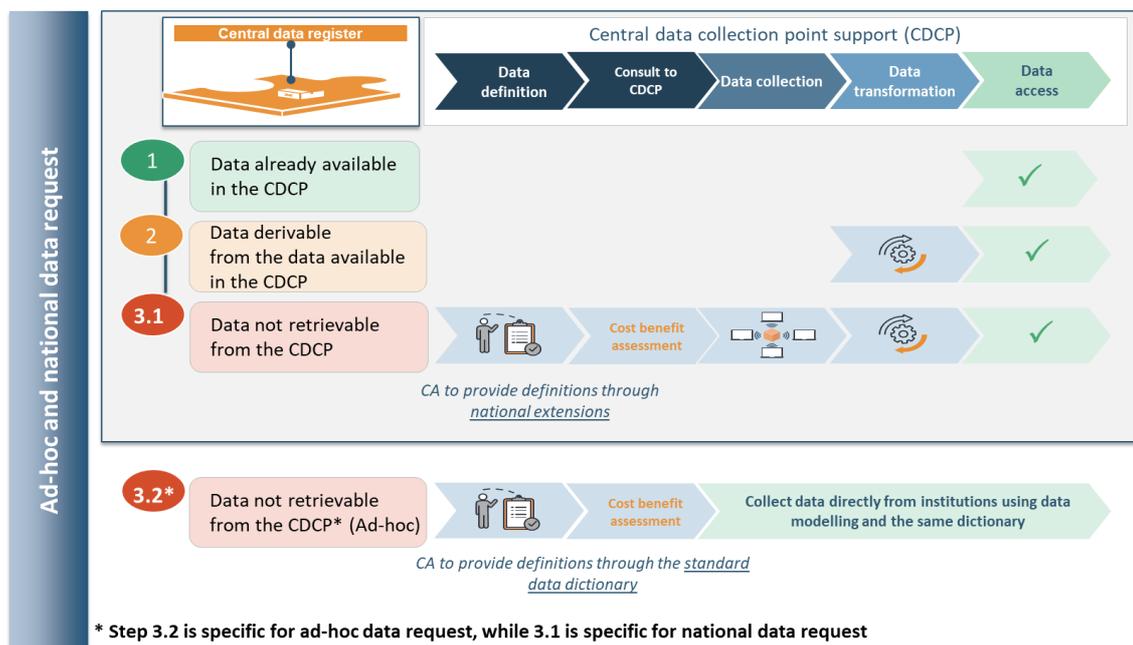
395 The agile coordination mechanism to process national and ad hoc data requests could consist of the following steps:

396 Each time the authority submits a data request (national request or ad hoc data request) through the Central data register or a network of authorities, there could be three possible options:

1) Direct access: the data requested is already available in the CDCP. The competent authority could have direct access to the data points requested.

2) Indirect access: the data requested could be derivable from the data available in the CDCP. The authority could access the data once some data transformations are done in order to comply with its data needs.

3.1) Data not retrievable - National data request: the data is not retrievable from the CDCP. In these cases, the authority would need to provide the definitions of the data points needed in order to complement the common data dictionary with national extensions that provide the definitions of the national requirements data needs. Once the data is defined, the authority would need to prepare a simplified cost-benefit assessment. After a swift review and assessment, the data could be collected through the CDCP, which would perform the necessary data transformations in case it is needed. The authority would have access to the data points requested after the data is defined, transformed and collected.

3.2) Data not retrievable – Ad hoc data request: the data is not retrievable from the CDCP. Once the data is defined, the authority should prepare a simplified cost-benefit assessment regarding the data collection (even though there might be exceptional circumstances where a proper cost-benefit assessment is not possible, e.g. particularly in stress periods there might not be enough time to perform a cost-benefit assessment). After the cost-benefit swift review, the authority would directly collect the data requirements after providing the data definitions.

In the cases explained above under points 3.1 and 3.2, a proper governance process for such situations should be considered as well as any possible exceptions.

397 As explained in Section 4 on Data dictionary, national extensions are a lighter version of the data dictionary and should be developed by national authorities following similar criteria as the common data dictionary.

398 There are also some pre-conditions for this coordination mechanism to work in an appropriate manner and be efficient:

- Agile approach: authorities should receive the data requirements for both national requests and ad hoc requests in a sufficiently timely manner to meet the needs of authorities' mandates. This may include the need for rapid responses, for instance during periods of market stress. Any administrative burden should be streamlined.

- Provide assistance to authorities: some guidelines and standards should be provided to authorities in order to follow harmonised standards for the modelling of definitions and formats along the process.

- Legal powers to request information: the introduction of such an agile coordination mechanism should not impact the authorities' legal powers to request information.

- Exclusions or ex-post inclusion: there may be situations where it merits excluding a request or including it only ex-post into the system, for example in a crisis situation

where data is very urgent and also very much institution specific the request may not merit standardisation.

399 Therefore, the EBA would recommend that any possible integrated reporting system should count on an agile process for the coordination of national and ad hoc data requests through the CDCP and at the same time guarantee quick access to the data requested.

8.3.1 Supporting processes for the coordination mechanism

400 The agile coordination mechanism could be supported by best practices developed at national level for the analysis of data requests. These practices could help in streamlining and improving the internal governance processes at the national authorities whereby new (regular or ad hoc) data requests are vetted and can be approved only after checking that the data is not already reported and assessing the burden on supervised entities.

401 Creation of committees in each national authority to perform a preliminary assessment at national level in order to:

- conduct an analysis of similar data requests already in place throughout the EU;

- search for similar definitions already present in the common data dictionary;

- explore the need to create new definitions in the common EU data dictionary or adapt the new data requests to the information currently available.

402 The scope of data requests to be included in the committees for discussion should be at least: statistical, prudential, resolution.

403 Each authority should have a register of the different data requests, and national request should be stored in a central data inventory.

404 Any new data requests should be processed by using national extensions of the common data dictionary. Existing national data requirements should also be included in the common data dictionary as national add-ons. Each authority should have the option to not include the existing national data requirements in the integrated reporting system where not feasible or in case of very specific data requests.

405 Each authority will be in charge of preparing the modelling for the data requests by making use of the harmonised standards for the modelling of definitions and formats along the process.

406 Each authority should have the option to not include the new data request in the integrated reporting system, or to include it at a later stage, provided it present a justification for that (very specific request, very short notice or short term, information collected for purposes other than statistical, prudential, resolution, etc.).

8.4   Joint Committee

407 The mandate on Article 430c CRR refers to the creation of a Joint Committee (JC) for the development and implementation of the integrated reporting system.

408 For the purposes of this study some general considerations are provided with regard to its main tasks and structure. However, further details on its specific functioning would need to be revisited once a clearer view on other technical priorities are covered and defined.

409 The JC would act as a forum of authorities involved in the efficient development and implementation of the integrated reporting system. Its main objective would be strengthening cooperation and coordination among authorities.

410 Considering its main objective, the JC could have a role in different dimensions that conform or would be a key part of the integrated system:

- coordination of data sharing between relevant authorities (statistical, resolution, prudential);

- implementation of the integrated reporting system and technical aspects to be covered in the implementation phase (common dictionary, architectures impact);

- exchange of views with regard to the systems, tools and technologies to be used in a possible implementation phase;

- outlook of the possible future actions on the development and implementation plan that may wish to be pursued after the feasibility study analysis.

411 The setting -up of the Joint Committee should be done once the feasibility study is concluded and should take as a basis the feasibility study conclusions. An informal coordination mechanism among authorities could be organised before setting up a Joint Committee in order to create a forum between authorities to discuss the main aspects of the Integrated reporting system.

Questions to consider

43) Do you agree with the suggested coordination mechanism for data requests? Do you see any benefits or disadvantages to this approach?
44) Please specify how the agile coordination mechanism for coordination of data requests could be further simplified and how your proposed measures could enhance coordination and avoid data duplication?
45) According to the reporting classification proposal included in Annex I: are there further reporting criteria to be taken into account under Category 3 reporting?
46) According to the reporting classification proposal included in Annex I: Do you agree with the proposed approach for non-recurring type of data Category 4?
47) What solutions could the EBA investigate that would reduce costs?

9. Technology impacts on prudential reporting

412 Authorities depend on access to high-quality data to fulfil their respective missions. The definition of the regulatory data is the first of a sequence of processes, which aim to gather the adequate prudential and financial data from institutions and make it available to different users of information, not only authorities, but also policy makers, institutions, investors, depositors, academia, researchers and others.

413 Achieving an end-to-end digitally supported chain of regulatory processes is key to achieve data transparency and integration, improve data availability and data comparability, increase efficiency and effectiveness of the processes and enhance the system interoperability.

414 Making use of the latest technologies and best practices in data management all along the reporting process chain (in the definition of regulatory requirements and in the collection, transformation and exploration of data) is a fundamental key stone aspect to consider.

415 Regulators have an important role in developing robust and adequate data standards, which are key elements in:

- supporting the digital processing of reporting requirements;

- promoting a transparent and harmonised playground for institutions in their regulatory compliance;

- facilitating open and trustworthy technology development.

416 This section looks at the use of technologies by institutions in complying with the prudential requirements and reporting, leveraging evidence from the Cost of Compliance study[62]. Using technology by the reporting institutions may be facilitated if the market offers solutions in this respect. More specifically, Regulatory technology (RegTech) is expected to help financial service providers with their compliance tasks through automation of process and decision-making. Leveraging the preliminary results of a survey, this section is looking to show the developments in this area.

---

[62] In particular Questions Q13 and Q14

9.1   RegTech at institutions

417 Regulators and supervisors, alongside the industry, are investigating possible solutions to adapt supervisory reporting requirements to take advantage of the latest advances of technology.

418 RegTech services have appeared as a new actor in the regulatory landscape, with the potential to help institutions to manage compliance tasks, including reporting, through an automated process.

419 In order to analyse the use of technology through RegTech services, the EBA has compiled preliminary feedback received from a survey on the use of RegTech services that has been answered by institutions and RegTech providers. The aim of the survey was to better understand how innovative technological solutions have been implemented for prudential reporting purposes, what are the main reasons and main impediments for its implementation. The information provided in this report counts only on the preliminary analysis of the results and only on the information provided by RegTech and institutions that have provided answers to the questionnaire submitted for this purpose. The preliminary analysis does not take into account information from institutions that are not making use of RegTech services.

Main reasons to use RegTech for reporting purposes

420 Reasons for using RegTech for prudential reporting are diverse. From a RegTech providers' perspective, the main reasons to implement a reporting solution are to facilitate the regulatory data integration, the implementation of ongoing regulatory changes and the need for efficiency.

421 From the institutions' perspective, the key reasons to use RegTech are to focus on the need to reduce human errors, enhance risk management, enhance system and data integration and enhance monitoring.

Main challenges

422 The survey highlighted some obstacles related to the use of RegTech for prudential reporting purposes such as the identified lack of standardisation (data interfaces, etc.), non-availability of data, data quality issues, lack of client API (Application Programming Interface) capabilities.

423 From RegTech providers' perspective, the challenges perceived as highly relevant are the lack of regulatory standards, lack of interoperability and the continuous changes of national and international regulation.

424 From institutions' perspective, the major obstacles and operational challenges for using RegTech solutions are data quality issues, changes in national and international regulation and integration with legacy systems.

425 Both financial institutions and RegTech providers agree that initiatives on regulatory data standardisation and data integration are key priorities in order to support the uptake of reporting RegTech solutions.

RegTech use through the reporting process chain

426 The figure below shows different parts of the reporting process chain and which steps of the process are covered by technology and hence have an increased level of automation.

Figure 22: Reporting process areas covered by RegTech solutions
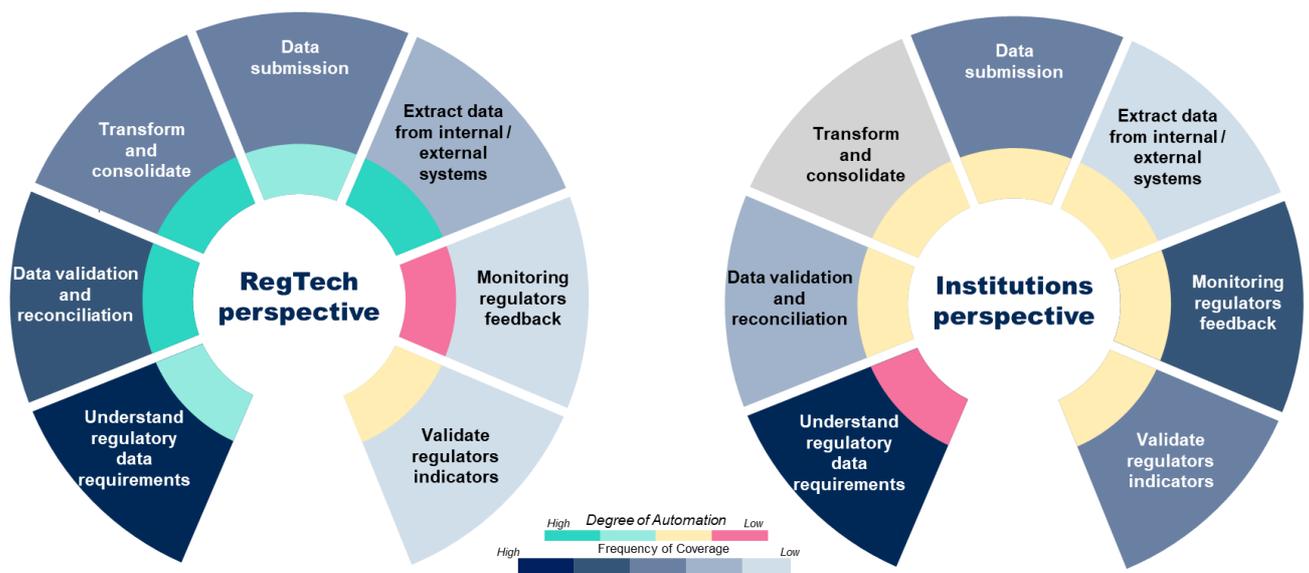
427 The evidence collected from RegTech providers shows that most of the RegTech services provided are focussed on the first steps of the reporting process, mainly on understanding the regulatory data requirements, data validation and data reconciliation. Similarly, those steps of the reporting process present a higher level of automation.

428 From the institutions' perspective, the use of technology is mainly focused on understanding the regulatory data requirements, monitoring the regulators' feedback and data submission. Very few new technologies are used in transforming and consolidating the data.

429 Under the current reporting landscape, the deployment of these tools has been motivated by the need for improvements in efficiency, including cost reduction and enhancements in regulatory compliance effectiveness.

430 From the preliminary analysis, fact finding and interviews, some first views could be inferred:

- Any strategy on simplifying and streamlining supervisory reporting requirements would have to take technology developments into account.

- Inconsistent definitions and insufficient standardisation of formats and processes hinder the development and application of new technologies. There is a need for common definitions ('define once') and further standardisation would be needed for developing standardised reporting requirements.

- The main aim for using RegTech is to improve the efficiency of the reporting processes.

9.2 The technologies used – evidence from the Cost of compliance study

431 The EBA Cost of Compliance study was looking to gather some evidence on the nature and complexity of the IT solutions used for regulatory compliance and associated reporting obligations. The objective was to identify how institutions perform regulatory reporting on top of their regulatory compliance, understand the general setup of the IT solutions used along the four different phases of the regulatory data chain: understand regulation, extract data from sources, calculate and reconcile data, report and monitor data feedback.

432 The industry responses received from the qualitative and quantitative questionnaires were 408 and 298, respectively. The responses cover 8.5% of all banks (27.6% of large banks, 3.3% of medium ones and 10.5% of small and non-complex banks). The following graph shows the distribution of institutions by country and size.

Figure 23: Industry responses
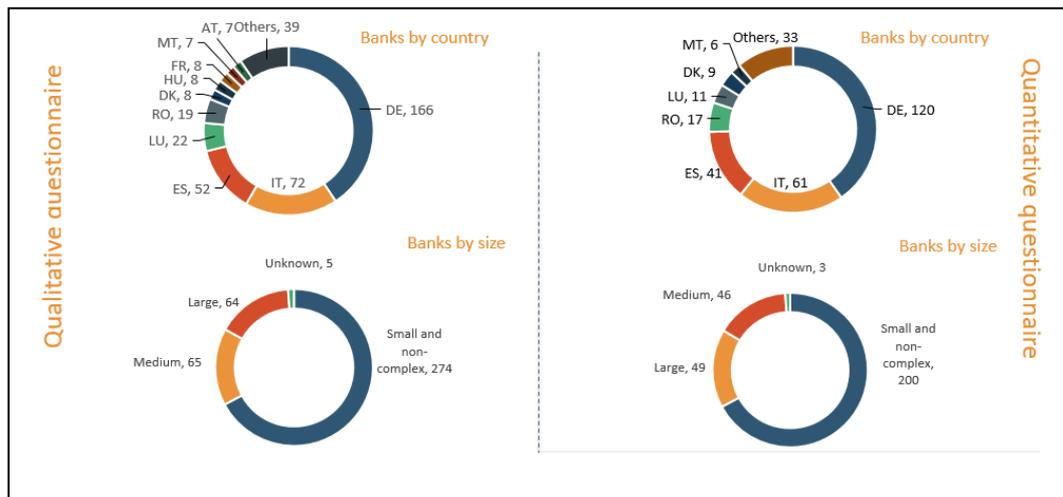
Differences in the processes of regulatory compliance and regulatory reporting

433 For the purpose of the CoC study, the institutions were asked to answer how they use technology and organise their data through the four principal phases of their processes of regulatory and reporting compliance: i) understanding regulation; ii) extracting data from sources; iii) calculating and reconciling data; and iv) reporting and monitoring data for compliance.

434 There are no significant differences in the type of IT solutions supporting both the overall regulatory compliance and the reporting processes. For all types of institutions (small and non-complex, medium, large) and in each of the different phases both compliance processes use similar types of technical solutions. This can be a sign of possible integration of reporting processes within the internal processes of compliance.

Setup of the IT solutions in the distinct phases of regulatory compliance

435 The general setup of the IT solutions is very different along the distinct phases in all types of institutions

436 In the initial phase of understanding regulation, small and medium-size institutions are mostly relying on service providers' solutions and, in second place, on internal solutions, being less

dependent on COTS[63] software. Large institutions have an equal usage of service providers, internal solutions and COTS Software. Other possible technologies are not relevant to any types of institutions.

437 In the phase of extracting data from the sources, small institutions are almost exclusively relying on service providers' solutions, few have internal solutions and the usage of COS is insignificant. Medium institutions are relying equally on service providers and internal solutions and the COTS software is less relevant. Large institutions use less service providers and have a strong implementation of internal IT solutions with some COTS software usage.

438 In data calculation and reconciliation small institutions are almost exclusively relying on service providers' solutions, few have internal solutions and the usage of COTS is insignificant. Medium institutions are relying more on service providers, but they still have a relevant number of internal solutions and also increase on COTS software in comparison with data extraction processes. Large institutions are using less service providers and have strong preferences for implementing internal IT solutions and use of COTS software. However, when comparing with the data extraction phase they are increasing all the three types of IT solutions, which can be a signal of a more intense use of IT solutions in this phase of regulatory and reporting preparation.

439 In report and monitoring processes small and medium-size institutions are almost exclusively relying on service providers' solutions, few have internal solutions and the usage of COTS is insignificant. Large institutions use some service providers, but are relying more strongly on internal IT solutions. The usage of COTS software is more significant in report and monitoring than in the other processes.

440 The cost of compliance study was looking to gather some evidence on banks' practices with respect to data architecture and metadata management process, the level of data integration achieved by banks, the degree of automation in data processing, data stewardship and ownership and the degree of automation of the data quality process.

Data architecture and metadata management process

441 A large part of responding institutions, irrespective of the size, have implemented to a large extent (medium-high and high) architecture standards for data acquisition, processing and provisioning in their institutions and have standard operation procedures (medium-high).

442 A large share of the responding institutions have achieved (at a medium level) the creation of a single, integrated business glossary that documents all business terms used for describing data inputs and outputs used for regulatory reporting, while medium institutions seem to have made better progress in this direction than large and small and non-complex institutions. Most of the responding banks marginally use specialised tools (e.g. natural language processing of legislation) in the management of the business glossary with the exception of a large share of

---

[63] Commercial-off-the-shelf (COTS) software is a term for software products that are ready-made and available for purchase in the commercial market.

the respondent institutions in Italy and Spain (irrespective of their size) that are making high use of such specialised tools. In terms of resources to maintain and manage the business glossary, while medium and large banks seem to be in a better position to have been able to identify appropriate roles and resources, small and non-complex institutions have been less able to cover their needs.

443 In terms of deploying a metadata management system (and a related management process), that documents all key data objects (data models, tables, columns) and dependencies, both large and smaller banks have achieved this at a medium level. While there are cases, where large and medium institutions have highly achieved the implementation of such a system, only 4 out of almost 130 small and noncomplex institutions have (highly) implemented them. The level of integration of bank's metadata management system with other metadata repositories (e.g. such as those defined by authorities) is medium to low. Few exceptions that responded high are banks in Spain and Italy that have a high integration with these, irrespective of the size. While large and medium banks seem better organised in maintaining references and data field specifications within a searchable central catalogue, most of the small and non-complex ones do not (medium to low).

Data integration

444 In terms of the degree to which institutions have managed to integrate the regulatory requirements (e.g. in a single or interoperable data lake) at the organisational level, a large share of respondent institutions seem to have achieved it at a medium level. Very few institutions have not achieved it at all which might show that integration was indeed an objective or that the market provides solutions in this respect. Most of the institutions have achieved real-time data collection and/or processing to a limited extent, while certain institutions (among the entire size category) report that they are highly capable of implementing such a real-time process. On the other hand, the process of archiving, backing up and retaining data has been largely automated at an organisational level in most cases irrespective of their size.

Data processing automations

445 Across bank size there seems to be no difference in terms of the level achieved in automating data processes. Most of the banks seem to have limited manual/human interventions at a moderate level (the majority at a medium high level) and have managed to a larger extent to introduce processes and procedures to facilitate and improve the implementation of change requests and/or developments (medium-high and even high). On the other hand, less automated processes and procedures have been set in place to identify dependency in the data and further link them through an automated process.

Data stewardship and ownership

446 Data governance aspects such as processes that include clear responsibility for the data lifecycle management are highly relevant for larger institutions and moderately high for medium

ones. Similarly, larger institution have been in a better position to identify business user profiles in the organisation to cover the role of data owner and data steward for a specific data category. On the other hand, many small and non-complex institutions implemented such governance aspects at a low level or outsourced them (the majority responded with 'I don't know ').

Data quality

447 Data quality standards are essential to ensure compliance with reporting and regulatory requirements (and may be more important to ensure any business decision is taken based on accurate information). A high degree of automation of the data quality process is essential to ensure the quality of the data (reduce manual checks), replicability and reusability of processes (easy to detect where inconsistencies may occur in the final data) and fast response to changes. The cost of compliance suggests that the degree of automation of the data quality process has been achieved to some extent across all types of institutions; however, interventions outside the reporting system are still needed. Compliance with regulatory validations is more highly automated.

448 Evidence from the cost of compliance shows that the majority of the institutions, irrespective of their size, have managed to automate the process of detecting inconsistencies in data (syntactic, semantic) to a high and medium-high level, however there remains a significant number of banks where the automation process remains at medium-low. Many of the small and non-complex institutions have outsourced this process.

449 Most of the institutions need to perform additional activities (medium-high), outside the reporting system to validate the data reliability. In terms of completeness of data, the majority of institutions don't see a significant need (medium-low) for the business user to intervene and integrated data from other systems in order to fill data gaps (e.g. empty fields, partial data availability, missing information), especially in the case of small and non-complex and large institutions. For medium institutions this aspect might raise more importance (medium-high), possibly higher data volumes/dispersion creates more complexity in the processes (compared to small ones) but the processes are not that automated (as might be the case for larger institutions). Therefore, manual intervention to fill data gaps for medium banks might be required more.

450 A governance aspect for the data quality process (in particular the data cleansing process) has only partially been implemented by the responding institutions. Most of the time the data cleansing cannot be performed in the phase of extracting and integrating data (medium-low) and additional inconsistencies are inferred probably later in the process, making the process more burdensome, as previous steps have to be re-run in order to detect the source. Large banks seem to have marginally achieved to clean the data at early stages in the process to a higher extent (medium-high) thus preventing low-quality data from being used downstream.

451 Meeting the compliance with the validation rules defined for the reporting framework or by the competent authority in a more automated way (have been integrated in the reporting

system) has been done to a high (medium-high) extent by large and medium institutions and at a medium-high (high) extent by small institutions.

9.3   Preliminary conclusions on technology

452 The digital transformation on regulatory compliance depends very much on how all the stakeholders and special institutions establish effective use of trustworthy technologies. The surveys the EBA is doing have demonstrated different situations and levels of technology maturities, but an increased awareness of the importance of data standardisation as a way to improve data availability and to streamline the processes of regulatory compliance.

453 On RegTech firms, we can verify an important increase in the products available in the market and a relative intense activity addressing the needs of the different areas of activity of financial institutions.

454 Institutions are making use of automated processes in various parts of their compliance and reporting processes. However, the degree of coverage is very different from one institution to other and few institutions admit to having implemented fully automated processes. In general, the small and non-complex institutions have outsourced (parts or all) of their compliance and reporting-related activities to third parties leveraging on the pooled experience and technology of a single provider (e.g. respondent institutions in Germany). On average there does not seem to be clear discrepancies between the size of the banks and the degree of automation of the processes. However, it is visible that the big and medium institutions prefer to implement solutions developed internally and use less service providers' solutions when compared with small institutions.

455 On other hand, the few responses received in the RegTech survey indicate there is an opportunity to improve the space between RegTech and institutions and leverage the potential of innovative technologies.

456 This discussion paper is looking to gather more evidence in this area and in addition to infer on the areas that regulators could work on together with all stakeholders involved, supporting the digital transformation of the regulatory compliance and reporting chain.

Questions to consider

48) Are you making use of RegTech for reporting purposes? If yes, please specify in which reporting process step. If not, please explain according to the options provided in the questionnaire.
49) Which of the reporting process steps would benefit more from RegTech development? (Options provided in the questionnaire section.)

Questions to consider

50) Do you agree with the main obstacles highlighted in the discussion paper? Do you see any further challenges?
51) Would you be keen to invest in RegTech for integration of different types of data?
52) How do you think RegTech can help in data integration?
53) Do you agree that data standardisation is the first step necessary for using RegTech?

10. Questions for consultation

The purpose of the Discussion Paper is to present preliminary analysis and options considered so far on the feasibility of creating an integrated reporting system for prudential, resolution and statistical data. The discussion paper aims to gather additional evidence and opinions on the topics presented and to serve as a basis for future discussions with various stakeholders.

The EBA is looking to receive feedback from reporting institutions and other stakeholders that consider they might be impacted by any topic or option outlined in this discussion paper or that might have relevant information that would help to form a complete picture on them. The EBA will prepare the Feasibility Study taking into account the feedback received. When reading the discussion paper please consider the questions below to which we ask for your feedback.

➢ **When providing your feedback please refer to the specific section or paragraph number**

➢ **If you agree to being contacted by the EBA in order to provide additional technical feedback through some interviews, please provide the email at which we may contact you.**

➢ **Please consider, in the reply questionnaire, each question will have an open field where you may provide further comments**

10.1 General questions

1) Please explain which institutions you think should be considered by the Feasibility Study.

2) Please explain which data collections you think should be considered by the Feasibility Study.

3) Do you consider that the issues identified, the options proposed and the assessment approach taken throughout the discussion paper are relevant and complete? If not, please explain.

4) What do you perceive as the key obstacles and operational challenges to develop an integrated Reporting Framework (for your institution)?

	Not relevant	Somewhat relevant	Relevant	Highly relevant
Training / additional staff (skills)				
IT changes				
Changes in processes				
Changes needed in the context of other				

counterparties / third-party providers				
Time required to find new solutions				
Other (please specify)				

10.2 Section 2. Stocktake

5) Do you confirm the findings presented in the stocktake? If you have additional information, please provide more specific details about the amount of data collected.

10.3 Section 3. Reporting process

6) Do you agree on the holistic approach used and the assessment done for the integration assessment (different steps of the reporting process chain and different levels of integration?

	Highly agree	Agree	Somewhat agree	Don't agree	Comments	What solutions should the EBA investigate in these areas that could help to reduce reporting costs?
Data Dictionary						
Semantic level						
Syntactic level						
Infrastructure level						
Data collection						
Semantic level						
Syntactic level						
Infrastructure level						
Data transformation						
Semantic level						
Syntactic level						

Infrastructure level	
Data exploration	
Semantic level	
Syntactic level	
Infrastructure level	

7) Please specify any further costs[64] or benefits you envisage related to the different stages of the reporting process chain:

Reporting process stages	Comments on the costs and benefits already identified	Additional costs identified	Additional benefits identified
Data definition			
Data collection			
Data transformation			
Data exploration			

10.4 **Section 4.** Data dictionary

8) Do you use one or more data dictionaries in your compliance and reporting processes?

9) What are the characteristics you think a data dictionary should have? Do you agree with the one referred to in this document? Do you think any characteristic is missing or should not be included?

10) What is the role you think the data dictionary can have in regulatory compliance and reporting?

11) How would a standard data dictionary help institutions to improve the processes of:

	Significantly	Moderately	Low	Please explain how:
Understanding reporting regulation				
Extracting data from internal system				

---

[64] 'Costs' refer to qualitative aspects as described in the relevant sectors throughout the report.

Processing data (including data reconciliation before reporting)
Exchanging data and monitoring regulators' feedback
Exploring regulatory data
Preparing regulatory disclosure compliance.
Other processes of institutions

12) How important is it for institutions to have a unique and standard data dictionary for all regulatory data with the aim of ensuring consistent use across the supervisory, resolution and statistical reporting?

      a) Highly important
      b) Important
      c) Somewhat important
      d) Not important
      e) Other

13) How much would it cost to move to a unique regulatory data dictionary?
      a) Highly costly
      b) Moderately costly
      c) Not very costly
      d) Not applicable

14) How much cost reduction is expected by integrating the national regulatory reporting together with the harmonised reporting regulation into a unique data dictionary?

      a) High cost reductions
      b) Moderate cost reductions
      c) Small cost reductions
      d) No cost reductions

15) How much cost reduction is expected by integrating ad hoc regulatory reporting with harmonised regulation into a unique data dictionary?

a) High cost reductions
b) Moderate cost reductions
c) Small cost reductions
d) No cost reductions

16) Do you agree with the costs and benefits highlighted in the chapter? Do you see other costs and benefits when implementing a standard data dictionary?

10.5 **Section 4.6** Data granularity

17) What would be the implication of granular data reporting on the institutions' compliance with BCBS 239 (also in the context of the options presented)?

18) For which reporting areas (prudential, statistical and resolution or modules/parts of these areas) may the use of granular data present a solution? (multiple choices)
   ▪ statistical
   ▪ resolution
   ▪ prudential

19) Which of the options regarding the granularity of the possible future integrated reporting do you think is feasible (given the challenges and constraints highlighted in the discussion paper and possibly others) and preferable for you? What are the main challenges and possible solutions to consider? Please rank potential challenges in ascending order (i.e. starting with the most challenging item in your view).

	Feasible	Preferable	Main challenges	Possible solutions
Option1				
Option 2				
Option 3				
Other, please specify				

20) In case of Option 2, please specify how should the granular collection layer be designed to your best advantage (and benefit of reporting more granularly)?

21) What are the main benefits and costs of implementing the option considered feasible and preferable by you from Question 19[65]?

Costs	Highly (1)	Medium (2)	Low (3)	No costs (4)
Collection/compilation of the granular data				
Additional aggregate calculations due to feedback loops and anchor values				
Costs of setting up a common set of transformations[66]				
Costs of executing[67] the common set of transformations				
Costs of maintaining a common set of transformations				
IT resources				
Human resources				
complexity of the regulatory reporting requirements				
Data duplication				
Other: please specify				

Benefits	Highly (1)	Medium (2)	Low (3)	No benefits (4)
Reducing the number of resubmissions				
Less additional national reporting requests				
Further cross-country harmonisation				

---

[65] The options as defined in the discussion paper take as given that a standard, common data dictionary would be implemented. Benefits and costs should then be analysed from the point of view of the granularity of this data dictionary

[66] E.g., participate with resources at setting up, testing and maintaining the transformations (please consider this in opposition to the current costs for the maintenance of unharmonised transformation rules).

[67] Please consider this in opposition to the current costs for executing internally unharmonised transformations.

and standardisation	
Level playing field in the application of the requirements	
Simplification of the internal reporting process	
Reduce data duplications	
Complexity of the reporting requirements	
Other: please specify	

22) What possible aspects related to the design of the option (Question 19) would make the costs for this option higher than the benefits and therefore not worth implementing?

23) If transformations are to be defined (as depicted in Option 2 or Option 3), who should be responsible for their definition (e.g. who takes responsibility for their correctness) and their execution?

      a) The authorities
      b) Reporting institutions
      c) Authorities and reporting institutions jointly
      d) Transformations would not be binding but the result of a cooperative effort and no responsibility attribution would be required.

24) If transformations are defined under different scenarios with respect to responsibilities, what are the major implications to the possibility of defining a more granular collection layer from a cost and benefit perspective also considering some of the challenges depicted in the paper (technical and legal, e.g. institutions need to remain responsible for all the data).

Responsibility for defining transformations	Costs	Benefits	Challenges	Design options/solutions
Authorities				

151

Reporting institutions	
Jointly authorities and reporting institutions	
Transformations are not binding	

25) How should the transformations be in terms of formalisation and readiness for digital processes?
   a) Harmonised and standardised, ready to be implemented by digital processes (fixed)
   b) Indicative instructions of calculation explaining the possible approaches (allowing for adaptations)

26) How could some of the challenges highlighted for defining transformations be overcome?

Manual adjustments

Consolidated/individual figures

Different valuations

Principle-based rules

Legal aspects

Others

27) What kind of data should be part of the feedback loops?

28) What other areas should the feasibility study investigate in terms of granularity and transformation rules?

10.6 **Section 5.** Central data collection point

29) Is your institution reporting to different authorities in your home country?
   a) Yes
   b) No
   c) Not applicable

30) Is your institution reporting to other authorities in host countries?
   a) Yes

     b)   No

     c)   Not applicable

Please comment: What problems arise from reporting to different authorities?

31) Are you using one or more data dictionaries for reporting? How?
     a)   One single dictionary
     b)   Multiple dictionaries

Please comment: how are you making use of them?

32) Are you using the same or different formats for prudential/resolution reporting and for statistical reporting?
     a)   The same format
     b)   Different formats

33) How important would it be, for your institution, to have access to a CDCP for all prudential, resolution and statistical reports? Why?
     a)   Very important
     b)   Important
     c)   Somehow important
     d)   Not important

34) What should be, in your view, the main characteristics of a CDCP?

35) What would you think could be the challenges, costs and benefits of changing to a CDCP?

System design	Costs	Benefits	Challenges
[…]			

36) What solutions could the EBA investigate that would reduce costs?

Aspect	Proposed solutions for the EBA to investigate to reduce costs
Architectures	

10.7 Section 6. Private-public collaboration/cost-sharing

37) Would the industry be prepared to bear the costs of integrated reporting?

a) Yes, to a large extent.

b) Yes, to a limited extent.

c) No.

38) Where (within the different elements discussed in the paper) do institutions and CAs see institutions' involvement and cost contribution as most valuable in the development of an integrated reporting framework? (**0= not valuable at all, 1= valuable to a degree, 2= valuable, 3= highly valuable**)

Aspect	Please number according to the perceived value		Please provide details on how and why
	Involvement	Cost contribution	
Data definition			
Date collection			
Data transformation			
Data exploration			
Data dictionary			
Granularity			
Architectures			
Governance			
Other			

39) On a best effort basis, please include any monetary cost estimate you may be able to provide (% of operational costs) related to the implementation of an integrated reporting system for your institution.

10.8 **Section 7.** Push vs Pull approach

40) Would you prefer the future integrated reporting system to be based on:

a) A pull approach

b) A push approach

c) A mixed (pull and push) approach

d) Other

41) What would be the main advantages and disadvantages to consider or/and what would be the main challenges to consider and what would be the possible design options (from both a technical and process perspective)?

Approach	Costs	Benefits	Challenges	Design options/solutions
Pull				

Push

Mixed

10.9 **Section 8.** Governance

42) Could you please specify any legal obstacles you would see related to the following?

	Obstacles/Challenge	Possible solutions
Feasibility of the central data collection point		
Collection and access to granular data by authorities		
The responsibility of the reporting institutions for the reported data, if the transformation of granular data is conducted by authorities (i.e. not the reporting institution)		
Other. Please specify:		

43) Do you agree with the suggested coordination mechanism for data requests? Do you see any benefits or disadvantages in this approach?
44) Please specify how the agile coordination mechanism for coordination of data requests could be further simplified and how your proposed measures could enhance coordination and avoid data duplication?
45) According to the reporting classification proposal included in Annex I: are there further reporting criteria to be taken into account under Category 3 reporting?
46) According to the reporting classification proposal included in Annex I: Do you agree with the proposed approach for non-recurring type of data Category 4?
47) What solutions could the EBA investigate that would reduce costs?

Aspect	Proposed solutions for the EBA to investigate to reduce costs
Governance	

10.10   Section 9. Technology

48) Are you making use of RegTech for reporting purposes?
    If yes, please specify in which reporting process step:

a. Data definition	b. Data collection	c. Data transformation	d. Data exploration
Yes/no	Yes/no	Yes/no	Yes/no

If no, please explain:
    a.   Not fully developed or useful for my needs
    b.   Others...please explain

49) Which of the reporting process steps would benefit more from RegTech development?
    a.   Data definition
    b.   Data collection
    c.   Data transformation
    d.   Data exploration

50) Do you agree with the main obstacles highlighted in the discussion paper? Do you see any further challenges?

51) Would you be keen to invest in RegTech for integration of different types of data?
    How would you develop such a technology?
            a)   in-house or
            b)   via a service provider

52) How do you think RegTech can help in data integration?

53) Do you agree that data standardisation is the first necessary step for using RegTech?
            a)   Yes
            b)   No

11. Accompanying documents

11.1 Annex 1: Classification of reporting requests

1. Although most of the regulatory reporting is harmonised, currently institutions provide different types of datasets to many different authorities, in some cases the data is requested through standard reporting templates, but also through non-standardised reports.

2. In order to scope the data contained in the CDCP mentioned in Article 430c CRR, it is necessary to provide an analysis of the characteristics of the different reports requested by authorities. Inclusion of data requests to the CDCP should be defined only once there is more clarity on the operation of the CDCP with necessary considerations for implementation aspects.

3. The figure below provides a reporting classification based on different types reporting:

Figure 1.: Classification of reporting request

	1 EBA ITS / ECB statistics	**2** Other regulatory data request (National + ECB/SSM requirements)	**3** Recurring data request	Non-recurring data (EU and national level)	
	Data request coming from regulation at EU level	Data request coming from regulation at national level	Recurrent data request that are not coming from regulations* (EU and national level)	Could be classified as: Type A and Type B	
				Type A According certain characteristics**	**Type B** rest of data request
To be included in the common data dictionary	✔	✔ Through national extensions	✔ Through national extensions	✔ Ex - ante modelling (if possible), should be modelled if the information would be requested in the future – Through national extensions	✖
To be included in the central inventory/register	✔	✔	✔	✔	✖
To be included in the CDCP	✔	✔	✔	**(?)** To be defined once there is more clarity on the operation of the CDCP	✖
Clarifications and examples	i.e.: EBA reporting framework, ECB statistical regulation	i.e.: Finrep solo regulation)	* Characteristics to be considered: time and number of institutions i.e.: STE and similar, EBA stress test, others	**Characteristics of a Type A: • general characteristics based on time and number of institutions • certain characteristics to be defined in a case by case basis	i.e.: unstructured reports, internal reports

4. Category 1 reporting includes the data requests coming from regulation at EU level. Examples of types of reports included in this category could be the EBA reporting framework and the ECB statistical framework.

5. These could be considered as harmonised reports which should be defined in the common data dictionary and be included in a central data register or inventory of data. In addition, these reports should also be included in the CDCP.

6. Category 2 reporting includes data requests coming from regulation at national level and the SSM/SRB data requirements. Some examples could be FINREP solo reporting requested via the ECB regulation. In this case, the SSM is considered at the same level as national reporting as it only applies to those countries under the scope of the SSM.

7. In these cases, the data will have to be defined through national extensions of the common data dictionary, included in a central data register or inventory and collected through the CDCP.

8. Category 3 reporting includes recurring data requests, at EU and national level that are not coming from regulations. Some aspects could be taken into account in order to categorise a report as Category 3. For instance, a report could be considered as a recurrent data request when it has been requested of more than five institutions or when the recurrence of the data request is indefinite. Examples of Category 3 reports could be the SSM Short Term Exercise, the EBA stress test or similar reporting which is a recurrent exercise.

9. Category 4 reporting includes non-recurring reporting which could be classified as Non-recurring type A and non-recurring type B reporting.

   - **Non recurrent – Type A:** general characteristics based on number of institutions and time.

      o Is considered non-recurrent type A, when it is requested more than 3 times and it has a fixed end date.

      o Is considered as non-recurrent type A, when it has been requested of more than five institutions.

      o Certain characteristics will need to be defined on a case-by-case basis.

   - **Non recurrent – Type B:** rest of data requests

      o Is considered non-recurrent type B, when it is requested less than 3 times.

      o Is considered as non-recurrent type B, when it has been requested of less than five institutions.

      o Examples: unstructured reports, internal reports.

10. In addition, there are other reports that do not fully fit into the previous classification.

11. In general, a proposed approach could be to integrate in phases. Once the reporting requirements from Category 1 to Category 3 are integrated and with the experience gained through its integration, it would be possible to have greater knowledge of the functioning of the CDCP. Therefore, this would help not only to determine what characteristics should be defined to classify the non-recurrent reporting (type 4.A and type 4.B) but also to determine what their degree of integration should be within the CDCP.

11.2 Annex 2: Data dictionary requirements:

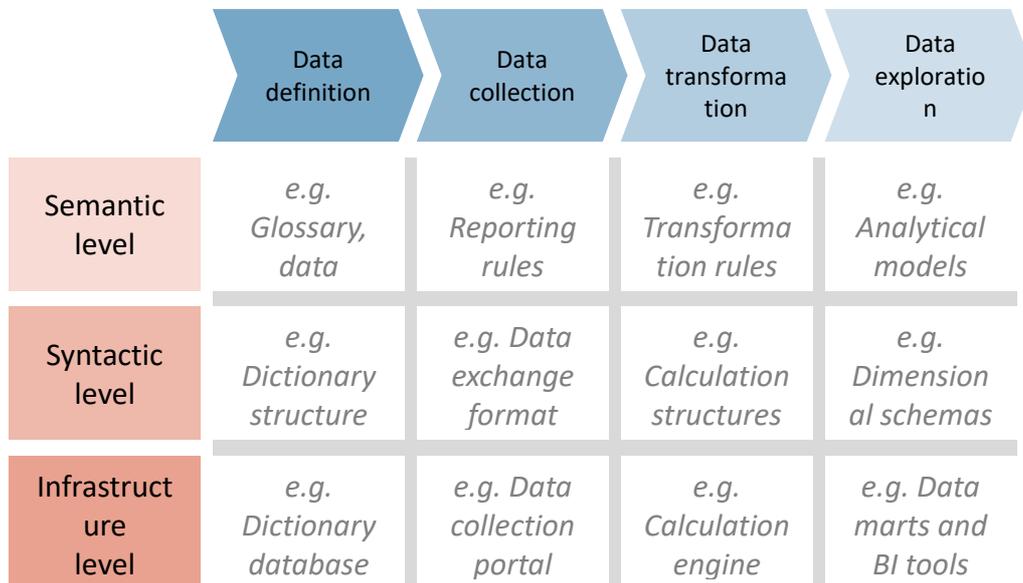## 11.3 Regulators and industry perspective

11.3.1 Overview

1. Reporting by institutions to authorities, be it in the area of supervision, resolution or statistics, provides these authorities and the regulators with crucial information on the compliance with regulatory requirements, and on the financial situation or risks and vulnerabilities. Originating from different sources and for different purposes, the reporting obligations are fragmented and a cause of inefficiencies and redundancies resulting in a burden on reporting institutions and unnecessary costs on authorities. The current situation also limits the overall level of benefits of data analysis which can be potentiated by improvements of data quality and data sharing.

2. The EBA has been assigned a mandate to carry out a feasibility study for an integrated reporting system in order to improve the benefits and reduce the costs of the current situation. More specifically the mandate addresses the feasibility of a central data collection point for the integrated reporting and the establishment of a standard dictionary of the data to be collected.

3. The EBA established a work stream with the supervision authorities and the ECB to analyse the characteristics and requirements of the data dictionary that could support the envisaged integrated system. This work stream also analysed the previous work regarding data dictionaries of integrated data collections in the ECB, the EBA, Bank of Italy, Bank of Spain and OeNB.

11.3.2 Data dictionary: holistic approach analysis

4. The discussion regarding the underlying requirements to such a dictionary covered the complete lifecycle of the data which includes: definition, collection, validation, transformation (including the calculation of derivable information), analysis, disclosure and dissemination of the data.

5. The integration requirements were analysed at three levels: semantic integration requirements, syntactic integration requirements and infrastructure integration.

Representation level of the reporting process chain

	Data definition	Data collection	Data transformation	Data exploration
Semantic level	*e.g. Glossary, data*	*e.g. Reporting rules*	*e.g. Transformation rules*	*e.g. Analytical models*
Syntactic level	*e.g. Dictionary structure*	*e.g. Data exchange format*	*e.g. Calculation structures*	*e.g. Dimensional schemas*
Infrastructure level	*e.g. Dictionary database*	*e.g. Data collection portal*	*e.g. Calculation engine*	*e.g. Data marts and BI tools*

6. At a semantic level the data dictionary addresses the text of business definitions of data, validation and calculation rules, having a limited value in terms of ability to use digital processes and contribute to the data collection, data transformation and data analysis effectiveness.

   Semantic integration is a huge effort that can be progressively achieved, by analysing one by one each reporting requirement and the different possibilities of integration. Due to its high diversity, it currently seems unrealistic to create a common and harmonised semantic data dictionary that would immediately encompass all reporting obligations. Instead, it may be preferable to start integrating those areas where convergence is most likely to be achieved.

   The prioritisation of the areas to be integrated should be decided by regulators bearing in mind the benefits to institutions, and knowing that the data dictionary can include all data concepts but not all transformations. In fact, some calculations are not feasible to implement as they imply extracting huge amounts of data and replicating huge parts of banks internal processes.

7. At a syntactic level, the data dictionary addresses the structure and formats of the different elements that enable the translation of semantic data definitions, validations and transformations. The formal and standard formats can then be used by digital processes, creating high value and effective integration support in data collection, data transformation and analysis.

   In the syntactic integration there are two different lines of work, first the preparation of the formal data standardisation mechanism with the dictionary structures and features; second the concrete syntactic translation of semantic data definitions and transformations. The first effort is feasible and has to be in place to be available from the starting of the new system and the first

integrated data, supporting the step by step approach on semantic integration. It represents an agreement on what are the formats and features of the data dictionary and other elements that will support an architecture for data integration. The second effort encompasses the progress on the incremental sematic integration.

8. At an infrastructural level the data dictionary addresses the concrete technologies and tools that will implement the integration across the several data processing steps.

11.3.3 The perspectives of regulators and the perspective of institutions

9. The analysis on data dictionary requirements was split into two different topics taking into consideration two different roles and perspectives: the regulators and authorities, and the reporting institutions.

10. Authorities have a specific data process chain and their perspective of what should be the ideal system should take into consideration the reporting institutions' data process chains and specificities, related to integration with their internal systems.

11.3.4 The data dictionary definition

11. A dictionary provides metadata about data elements. The metadata included in a dictionary can assist in defining the scope and characteristics of data elements, as well as the rules for their use and application.

12. The essential components of a data dictionary are:

Dictionary of vocabulary and concepts useful to describe the collected phenomena. It is a collection of names, definitions and attributes (i.e. domains, members, variables) about data elements that are being used or captured in a system. It describes the meanings and purposes of data elements within a certain context, and the links between concepts providing guidance on interpretation, accepted meanings and representation. As examples of other components of a dictionary there are hierarchies, dataset structures rendering, report grouping, specific mappings to exchange standards.

Dictionary of validations and transformations useful to describe data requirements and support specific data processes such as the dictionary of validations with its links and restrictions (i.e. hierarchies) and the dictionary of transformations that describe algorithms and processes that transform data into other data (i.e. aggregations or other calculations) are also included.

13. Dictionaries are useful for a number of reasons. In short, they:

- assist in avoiding data inconsistencies across a data universe;
- help define conventions that are to be used across one data universe;
- provide consistency in the collection and use of data across multiple organisations and people;
- make data easier to analyse;

- enforce the use of data standards.

11.3.5 The authorities' perspective

14. Despite the integration harmonised EU frameworks, it is a fact that for reasons of prudential, resolution and statistical reporting the requests for financial data can overlap for the same institutions. When considering possible national discretions and ad hoc collections in any of the three areas- prudential, resolution and statistical data- the overlapping burden can increase significantly.

15. Prudential and resolution frameworks integrate the data using standard data dictionaries that make it possible to identify and compare all data collected.

16. The semantic and syntactic integration of concepts is best achieved when authorities design reporting requirements. The effort should occur the very first moments of the regulatory definition. In fact, as the authorities have different reasons to collect data, it is easy to require different data on the same original data from reporting institutions.

17. In order to integrate data and avoid regulatory data overlaps it is necessary to have a central common data dictionary that should be:

- **Comprehensive** with all semantic integrated concepts available to any regulators to know if data was already requested in other frameworks. Integration of data is only feasible if the data is referred to in the data dictionary.

- **Complete:** have all the expected data elements that enable digital processing of reported data along the different processes of the data chain.

- **Centrally managed:** ensures the technical standardisation which creates an effective data consistency and data comparability.

- **Centred in a common and unique vocabulary:** integrated concepts are using only one common dictionary. Each inserted concept has a clear definition sharing the common formal vocabulary with no room for interpretation. Having to map concepts from different data dictionaries is costly and not adequate for the high dynamic evolution of financial regulation.

- **Ready for digital processing:** as the central piece of a data-driven system architecture, it should be available at a formal syntactic level with all elements necessary for system interoperability and communication.

- **Ready for human interface:** easy to be used and understandable by business users

18. The data dictionary should be able to identify and compare all data definitions providing the formal and standard categorisation, used by digital tools to analyse how similar the data requested by different regulators are. The identification of similar data is the first step towards

identifying data converging possibilities. Coordinated mechanisms should be in place to ensure further specific data convergence efforts between regulators.

19. The same common data dictionary should be able to define and identify different levels of granularity of the same original data.

20. Different regulatory frameworks can request the same data with different granularities. In this situation regulators should converge, agreeing on a unique granular requests and derive the more aggregated data. However, it is possible that for regulatory reasons the convergence is impossible to be achieved or be adopted immediately by frameworks already in place. The above implies the need to describe in the data dictionary the different aggregates with distinct levels of granularity of the same original data.

21. Even if common granular reporting is possible the data dictionary should be able to store different granularity levels, not only for the data definitions of the granular data but also the definitions of the calculated regulatory requirements. In this case, the data dictionary should incorporate the definitions of the necessary transformation from granular to calculated data.

22. The same central and unique data dictionary should be able to identify and define different types of breakdowns. Different regulatory needs can require different data breakdowns. In this situation, regulators should try to converge and harmonise breakdowns and simplify the reporting obligations. However, breakdowns harmonisation is not always possible and different data aggregations on the same data can coexist as different reporting obligations under the scope of integration.

23. The common data dictionary should include the validation that define the criteria for the quality of data received within the data collection process. The definition of the quality criteria and the relationship between reported values are closely linked with the definition of characteristics of reported concepts.

    The validation rules have their own lifecycle and should be considered as an autonomous piece in the data dictionary. It has to be consistent and evolve with the dictionary, but it changes much more often than the rest of the dictionary's contents. Keeping the necessary independence will ensure greater flexibility and more stability on the other data dictionary parts.

24. The data dictionary should include the common data transformation definitions to be shared by authorities on:

- data enriching processes usually linked with the data analysis and the need to create new derived data;

- data calculating processes linked with the regulatory data preparation and the need to derive regulatory data from more granular reported data.

25. The common data dictionary enables two types of data integration:

- the semantic definition of data concepts using the same shared vocabulary;

- the transformation definition using formulae to relate two different data concepts.

26. The common dictionary has to encompass the different change and versioning requirements of its different frameworks. Supervision reporting is one of the more demanding frameworks with higher frequency of changes. These changes are dictated by external triggers and imply highly flexible and extensible data dictionary capabilities to incorporate new definitions, validations and calculations and support the need to shorten the time of reporting implementations.

27. In supervision more granular reporting is not a guarantee that reporting requirements are stable and that future needs can be satisfied without additional costs. The experience shows that with each new framework the institutions have to select new data or classify data accordingly with the specificities and goals of the new framework. Even when data required is granular and two frameworks already agreed on a common granular collection, new future requirements can still involve new data, new classification of the same data or even lower levels of granularity that will imply costs on institutions side.

28. However, in some cases when a new framework doesn't require new original data, new classification of the same data or lower granularity, then the more granular reporting can enable the calculation of a larger amount of other new derivate data without implying that institutions have extra costs.

29. On reporting integration, the common data dictionary defines **what** are the data concepts and their definition. It needs to be complemented by:

- a common master data system defining **which** are the institutions and groups the data is referring to;

- the calendar of reporting obligations defining **when** the reported values are expected.

30. The data dictionary should enable national extensions providing a transparent and complete view of reporting obligations making it possible to identify and avoid data requests from overlapping.

31. The data dictionary should enable ad hoc request extensions providing a transparent and complete view of reporting obligations, making possible to identify and avoid any data requests from overlapping. Ad hoc data is usually requested with short notice, which can demand a two-step approach with a previous agile data definition and a later step with more complete dictionary definition.

32. At syntactic level the dictionary should have the following general requirements:

- High frequency of change of data concepts in terms of scope extensions and data definitions updates, encompassing regulatory work and financial system change environment where a permanent need to introduce new data definitions and changing the existing ones is expected.

- Fast and easy support to shorten the processes of developing and updating the data dictionary

- Comparability and consistency of concepts definition in a context of integration of different frameworks.

- Unequivocal identification of the elements included in the data dictionary

- Easy and clear access of different types of users with powerful browsing and self-explained data definitions and relationships.

- Access rights to dictionary contents allowing different roles: only read, read and write permissions.

- The dictionary has to be referenced by from and to dates and make data versioning possible.

- It should be able to represent the relationships that are particularly necessary for granular data and support the definition of aggregated data requirements.

- It should be agnostic to any data exchange formats (SDMX, XBRL, CSV, etc.) and suitable for representing all data elements and their relationships.

11.3.6 The industry perspective: Stability in a high change environment

Too many amendments

33. Having access to a dictionary enables access to data already requested and reduces new reporting requirements. This will be particularly relevant for assuring that national and ad hoc regulatory reporting are not redundant and are not a source of reporting burden.

34. Going granular can introduce more stability on requirements if new requests reuse the same data and nothing new is required. There are two types of new reporting requirements: a) New concepts are needed implying getting new data from institutions systems or new criteria for data classification; b) Nothing new is asked and new data aggregates can be obtained without asking institutions for more data. Granularity only increases stability in the case of b).

35. Templates/tables normalisation and atomic items. Normalised templates/tables can introduce more stability on requirements, because it will isolate data requirements and will make it possible to change only smaller parts of data requirements. Use of atomic items instead of complex composite items can introduce clarity and facilitate the change, because it will isolate what is new. Describing the relationship between atomic and composite items should be part of the data dictionary, however, it doesn't ensure the stability reporting requirements if new data is required.

Long term analysis

36. Financial systems are characterised by fast evolution which impacts the accelerated rhythm of regulatory amendments. A data dictionary should manage the time evolution of all the relevant content elements in order to record and keep track of data evolution and make it possible to analyse data evolution and its data disruptions (historical data / versioning of data).

37. Time management is a must on data integration and a very demanding requirement in supervision (prudential/resolution) where data is changing faster and more frequently when compared with statistical areas. Both areas have time versioning requirements and need to ensure long-term analysis, but they differ and have a different impact and reasoning.

38. Institutions will benefit from having a common dictionary with all regulatory data referenced with the same unique vocabulary. The dictionary has to ensure easy and fast incorporation of new reporting obligations in a consistent and integrated way.

Lack of predictability

39. Add dictionary definitions to regulatory text within public consultations. In supervision the regular regulatory frameworks go through public consultation processes which usual run for more than one or two years prior. As a dictionary can better explain the requirements, it would be advisable to include its definitions together with the regulatory text in public consultations. This can be applicable as a principle to all data requirements.

Short implementation period

40. Each institution manages data differently having different internal aggregations depending on their size and organisation, but a data dictionary can help institutions shorten their implementation costs by having:

41. A common data dictionary and common vocabulary enabling data comparability. Data comparability is a very important instrument to avoid data redundancy and to help institutions to be compliant with regulatory reporting. When data is defined using the same vocabulary the identities or similarities of different data concepts are visible. This helps institutions to map the data reporting obligations with their internal systems and reduces their costs of reporting compliance. The definition of a common data dictionary and vocabulary for all regulatory obligations is not controversial and a very important step to facilitate institutions' reporting compliance.

42. The decision to implement a common granular reporting to integrate reporting frameworks implies not only a common data dictionary but also a common calculation dictionary to achieve any regulatory aggregated or calculated data. This way it will be possible to have reporting formats closer to institutions' internal systems and deliver regulatory aggregation/calculation rules in a clear and transparent manner.

43. It is reasonable to believe that the alignment of different reporting frameworks would decrease the costs of reporting compliance, but some relevant aspects should be taken into account:

- The definition of the common granular input data dictionary, fitting all the possible institutions internal systems is difficult to achieve as each institution manages data on a different level of granularity and has different internal aggregations depending on their size and organisation. Some institutions are expected to be happy with their existing systems and they would have costs of changing to a different granular setup which would very unlikely be better than the one they already have. Other institutions would prefer to map their systems with a more granular reporting framework.

- A significant part of the reporting obligations cannot be reported at more granular level as the institutions have to reconcile data running iteratively internal processes of consolidation and adjustments that are impossible to replicate outside their systems. Institutions have different processes, mixing data collected from their operational and risk analysis systems which depend on of their specific strategies and managing capabilities. This way, different levels of granularity for the same data can coexist in the same data dictionary even if it is impossible to define a formal and precise formula to link them.

- Even when a common granular reporting is possible, institutions are still responsible for their ultimate regulatory compliance and they are obliged to calculate (or at least to confirm) the calculated data. This way it is not obvious that institutions will reduce their overall costs. In fact, they can have increased costs on adapting their systems to the new granular reporting,

and they will continue to support costs to calculate or to confirm the calculated data. It is very unlikely that institutions would accept extra costs unless going granular brings clear benefits increasing transparency and inhibited duplication of data requests.

Under these terms, the cost of implementing a more granular reporting should be investigated and compared with the current costs, taking into account the implementation costs and the maintenance costs.

- Sharing granular reporting implies aligning reporting frequencies, reference dates, statistical estimation methods and supervisory consolidation/reconciliation mechanisms.

44. The decision on going towards a granular approach should be based on a case by case cost-benefit analysis that can evaluate the impacts on all stakeholders involved: institutions, regulators, supervisors and others. A more granular approach would imply extra implementation and data maintenance costs, which need to be justified and clearly identified. A correct evaluation of the benefits is absolutely necessary in order to justify and distribute the inherent costs.

45. **Common data dictionary for national level reporting and ad hoc reporting,** integrating all the data requested in order to avoid duplicated data requests. Only by registering and defining all data required would it be possible to know if a new request was already reported in the past. In fact, a pre-condition of the reporting once principle is that all the reported data should be integrated into the Data Dictionary.

46. The European prudential/resolution framework is currently syntactically integrated and all the reporting data concepts are identified and distinct at dictionary level. Besides, this assurance on no redundant requirements, in few isolated cases some duplications were intentionally implemented:

- The same data point might be requested in different reporting modules (i.e. Resolution vs Own Funds)
- A data point that is clearly (or by design) the sum of other data points (i.e. Template 4 FinRep)

47. Furthermore, it would be necessary to extend the integration scope to other eventual sources of reporting duplications enlarging the scope to the ESCB statistical world and to the national reporting and the ad hoc data requests.

48. Managing calculation metadata makes it possible to adapt calculations to the specificity of data reported and achieve the correct calculated figures. Calculated figures can evolve in the following circumstances:

- Resubmissions of granular reporting implies recalculating the figures, which are currently processed by institutions in their internal reconciliation processes.

- Changes of calculation rules. Like validation rules in reporting, the calculation rules can change if new input data makes evident the need to adapt the rule.

Lack of quality

49. Reporting quality is achieved by a sound syntactic model and by defining a special set of validation rules that restrict the conditions of the values reported. Validation rules are of diverse types depending on their nature (i.e. signs, sums, values restrictions) and can include data calculations. Data quality processes are strongly related to the data definition processes but their evolution has a distinct lifecycle and depends very much on the time schedule and data reporting processes.

50. Validation rules' metadata should be part a of the data dictionary and should reflect the common quality criteria available for all stakeholders involved: institutions and all authorities. A data dictionary with clear and universal validation rules is essential to improve the data quality.

51. In case of granular reporting, the regulators and supervisors should ensure the quality of data reported by defining and publishing the validation rules for the granular reporting data, the calculation rules to transform reporting data in regulatory data and the validation rules on calculated data. These three levels of data quality definition should be in the common data dictionary and transparently shared with reporting institutions, which by remaining responsible for data quality of the prudential and resolution derived data, have to ensure both the quality of the granular input data and the resulting calculated data.

52. The set of validation and calculation rules, defined via the same language and agreed by authorities should be transparent and available to institutions at the same time that regulatory data definitions are communicated. This will enable institutions to use the same harmonised criteria for data quality and have time to prepare their data checks processes that will run on data before reporting submissions.

53. Depending on the type of validation rule, these can be manually defined or automatically derived from data dictionary. The validation rules should have different levels of severity depending on their relevance and universal validity.

54. The common set of validation and calculation rules agreed by authorities can be extended with an additional set of validations and calculations in order to further deepen the quality of the data and its relationships.

Clarity and complexity

Unclear and non-consistent requirements

55. Unclear regulatory requirements are related to the lack of a common and clear vocabulary and the existence of a standardised data dictionary. A consistent common vocabulary shared and applied consistently to the different data concepts, is imperative to have clear requirements.

Requirements give room for interpretation

56. Clarity of complex requirements can be improved by using the dictionary at early stages of regulatory definition. The data dictionary can provide a formal language that enables regulators to reuse concepts and create regulatory text that is closer to machine readable regulation. The use of a formal approach should cover data definitions, validation rules and calculation definitions, improving the precision and lowering room for interpretation.

11.3.7 Alignment of concepts

Non-harmonised concepts

57. Concepts can only be harmonised if defined using the common vocabulary and integrated into a common data dictionary. Apart from the inclusion of regular harmonised regulation, the national discretions and ad hoc data have to be integrated and included in the dictionary to be possible to investigate any redundant requests.

58. A common vocabulary, shared and applied consistently along the different reporting frameworks make the integrated concepts comparable and harmonised.

59. Data harmonisation doesn't require data granularity. Data concept alignment can be defined and implemented on any level of granularity, using only the dictionary common definitions or completing with calculation rules mapping data with different levels of granularity.

Overlap of concepts

60. The best way to reduce costs and avoid overlap of concepts is to improve data comparability. By using a common dictionary, with a common vocabulary, authorities and institutions know how similar or different the concepts are. Regulators should integrate most of their regulatory requests and compare data and converge avoiding unnecessary differences that can be burdensome for institutions.

61. In case regulators cannot converge, institutions can also compare concepts and know how similar or different they are and facilitate the data preparation.

Alignment of reporting standards.

Different reporting formats

62. The dictionary should be agnostic to any technology or related reporting exchange formats, being the common platform that can be translated and used in any possible exchange format. The exchange mechanisms should be reviewed to reduce the burden on institutions and data dictionaries should cope with any exchange.

11.3.8 Spread and quantity of reporting requirements

Spread of requirements along legislative text

63. The dictionary has the necessary detailed data in order to provide institutions with a comprehensive and consolidated understanding of all their compliance obligations. This information can be the concrete calendar of all reporting obligations of a certain institution in different integrated frameworks or a list of compliance obligations of a generic type of institutions with a certain set of activities and characteristics.

11.3.9 Transparency and returned value

No returned value for reporting institutions

64. A common data dictionary is a fundamental tool to support a common language to be shared by different stakeholders along the different processes in the regulatory data lifecycle: reporting, validation, calculation, analysis and data disclosure processes.

65. The dictionary in an integrated central system can vastly improve data analysis by providing the reference and the meaning necessary in order to produce the new data and the valuable benchmarks for the different stakeholders interested.

11.4 Annex 3: Integration experiences at national level

11.4.1 Integration experience in Austria

Competent Authority	OeNB / FMA
Institutions covered by the reporting model:	495 institutions
Types of report covered:	Supervisory, Statistical, Resolution reporting
Status of the reporting model	Fully in use since 2015
Reporting process phases covered	Data definition, data collection, data transformation
Type of collaboration	Public-private development

OeNB's reporting data model

The OeNB's reporting data model aims to give a complete, single description of the contents of the reporting data and in this way to minimise the room for interpretation. The objective of the OeNB's reporting data model is to formally describe the reporting data flow starting from the core banking system to primary reporting to the OeNB. To this end, the data model features a granular entity-relationship model (ER model) as a central element that captures all information needed to fulfil reporting requirements. Through this data model, semantic integration has been achieved.

This model, which is referred to as the basic cube, was developed jointly by the OeNB and Austrian banks. The OeNB's reporting data model also comprises algorithms in a formal pseudo code that enrich the basic cube and generate the following primary reporting frameworks:

1) Integrated reporting frameworks (smart cubes). Smart cubes are multidimensional reporting frameworks that use data collected by the OeNB to generate various secondary statistics. The description of these reporting frameworks forms part of the OeNB's reporting data model. This is how data integration on the collection level has been achieved.

2) Supervisory and resolution reporting requirements such as those of the EBA. The OeNB collects these data in the form of data templates.

A joint reporting software used by more than 90% of the Austrian banking market, which is a 1:1 IT implementation of OeNB's reporting data model, is running on a joint infrastructure. This infrastructure is operated and maintained by AuRep (Austrian reporting Services) that is a subsidiary of 90% of the Austrian banks. OeNB does not hold a share in the AuRep.

Data dictionary and semantic integration

Semantic and metadata integration has been achieved by establishing OeNB's reporting data model with its formal description of data, by creating a standardised input layer with a single, complete description of reporting data that is redundancy free (Basic Cube) and by using harmonised definitions and transformation/validation rules in a formal language. Currently there is an ongoing project that aims to formalise the reporting documentation of the data model (in the form of a metadata model). Infrastructure integration has been achieved as in Austria the OeNB acts as a single data collection point for statistical, supervisory and resolution reporting of institutions.

Centralised/decentralised model

Integration has been achieved through collaboration (e.g. workshops, ongoing discussions) with all involved, stakeholders such as commercial banks and experts within the competent authority and the central bank. The achievements of the discussions result in the data model documentation.

OeNB acts as a centralised single data collection point for statistical, supervisory and resolution reporting. The processing of all mentioned reporting requirements is performed within a single integrated IT system landscape.

Benefits achieved

Integration of reporting requirements (collection level):
* Collect data only once: reduce redundancies
* Multi use of data
* Granularity: reduce ad hoc requests
Integration of data definitions:
* Ensures consistency of collected data
* Better data quality through more formal reporting instructions and data standardisation

Areas not integrated

On the level of data definition there has been a large degree of data integration achieved and there is a work plan of further integration of additional requirements e.g. AnaCredit, Liability data report and FINREP solo are fully integrated, FINREP consolidated is work in progress, COREP credit risk is planned. On the level of data collection statistical requirements of the ECB are collected based on integrated datasets (cubes). Supervisory and resolution data are collected in the form of data templates (no integration on collection level).

OeNB's data reporting model does not contain consolidation rules and algorithms for computation of risk parameters (e.g. risk-weighted asset (RWA))

11.4.2 Integration experience in France

Competent Authority	ACPR)
Institutions covered by the reporting model:	722 entities (of which 351 banks)
Types of report covered:	Supervisory, Statistical reporting
Status of the reporting model	Fully in use since 2010
Reporting process phases covered	Data definition, data collection
Type of collaboration	Public development

SURFI reporting data model

The French integrated reporting covers European quarterly statistical data and a part of national supervisory data. The methodology used is based on national GAAP definitions. These definitions comply with the principles set up by the European system of accounts (ESA 2010).
This current national reporting is called SURFI and will be replaced by a new taxonomy named RUBA (unified reporting for banks and equivalent entities) from the reference date of 31st January 2022.

Data dictionary and semantic integration

The dictionary and the metadata are part of one single XBRL taxonomy. The taxonomy is based on a nomenclature of items accompanied by detailed definitions. There are about 3,000 different items, that are reported only once in the relevant entry point. Items used in integrated templates encompass regulatory and statistical references. The nomenclature includes specific items when necessary for supervisory or statistical purposes. This dictionary is currently being recast in order to adopt the DPM format. One single DPM will be used for supervisory and statistical data. This adaptation to the DPM format is made in the framework of the replacement of SURFI by RUBA in 2022.

Centralised/decentralised model

A close cooperation between Directorate of statistics and ACPR enables the definition of common templates. These templates are submitted only once by the institutions to a centralised portal, and are managed in a centralised application. In addition, the data are disseminated to two databases: supervisory database in ACPR and statistical database in Banque de France.

Benefits achieved

A close cooperation between Directorate of statistics and ACPR enables the definition of common templates. These templates are submitted only once by the institutions to a centralised portal, and are managed in a centralised application. In addition, the data are disseminated to two databases: supervisory database in ACPR and statistical database in Banque de France.

Areas not integrated

Prudential requirements defined in the CRR could not be integrated, due to difficulties in reconciling prudential and accounting definitions of data. Moreover in COREP the classification of products and counterparties could be tailored for the risk approach, which is not convenient for statistical or accounting purposes.

There is no central tool for data transformation and data exploration between prudential and statistical data, which are based respectively on the tools BASETU and ROSTAM/OKAPI.

11.4.3 Integration experience in Hungary

Competent Authority	Central Bank of Hungary
Institutions covered by the reporting model:	40 entities
Types of report covered:	Supervisory, Statistical, Resolution reporting
Status of the reporting model	Fully in use since 2020
Reporting process phases covered	Data collection, data transformation
Type of collaboration	Public-private development

Integrated Data Collection System

According to Act CXXXIX of 2013 on the Magyar Nemzeti Bank (MNB, the central bank of Hungary), from 1 October 2013, the MNB has been responsible for the supervision of financial institutions and financial consumer protection. (The supervision of the banking system was previously the responsibility of the Hungarian Financial Supervisory Authority (HFSA).) In addition to the supervisory powers, the authority's functions related to financial resolution have endowed the MNB with a wide range of competence. The coordination of macro and microprudential regulations as well as the creation of the resolution function in its unique way made it possible for the central bank to assess and effectively manage the risks of the financial system. Then the MNB has created a project aimed at realising an integrated statistical, supervisory and resolution data collection system and new data warehouse for storing the data. The new system with a wide range of the integrated functions was launched in 2020. Data exploration systems are still different in all directorates, but for the publication of the statistical and supervisory data we use the same statistical application supplemented by the Microsoft Power BI tool.

Data dictionary and semantic integration

Semantic and metadata integration has not been achieved from the metadata perspective but it has been achieved from the data receiving channel, data storing and data publication perspectives. We wanted to avoid data duplication from the statistical and supervisory perspectives, but we have come to the conclusion that it is necessary to maintain many separate statistical and supervisory data requests. There are clear methodological differences which hinder the integration of data sources.

Centralised/decentralised model

MNB has a centralised infrastructure model, with one single data collection channel for all types of reports to be submitted to MNB by the credit institutions. (EBA XBRL taxonomy reports and national reports are also received on the same - integrated - channel.) MNB has created a project aimed at realising an integrated statistical, supervisory and resolution data collection system and new data warehouse for storing the data. The new system with a wide range of the integrated functions was launched in 2020.

Benefits achieved

Integrated Data Collection System and integrated analytical database for all types of reports results in more efficient data quality control, possibility for defining automated validation rules between different types of reports. The same data points can be used for different analytical purposes from the integrated database.

Areas not integrated

The semantic layer could not be integrated; there is no integrated data point model for all (EU and national) types of reports.

11.4.4 Integration experience in Italy

Competent Authority	Banca d'Italia
Institutions covered by the reporting model:	392 entities
Types of report covered:	Supervisory, Statistical, Resolution reporting
Status of the reporting model	Fully in use since 2009
Reporting process phases covered	Data definition, data collection, data transformation, data exploration
Type of collaboration	Public development

Infostat system

In general, the solution covers all aspects of data lifecycle process. Definition of reporting requirements, collection, transformation and exploration. The methodology used is a multidimensional generalised model where the main components are artefacts called cubes. An observed phenomenon described with a cube may be the reporting requirement defined by a matrix representation, an elementary cube. A cube describes also aggregation between elementary cubes. Finally, cubes are suitable to describe archives (tables in a relational database) that are groups of cubes (cubesets). Each cube is formally described with a cube structure where every phenomena is described by its variables and sets.

Data dictionary and semantic integration

At national level the integrated approach is the result of the synergy of this elements:
- a coordination body that ensures the consistency of the statistical and supervisory reporting frameworks (statistics committee)
- a single dictionary that documents the whole data management and definition process (Matrix metamodel)
- a directorate dedicated to the entire data lifecycle process (metadata definition, data collection, data quality assurance, data transformation)
The Statistics committee is an institutional body with coordination and direction functions concerning Bank of Italy's statistical function that is connected with Bank of Italy's institutional duties. The Statistics Committee has a key role on coordination, analysing the impact of new reporting requirements with the goal of minimising the reporting burden. All information requests are filtered by the Statistic Committee. In this way, the body ensures that there are no duplications on information requests.
All national reporting requirements are then coded in the data dictionary according to the matrix model. If a concept (variable, element) already exists, it is reused in a new requirement without changing its codification; new concepts instead are coded by a specialised team in the Statistics Directorate (a 'meta-data and tranformations' management division) that ensures that there is no duplication of concepts.

Centralised/decentralised model

The process of integration in Banca d'Italia started in the mid '90s using a single data dictionary and a single metamodel, gradually migrating all data collections that were collected using different platforms to a single platform called PRISMA, all under the same data dictionary definitions. After PRISMA Banca d'Italia migrated to a more modern infrastructure called Infostat using the same dictionary and metamodel.

Benefits achieved

The tools developed for the statistical treatments are highly independent from the nature of the data to be processed (quantitative/qualitative, supervisory/statistical/resolution, banks/other intermediaries, universe/sample, etc.) that there's a significant leverage on the investment made. This makes it possible to reuse the same tools for different reporting frameworks. In addition the widespread use of the same tools also allows an optimal and flexible allocation of human resources dedicated to the statistical processes (i.e. given that the tools are the same, an employee in statistical data collection can work also in supervisory data collection without additional training).

Areas not integrated

In order to integrate at least formally a reporting requirement a necessary condition is that they shall be structured (formally defined like the DPM or the SDMX), therefore is not possible to integrate ad hoc data collections that are generally unstable and not formally structured.
Banca d'Italia has been capable of collecting, exploring and transforming every type of structured reporting requirement (national and European) exploiting the cube methodology, although the different definitions and codification systems used by the European authorities make the national and European data collections not comparable between each other, leading to a missing semantic integration of 'foreign' statistical and supervisory concepts. This means that all structured reporting requirements (national and European) are formally in the same dictionary, but only the national portion is fully integrated.

11.4.5 Integration experience in Lithuania

Competent Authority	Bank of Lithuania
Institutions covered by the reporting model:	10 banks + 7 foreign bank branches
Types of report covered:	Supervisory, Statistical reporting
Status of the reporting model	Fully in use since 2015
Reporting process phases covered	Data collection
Type of collaboration	Public development

CRD IV Reporting Solution (Banking sector)

The reporting system for statistical and supervisory purposes has not yet integrated,. but to gather all data we use the same data collection system. Reporting solution covers integration of the EBA data model with Bank Reporting System

Data dictionary and semantic integration

On a semantic level, we created ETL process, which translates the EBA access database data into our internal Oracle Database. On a syntactic level, we fully reused the DPM metadata concepts.

Centralised/decentralised model

The integration was achieved by creating additional data layer (tables) which connects the DPM metadata with internal reporting system (database) concepts. This data layer consists of additional new tables that can be presented as an information gate for client applications. Client applications use API library for communication with reporting system database.

Benefits achieved

This solution lets us significantly decrease system maintenance time and effort. This solution helped us almost automatically to keep updates of clients' applications for each new DPM version.

Areas not integrated

Until the patterns of coding the DPM model have changed, we can keep our current solution without changes.

11.4.6 Integration experience in Norway

Competent Authority	Finanstilsynet – FSA Norway
Institutions covered by the reporting model:	212 entities
Types of report covered:	Supervisory, Statistical reporting
Status of the reporting model	Fully in use since 1986
Reporting process phases covered	Data definition, data collection, data transformation
Type of collaboration	Public development

ORBOF system

It is a long lasting cooperation between Statistics Norway, Norges Bank (the Central Bank) and Finanstilsynet.
The data collected are used for statistical purposes (National Accounts, Foreign Accounts, etc.) and some supervisory and analytical purposes, both on single banks and aggregated (interest rate statistics, accounting data, loans, loan losses, etc.). The data collection and the automated controls are performed by Statistics Norway before data is loaded into the database (from 1986!). More and more of the data are being extracted and processed in other data systems (SAS, SQL, Python and Excel)

Data dictionary and semantic integration

In 1986 Finanstilsynet and Statistics Norway started the cooperation, and the majority of data collected from credit institutions, were collected to this ORBOF system.

The system is a quite granular matrix system, where the intention was to collect information only once. That is all parts of the balance sheet, e.g. loans, type of loan, borrower sector, industry classification, currency (NOK or Foreign) in one single report. All accounting information could be linked to the banks' official quarterly and annual reports.

Centralised/decentralised model

The data collection is centralised in Statistics Norway, while Statistics Norway, Norges Bank and Finanstilsynet cooperate on the reporting instructions. The legal basis for the data collection is finanstilsynsloven (The FSA Act)

Benefits achieved

Lower reporting burdens for the banks. Better data quality. Less work for the governmental institutions (Statistics Norway, Norges Bank and Finanstilsynet).

This structure for data collection facilitates the institutes which only has to report this data to one authority. The same data will, according to this, only have to be reported once. This minimises or eliminates the risk of different definitions of the same data in different reports. Statistics Norway is responsible for the automated examination of data quality and the day-to-day contacts with the institutes. Finanstilsynet is the formal owner of the data, but the data is used by Statistics Norway, Norges Bank and Finanstilsynet.

Areas not integrated

The data is only on a non-consolidated level. (Finanstilsynet collects a few more data on a consolidated level, in addition to the FINREP data). Specialised prudential data is not included, since it is collected in the CRR reporting.

11.4.7 Integration experience in Portugal

Competent Authority	Banco Portugal
Institutions covered by the reporting model:	175 entities
Types of report covered:	Supervisory, Statistical reporting
Status of the reporting model	Fully in use since 2018
Reporting process phases covered	Data definition, data collection, data transformation, data exploration
Type of collaboration	Public development

CRC5G system

For this exercise we are presenting the Portuguese Central Credit Registers system covering data definition, data collection, data transformation and data exploration. Data definition is defined in the information dictionary containing both business concepts and data definitions. The financial credit institutions report the requested data to a centralised repository where the physical and logical processes run to validate and integrate data in the system. ETL procedures are applied to feed the analytical system where data is available to be analysed and used in several procedures.

Data dictionary and semantic integration

BdP uses one information dictionary of concepts, structured by domain and subdomain of business data. From the syntactic perspective we use one single metadata model. The information dictionary is a product customised for BdP.

Centralised/decentralised model

The integration was achieved with all the definitions, rules and technology organised in a centralised infrastructure model. The interaction with all users was crucial to ensure a single definition for each variable and to define the rules that should be applied.

Benefits achieved

The CCR system aims to be a single entry-point for credit data with a harmonisation of concepts, granularity, frequency, timeliness and reporting formats. This has the advantage of rationalising reports, is a relevant contribution for reducing the reporting burden of financial institutions and minimising efforts to ensure the coherence of credit data used for different purposes.

Areas not integrated

N/A

11.4.8 Integration experience in Sweden

Competent Authority	Finansinspektionen
Institutions covered by the reporting model:	All banks
Types of report covered:	Supervisory, statistical and resolution reporting
Status of the reporting model	Fully in use since 2014
Reporting process phases covered	Data collection
Type of collaboration	Public development

PI2 (Periodic reporting 2)

Finansinspektionen (The Swedish FSA) has a reporting system used for submission of statistical, resolution and prudential (supervisory) data.
Data are usually shared as it is reported (XBRL/XML/etc.) but aggregated data are shared in other formats.

Data dictionary and semantic integration

One example is that data which is used for statistical purposes is incorporated in national reported supervisory data.
Both definitions from the EBA DPM and another format agreed with other authorities are used.

Centralised/decentralised model

There is a cooperation developed for collection of data which is regulated in Swedish legislation or in a separate agreement between authorities. The Swedish FSA is the one authority to which data is reported and which has an agreement for sharing this data to authorities responsible for resolution and statistics.

Benefits achieved

This structure for data collection facilitates for the institutes which only has to report data to one authority. The same data will, according to this, only have to be reported once. This minimises or eliminates the risk of different definitions of the same data in different reports. The Swedish FSA is responsible for the examination of data quality, contacts with the institutes and support in all kinds of reporting questions etc. This single point of contact means that there is a clear and obvious way for the institutes regarding all matters of reporting.

Areas not integrated

N/A

11.4.9 Integration experience in Slovenia

Competent Authority	Banka Slovenije
Institutions covered by the reporting model:	16 (all banks, saving banks and branches, which in case of MS's parent banks are not obliged to report whole set of data))
Types of report covered:	Supervisory, statistical reporting
Status of the reporting model	Fully in use since 2011
Reporting process phases covered	Data definition, data collection, data transformation
Type of collaboration	Public development

Multipurpose reporting system in Bank of Slovenia (Porfi)

The reporting system includes:
- reporting of balance sheet and statement of comprehensive income items (including profit and loss items) for institutions on a solo basis, aligned with ITS FINREP, statistical requirements for statistics on Financial accounts and External statistics, data according to ECB AnaCredit.
- SHSG data, which is according to ECB's regulations on collection of detailed data on credits and credit risk and statistics on securities holdings of banking groups collected on consolidated basis from three banking groups only.
- macroprudential data regarding exposures to households for house purchase and consumer loans.
Data is collected using xml taxonomy.
All related data needs (such as those linked to the balance sheet) are organised in a two-dimensional table as a common reporting scheme, where rows represent reporting items and columns attributes using specified code list (for example countries, ESA sectors, currency, status of default and performance etc.). Reporting requests are defined in multidimensional matrixes and described in detailed instructions.

Data dictionary and semantic integration

From the data dictionary perspective the data integration is achieved on all levels – one document of instructions, 12 reports where items are reported using common attributes (variable, depending on the item) in the view of a matrix. The data reported by the banks in xml are transformed to the database in the Oracle data warehouse.

Centralised/decentralised model

Integration was achieved in a centralised infrastructure model.

Benefits achieved

Data is collected once for different purposes (AnaCredit, statistical, supervisory, macroprudential) using standardised definitions. There is no ex-post reconciliation among the datasets and no redundancies in statistical requirements. The production cycle is standardised. Reporting is organised in a single department

(statistical) and a system of automatic controls is developed. Collected data is used for various analyses and for several purposes, including ad hoc requests.

Areas not integrated

The EBA ITS is not integrated. The data in the reporting system is only on a solo level for FINREP items. The dimensions are not completely aligned with ITS due to the use of the reporting system for different purposes. The data is reported granularly by the banks and is aggregated by the NCA, which is different from the ITS, where data is reported on an aggregated level and the banks have to ensure the quality of the aggregated data.

The required data covers primarily statistical and financial data and not data from a risk perspective. It would be too complex to include ITS data (COREP, liquidity etc., level of consolidation) with the same granularity, i.e. data at the contract level.

11.4.10 Integration experience in Finland

Competent Authority	Finnish Financial Supervisory Authority (FIN-FSA)
Institutions covered by the reporting model:	256 institutions
Types of report covered:	Supervisory reporting
Status of the reporting model	Fully in use since 1996
Reporting process phases covered	Data definition, data collection, data transformation, data exploration
Type of collaboration	Public-private development

Riski

FIN-FSA have a reporting and analysis system, which has been developed for years. The core for data collection exists in the national data collections. Currently there are separate pipelines for the national data collections and the DPM data collection. For both, an Excel-based data collection tool is produced. The legacy format is csv which may be produced with Excel macros. The DPM/XBRL tool is an Excel AddIn. Encrypted reports are sent to FIN-FSA by email. Reporters receive automated feedback. The analysis system is renewed totally after the DPM-modelled data collections started. MS SSAS OLAP Cubes used to be core for analysis. Analysis database was introduced in 2019 and compilation needed analysis and dashboard production is shifting to that. Excel-based table view solution was also introduced in 2019 for the EBA and S2 data collections. Cubes still used for browsing purposes. FIN-FSA mainly uses MS-based technology, XWand is the XBRL engine. FIN-FSA has started a project for the renewal of a data collection system. Target is to have a modern portal for Reporters and get rid of Excel-based data collection tools and emailing of reports. The analysis part is not included in the project.

Data dictionary and semantic integration

The FIN-FSA uses the EBA DPM as a base for the database model. The EIOPA DPM is fitted to the same model, as is the European Securities and Markets Authority (ESMA) Alternative Investment Fund Managers Directive (AIFMD). The FIN-FSA DPM reporting database is updated with developer tools from the EBA and the EIOPA DPM publications. XWand is updated with taxonomy files and with a separate EFR component by Fujitsu. The DPM metadata is also used for analysis.

Centralised/decentralised model

Integration is an in-house production with the help of consultant IT work. Integration has a strong path dependence on Excel-based data collection templates and supportive tools.

Benefits achieved

There is an interoperability of data collection and analysis tools for the EBA, EIOPA and ESMA AIFMD.

Areas not integrated

There have been discussions on whether the Bank of Finland statistical and FIN-FSA supervisory data collections would be unified organisationally and technically. So far, this has not been feasible.