



16 December 2021

Annexes

On Final Report on Feasibility study on Integrated reporting system

Contents

List of figures	3
Annex 1: Feedback received on the Discussion Paper and main changes in the report	4
Annex 2: Links with other EU initiatives and projects	5
Annex 3: Stock-take on current data requirements	7
Harmonised reporting frameworks	8
Additional data requests for prudential purposes	9
Resolution data	12
Statistical data	15
Evidence from the Cost of Compliance study	16
Preliminary conclusions	17
Annex 4: On the different types of data dictionaries coming from the feedback received on the EBA Discussion Paper	18
Annex 5: Data dictionary requirements – Regulators and industry perspective	20
Annex 6: Reporting Frameworks – Current state of integration	33
Annex 7: The data dictionary – The framework differences	36
Annex 8: Experiences on European-wide data dictionaries	37
Annex 9: The going granular impacts on a common data dictionary	38
Annex 10: Data dictionary – Cost-benefit assessment	39
Annex 11: The granularity level of the reporting requirements	43
Annex 12: Granularity – Cost-benefit assessment	49
Annex 13: Tasks of the decision-making bodies (European authorities, national authorities)	51
Annex 14: Tasks where coordination, collaboration and/or joint work is needed and where the JRC could have a role	52
Annex 15: Best practices on governance ad hoc and national data request	53
Annex 16: On the push and pull mechanism	56

List of figures

Figure 1: Number of additional data requests to institutions by country (quantitative and qualitative)	10
Figure 2: Number of additional data requests by frequency (qualitative and quantitative)	11
Figure 3: Number of additional data requests per main risk areas (qualitative and quantitative)	12
Figure 4: Number of additional data requests by country	14
Figure 5: Granularity of the additional data requests	15
Figure 6: Representation of the reporting process chain	21
Figure 7: Granularity and data transformations	47
Figure 8: Coordination mechanism for data requests	55

Annex 1: Feedback received on the Discussion Paper and main changes in the report

- 1 The EBA discussion paper on integrated reporting published for public consultation between March and June 2021 was broadly welcomed by stakeholders who acknowledged the benefits of such a system (reusability, efficiency gains, increased understanding of reporting processes, common understanding across authorities) and encourages European authorities to progress in this arena for the contribution of a fully-fledged Banking Union.
- 2 The standardisation of data was seen as a necessary step towards an integrated reporting system and although it was noted by respondents that the integrated reporting would be difficult and that many resources would be required, it would result in efficiency gains and cost savings both for financial institutions and authorities in the medium-to-long term.
- 3 Most respondents supported the holistic approach used for the integration analysis of the reporting process chain. The main elements of the integrated reporting system were seen as covered although further detail was requested with regard to the Governance aspects of the system. Related to the common dictionary, the respondents generally support the design following the “define once principle” and acknowledges the difficulty of the endeavour as long as it requires the alignment of regulatory definitions.
- 4 In general, respondents welcomed the inclusion of key stakeholders’ perspectives and suggested continued engagement with other data transformation projects in order to share findings and lessons learnt, as a key input for the success of the integration project.
- 5 Cross-cutting comments concerned the challenges faced by institutions, particularly in terms of ad hoc requests. Respondents acknowledge that many such exercises arise from idiosyncratic reasons that are difficult to define upfront. However, respondents encourage further discipline on data requests as the integrated system should rather achieve harmonisation of definitions, smoother submission processes, and overall reduced complexity compared to the current framework.
- 6 Some proposals appeared to respondents as challenging but achievable over time, while some highlighted the need to overcome some concrete actions first in a proportionate and gradual manner. The setting up of the Joint Reporting Committee and work related to the data dictionary have been identified as prior steps before the implementation of any changes to the reporting architecture. The need for coordination and consistency between the various (EU and international) initiatives was also flagged.

- 7 The comments received have led to adjustments in the EBA's integration assessment. This includes providing more clarity on the objectives and actions pursued, more details on the timeline and proposals for actions to be taken after the feasibility study is completed.

Annex 2: Links with other EU initiatives and projects

- 8 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 institutions' compliance with both. In addition, the EBA has a mandate to continue enhancing proportionality and perform an analysis on the costs of reporting which has been provided in a cost-of-compliance report¹ 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 IReF², 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)³, 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 institutions' 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 SMCube Information Model⁴. The BIRD currently covers the reporting requirements of

¹ <https://www.eba.europa.eu/regulation-and-policy/supervisory-reporting/cost-compliance-supervisory-reporting>

² https://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/reporting/html/index.en.html#IReF

³ https://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/reporting/html/index.en.html#BIRD

⁴ https://www.ecb.europa.eu/stats/ecb_statistics/co-operation_and_standards/smcube/html/index.en.html

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 legislation⁵. 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 Strategy⁶ and the Digital Finance Strategy⁷.
 - 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. On 15th December the EC published the European supervisory data strategy⁸ which aims to provide the general guidelines to improve and modernise the EU supervisory reporting and put in place a system that delivers accurate, consistent, and timely data to supervisory authorities at EU and national level, while minimising the aggregate reporting burden for all parties.
 - 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.
- 9 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.

⁵ https://ec.europa.eu/info/publications/191107-fitness-check-supervisory-reporting_en

⁶ <https://ec.europa.eu/digital-single-market/en/european-strategy-data>

⁷ https://ec.europa.eu/info/publications/200924-digital-finance-proposals_en

⁸ Strategy on supervisory data in EU financial services', European Commission, COM (2021)798

Annex 3: Stocktake on current data requirements

- 10 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 reporting institutions.
- 11 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.
- 12 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.
- 13 This stocktake leverages firstly 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 databases⁹ are henceforth referred to as a '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 institutions in the jurisdiction, as the larger and more complex institutions tend to report more data. Additionally, the analysis may be based on differing methodologies, varying by jurisdiction. For some countries and frameworks, as for SSM countries supervisory data, 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 institutions.

⁹ 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.

However, in order to reduce the burden on the competent authorities, the EBA decided to use the available data as an approximation in this note.

- 14 The statistical requests are described using the information made available to the EBA by the ESCB¹⁰.

Harmonised reporting frameworks

- 15 An overview by the number of defined data points of the harmonised reporting frameworks can be found in Table 1. It should be noted that due to different ways of collecting data in different frameworks, the numbers are not always directly comparable.
- 16 The institutions' 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 tables¹¹. Thus, for example, Credit Risk Benchmarking and Large Exposures reporting comprises the majority of the reported values for some institutions due to the open data points concept.
- 17 In terms of the actual values reported in the EBA reporting frameworks, institutions on average¹² 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 institutions 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
Funding Plans	1,062
FINREP – COVID-19	600

¹⁰ 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?743bc2defe61abe865e1857ab1a98337>.

¹¹ 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.

¹² The average is referring to the EBA's reporting sample of institutions, available here: https://eba.europa.eu/sites/default/documents/files/document_library/882861/SCOP%202020%2023%20rev1%20%28Update%20of%20EBA%20List%20of%20Institutions%20for%20Supervisory%20Reporting%29.xlsx. The average takes into account periods from 2014 to 2019.

Remuneration	380
COREP Leverage Ratio	276
Benchmarking of internal models	153
COREP Large Exposures	77
G-SII identification and buffer rates	17
Total¹³	96,180

18 Prudential information is collected from EU supervised entities¹⁴ 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 institutions 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, institutions¹⁵ 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.

Additional data requests for prudential purposes

19 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 institutions.

20 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.

¹³ 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 some definitions (and thus data points) are reused across frameworks.

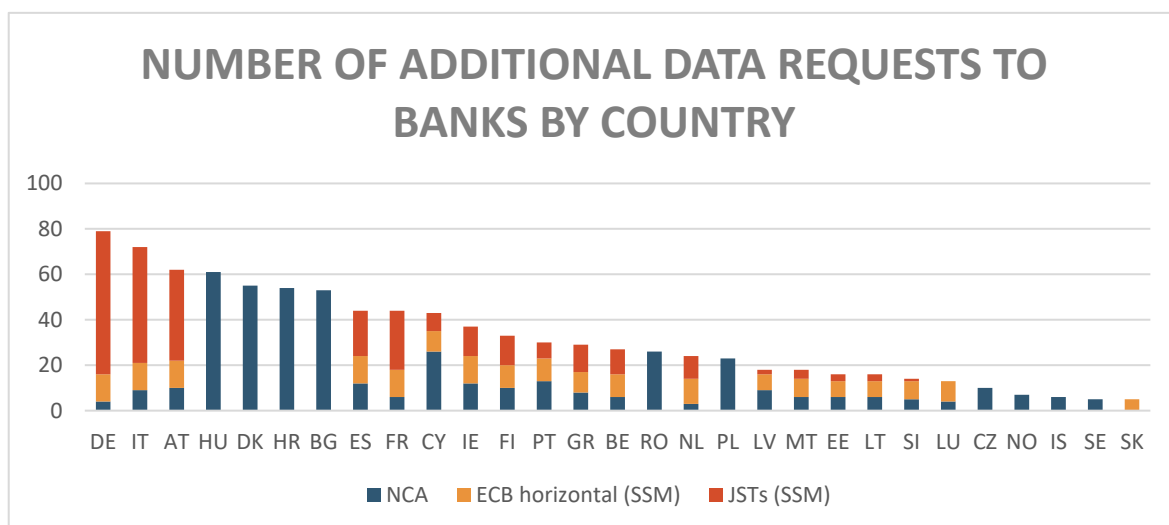
¹⁴ Credit institutions, financial and mixed financial holdings, banking groups and branches of EU and non-EU institutions.

¹⁵ 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.

Number of additional data requests for prudential purposes

- 21 When considering simply the number of additional data requests, most of them, roughly 59%, are of a quantitative nature. Only qualitative information is collected in 2% of the cases¹⁶, and the remaining 39% contains both types of data.
- 22 The inventory of data requests suggests 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 institutions in a jurisdiction requiring elevated prudential focus.

Figure 1: Number of additional data requests to institutions by country (quantitative and qualitative)¹⁷¹⁸



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

Frequency of additional requests

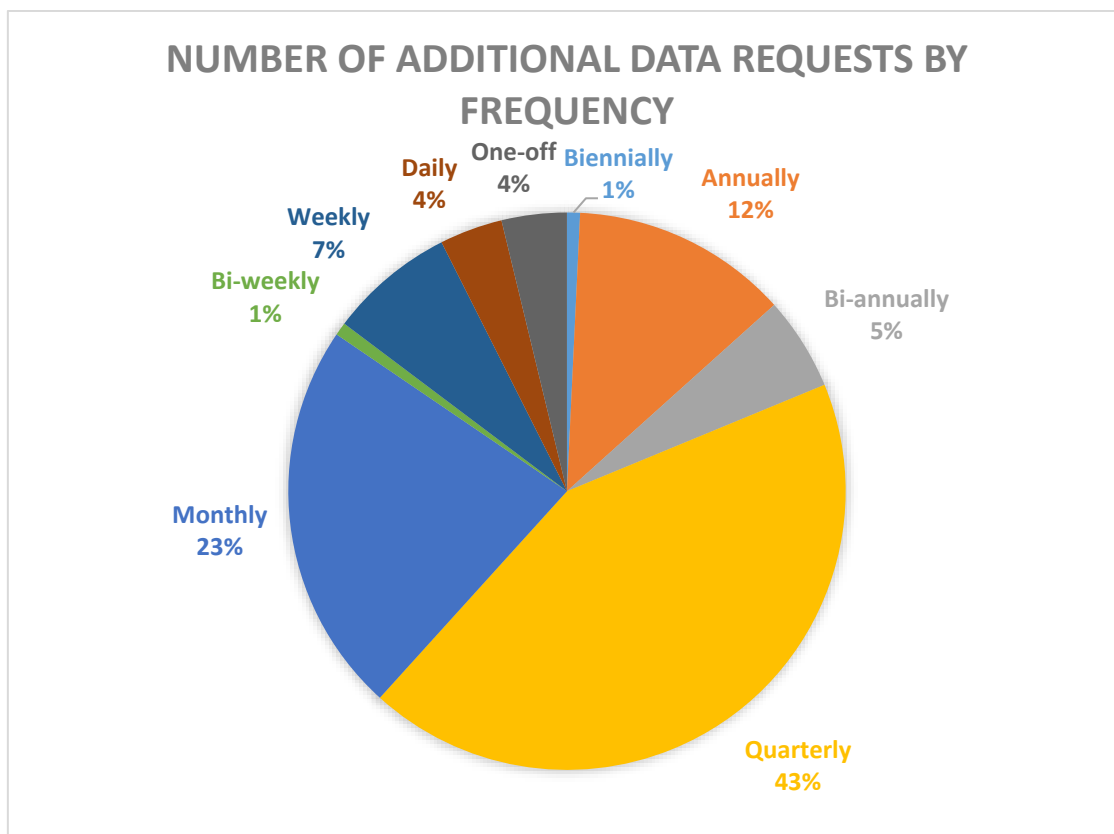
¹⁶ There may be overlap, as some requests refer both to quantitative and qualitative information. In such cases, the request is counted both as quantitative and qualitative.

¹⁷ 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.

¹⁸ 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.

23 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 2: 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.

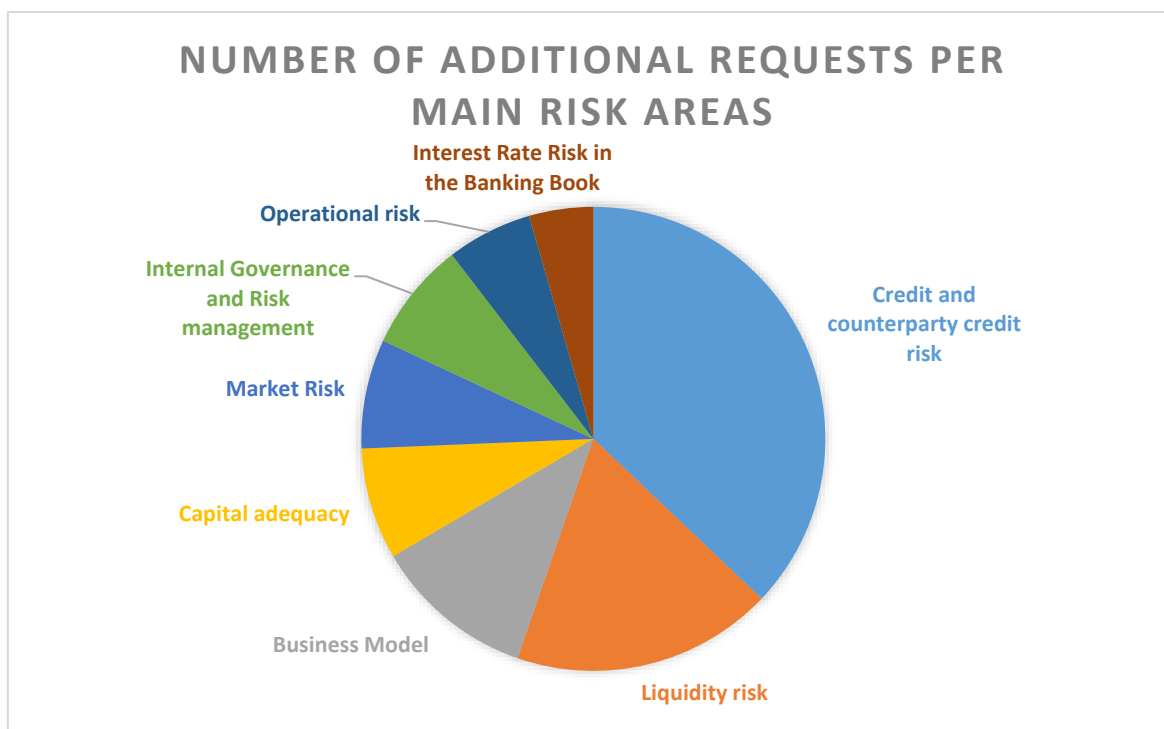
24 It is understood that, for specific purposes or entities, monitoring the situation with a higher frequency than what is specified in the prudential reporting might be necessary, e.g. monitoring the liquidity situation of a troubled institution. However, such requests should be restricted to the troubled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

Additional requests by risk area

25 As can be seen from Figure 4, the bulk of the additional requests concern credit and counterparty credit risk, followed by liquidity monitoring, which together account for 55% of all data requests¹⁹.

¹⁹ Only the main risk areas are taken into account.

Figure 3: Number of additional data requests per main risk areas (qualitative and quantitative)²⁰²¹



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

- 26 This picture seems to complement the harmonised EBA reporting framework, where credit risk monitoring comprises over a quarter of all existing data points²². 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.

Resolution data

- 27 Resolution reporting has only recently been included in harmonised frameworks across the EU with the EBA and SRB frameworks²³. Just under 1,000 defined data points exist in the EBA framework, of which institutions on average²⁴ 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

²⁰ SSM requests concern only Significant Institutions, while requests made by SSM NCAs concern only Less Significant Institutions.

²¹ The figure refers simply to the number of additional requests in a jurisdiction, without differentiating between requests with a 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.

²² Analysis is done on the defined DPM data points, meaning that data reported in open tables are counted only once.

²³ The two frameworks differ in the requested data, however, follow the same structure and format, aligned with supervisory reporting.

²⁴ Average for the sample of institutions currently reporting to the EBA (Largest institutions).

minimum harmonisation principle applies, the frameworks also differ in scope, and there may be further data collected by the national resolution authorities.

28 In this respect, a data inventory exercise was carried out in October 2021. The aim was to collect evidence on the data collected by resolution authorities on top of the harmonised EBA and SRB reporting requirements²⁵.

29 The SRB collects six reports besides the Resolution Reporting Requirements:

Table 2: Data collections initiated by the SRB outside the harmonised framework

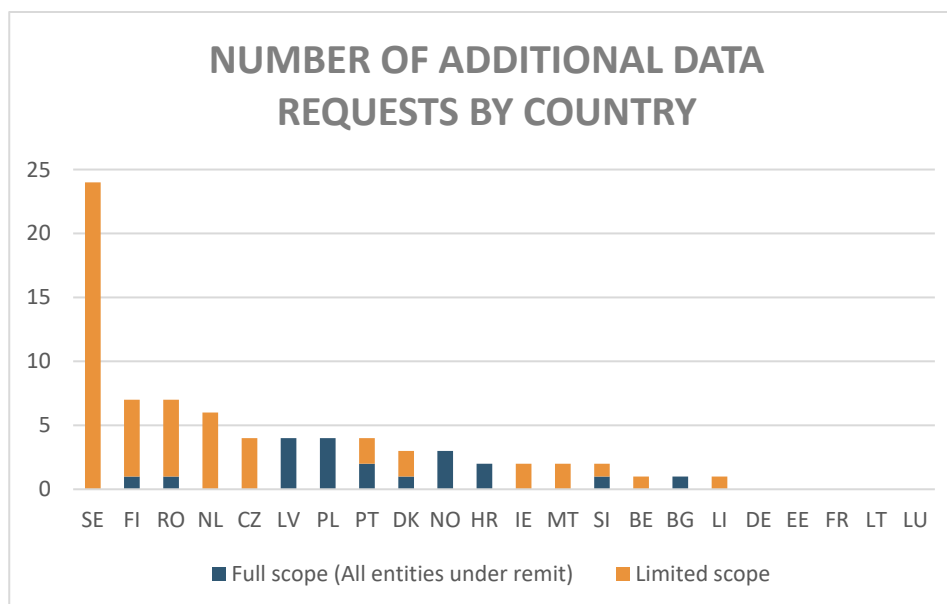
Sample	Name of request	Comments / description of data request
Limited scope: resolution groups under MPE strategy and resolution groups where the deconsolidation of mortgage credit institutions (MCI) applies	Additional Liability Report (ALR)	Collection of additional resolution data not reported in the EBA ITS on MREL and TLAC report.
Full scope: all credit institutions of the Banking Union and a limited subset of investment firms	Single Resolution Fund (SRF) ex-ante contributions reporting	Collection of FINREP-COREP and additional data in relation with the calculation of ex-ante contributions to the SRF.
Limited scope: entities with Strategy "Resolution"	ID card	Information collected in cooperation between NRAs and SRB as part of the assessment process. Collection of basic information on balance and P&L, resolution objectives such as credibility and feasibility of liquidation as well as of critical functions. In addition, collection of MREL, iMREL data.
Limited scope: LSIs	Quantitative assessment in accordance with Annex I of Commission Delegated Regulation 2019/348	Information collected from NRAs as part of the assessment process. NRAs shall conduct a quantitative and qualitative assessment including number of quantitative criteria measured on the basis of the indicators referred to in Article 1 and Article 3 SO DR, i.e. the score used when determining the other systemically important institutions (O-SIIs).
Limited scope: Resolution entities	MREL quarterly data collection	Collection of additional resolution data enabling the SRB to monitor quarterly the MREL build-up: gross issuances/outflows, outstanding amount of MREL instruments issued under UK law and yearly forecast on MREL issuances.

30 The dispersion of data requests is wide across the different jurisdictions²⁶. While 5 countries among the respondents do not have additional reporting requirements to those set out by the EBA and SRB, other countries have as many as 24 data collections. In addition to national and ad hoc data collections, the numbers presented below also include data requests originating from the EBA and/or the SRB that are extended by national resolution authorities to entities not in scope for reporting to the EBA/SRB. Not all competent authorities submitted information for smaller institutions. In general, the average number of data points is expected to be much lower, since in general, fewer templates are reported and for the majority of institutions no resolution templates are collected, and waivers are often granted for prudential reporting.

²⁵ Several respondents reported data collections that are part of the EBA and SRB frameworks. These requests were excluded from the data inventory, as out of scope.

²⁶ Austria, Cyprus, Spain, Greece and Italy have not submitted the resolution data inventory, and are excluded from the analysis.

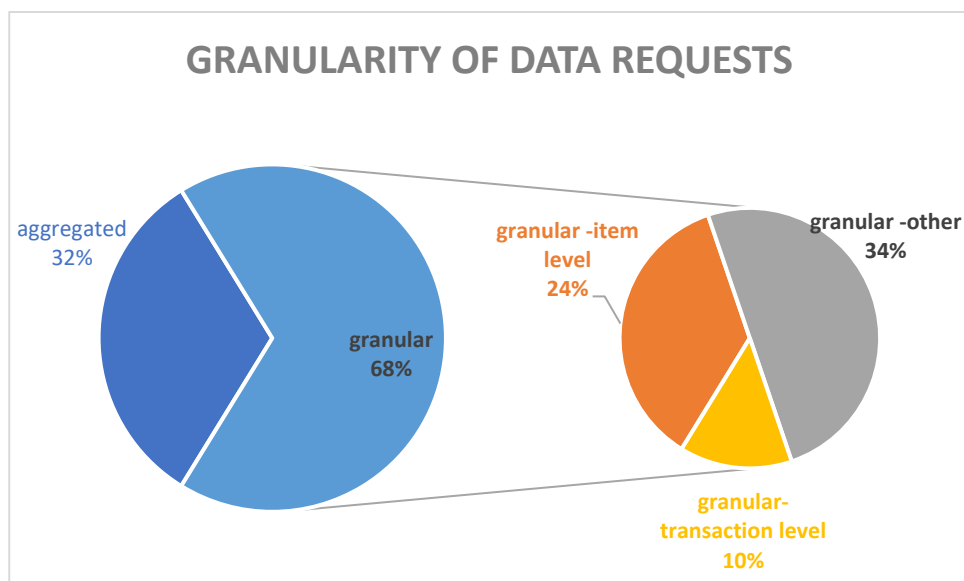
Figure 4: Number of additional data requests by country



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

- 31 On average additional data collections encompass 1,980 data points, with a maximum value reported corresponding to 42,576. As far as open tables are concerned, the average number of attributes is 514, and the maximum is 9,900.
- 32 When considering simply the number of additional data requests, 57% of them are both of a qualitative and quantitative nature. The proportion of only qualitative or quantitative data collections is very similar (22% and 21%, respectively).
- 33 Most additional requests (86%) are on an annual basis, aligned with the EBA and the SRB reporting. Quarterly frequency applies to 10% of the data collections. Unlike prudential data, no data collections are carried out with a higher frequency (monthly, weekly or daily).
- 34 Almost two thirds of the requests involve data requested in a granular form, for instance detailed by item and transaction. The remaining third is on an aggregate basis. Most of the granular data collection is carried out on an annual basis.

Figure 5: Granularity of the additional data requests



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

Statistical data

- 35 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 information made available by the ESCB on the harmonised statistical reporting for ESCB purposes. These amount, on average, to over 8,000 data points or attributes²⁷ per institution being collected for statistical purposes to fulfil the mandates of the ESCB only. This reporting is harmonised across the EU, however, there exist additional requests on a national level, which likely broaden the scope of data collected and reduce the level of harmonisation.
- 36 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 frequency²⁸. 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 institutions²⁹, with mixed and aggregated granularities, and monthly frequency as a minimum.

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 annually. 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 4,600 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.

Evidence from the cost-of-compliance study

- 37 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 different³⁰.
- 38 The cost-of-compliance study³¹ 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).
- 39 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 institutions³² (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. Sixty-eight per cent of participating institutions identify overlaps between EBA/standardised, regular reporting requirements and reporting requirements of a non-standardised/non-regular nature (ad hoc requests) as heavy contributors to the cost of reporting, providing evidence for the benefits of further integration.
- 40 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 view 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 institutions that agree is higher (46%).
- 41 There is no clear indication on how burdensome the interaction with the data recipient after submission and resubmissions is for institutions, 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

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.

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 institutions' own self-assessment.

format for national/ad hoc requests and EU ones would be beneficial for those reporting agents.

Preliminary conclusions

- 42 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 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.
- 43 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 institutions 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 frameworks³³ 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.
 - Institutions not being requested to report the same data multiple times.
- 44 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.

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

Annex 4: On the different types of data dictionaries coming from the feedback received on the EBA Discussion Paper

- 45 In the feedback on the discussion paper both authorities and institutions agreed that a common data dictionary is needed for regulators to define and communicate the regulatory data requirements to institutions. This common data dictionary would be necessary using the same methodology and vocabulary whatever is the regulatory framework and the authority responsible. Institutions need to understand the data they have to be compliant with and want to have the regulators integrated requirements in one only common data dictionary.
- 46 The common data dictionary has been referred as particularly relevant to the digital integration of data of different regulators and distinct processes along the regulatory data chain –the collection, validation, storage, transformation, exploration, and disclosure of data. All respondents see the need to have a common regulatory data dictionary at syntactic level, to facilitate the digital integration of the processes towards an effective integrated reporting solution.
- 47 Regarding a different type of a data dictionary, an input layer that would help institution to comply with their reporting requirements, based on the feedback on the discussion paper it is still not clear if institutions are interested in having it (which implies defining a common database to be populated as a condition to run common transformations and produce the data requested by regulators). However, some associations refer to the importance for institutions to have such a common system. Institutions explain that they use different internal data dictionaries depending on their way of managing data and how they manage the integration their day-to-day operations systems with their internal risk management, decisions and compliance systems.
- 48 Further assessment of this topic is necessary. Creating a data dictionary to ensure institutions' compliance implies the clarification of the following important aspects on:
- The regulators' degree of participation, taking into consideration the system would essentially be determined by institutions which would i) define the format of the database that they want in order for it to be easier to populate with their data; ii) define the transformations that would make it possible to go from this initial database to create the regulatory data requested by regulators.
 - How much of the costs of compliance of institutions will be transferred to the taxpayers. This depends on how much institutions are willing to pay the costs of moving from the current situation and on the regulators' perspective and

justification to have a public/private initiative – a clear position should be enunciated by regulators about their role and their contribution.

- How to ensure a fair playing field for all possible different service providers and their solutions in case a public/private initiative is considered. The regulatory compliance systems usually involve high expertise and IT resources, and in the current market situation, special measures should be taken to ensure proportionate equal conditions for smaller Fintech, Regtech and Suptech providers.
- How to ensure the integration with the common data dictionary defined by the authorities (the output layer dictionary that has to be reported by institutions) and the availability for all interested. The use of the same common methodology and vocabulary would be advisable, thus avoiding the costs to translate the same concepts in different approaches and the need to maintain two duplicated ways of defining data. The data definitions of the data dictionary used for the compliance with the reporting requirements should be available for the general public and not hidden under the processes or tools of any specific player or tool.

49 The benefits of implementing a common approach for compliance (input approach) needs to be subject to further investigation, as the interest of institutions were not explicitly stated in the feedback on the EBA discussion paper. However, from their responses it can be expected that some institutions could have some restrictions on moving to this common compliance solution (input approach) as they have to maintain their internal processes and data dictionaries, to ensure a more overall integration between their business management systems with their compliance systems and it would not be an option to separate both.

50 The implementation of the common input approach for compliance implies huge costs, related to the complexity of transformation definition. Among the challenges, institutions mentioned the difficulties to define calculated data (according to internal models) in a unique way, the difficulties in aligning local practices (each national authority allows some exemptions which creates big differences), as well as the alignment of reporting frequency and deadlines.

51 The costs associated with this option are related to the fact that the need to update the input database and the correspondent transformations in each new evolution of regulatory reporting has to be agreed and decided by all the institutions. A cost-benefit assessment would need to be accurate and define clearly who will bear the costs of defining and maintaining the database and the system always aligned. Service providers are also involved and costs related to ensuring the level playing field for all different service providers are likely to arise.

Annex 5: Data dictionary requirements – Regulators and industry perspective

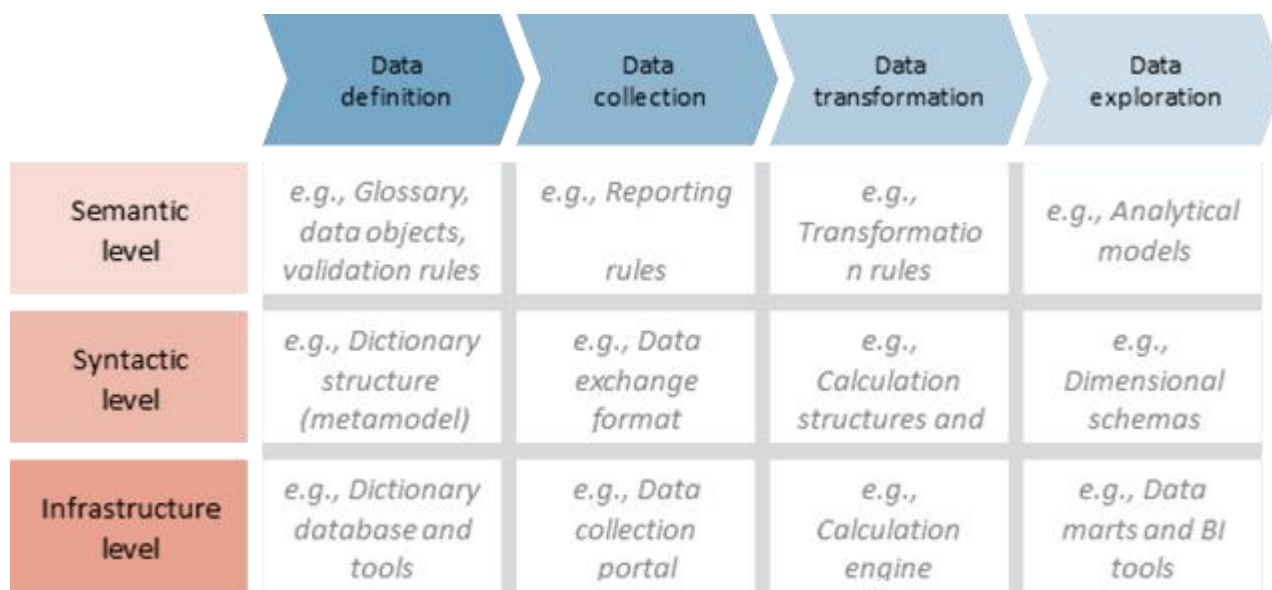
Overview

- 52 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.
- 53 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.
- 54 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.

Data dictionary: holistic approach analysis

- 55 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.
- 56 The integration requirements were analysed at three levels: semantic integration requirements, syntactic integration requirements and infrastructure integration.

Figure 6: Representation of the reporting process chain



- 57 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.
- 58 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.
- 59 The semantic definition of a concept needs to follow a consistent methodology for the selection of the characteristics that will better describe its meaning. The concepts should share these characteristics with other comparable data, and enable the identification of interlinkages between different data elements.
- 60 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 institutions' internal processes.
- 61 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.

- 62 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 the formats and features of the data dictionary and other elements that will support the architecture for data integration. The second effort encompasses the progress on the incremental semantic integration.
- 63 At an infrastructural level, the data dictionary addresses the concrete technologies and tools that will implement the integration across several data processing steps.
- 64 The perspectives of regulators and the perspective of institutions.
- 65 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.
- 66 Authorities have a specific data process chain and their perspective of what the ideal system should be and should take into consideration the reporting institutions' data process chains and specificities related to integration with their internal systems.
- 67 The data dictionary definition.
- 68 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.
- 69 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 and 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.
- 70 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.

The authorities' perspective

- 71 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.
- 72 Prudential and resolution frameworks integrate the data using standard data dictionaries that make it possible to identify and compare all data collected.
- 73 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.
- 74 In order to integrate data and avoid regulatory data overlaps, it is necessary to have a central common data dictionary that should be:
- a. **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.
 - b. **Complete:** have all the expected data elements that enable digital processing of reported data along the different processes of the data chain.
 - c. **Centrally managed:** ensures the technical standardisation which creates an effective data consistency and data comparability.
 - d. **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.
 - e. **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.

- f. **Ready for human interface:** easy to be used and understandable by business users
- 75 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.
- 76 The same common data dictionary should be able to define and identify different levels of granularity of the same original data.
- 77 Different regulatory frameworks can request the same data with different granularities. In this situation, regulators should converge, agreeing on 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.
- 78 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.
- 79 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.
- 80 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.
- 81 The validation rules have their own lifecycle and should be considered 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.
- 82 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.
- 83 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.
- 84 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.
- 85 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 by institutions.
- 86 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.
- 87 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.
- 88 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.
- 89 The data dictionary should enable ad hoc request extensions providing a transparent and complete view of reporting obligations, making it possible to identify and avoid any data requests from overlapping. Ad hoc data is usually requested at short notice, which can

demand a two-step approach with a previous agile data definition and a later step with a more complete dictionary definition.

90 At a syntactic level the dictionary should have the following general requirements:

- High frequency of change of data concepts in terms of scope extensions and data definition 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: read-only, 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.

The industry perspective: Stability in a high-change environment

Too many amendments

91 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.

92 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).

93 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

- 94 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).
- 95 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.
- 96 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

- 97 Add dictionary definitions to regulatory text within public consultations. In supervision, the regular regulatory frameworks go through public consultation processes which usually run for more than 1 or 2 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

- 98 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:
- 99 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.

100 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.

101 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 for 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 a 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 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 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 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 prudential consolidation/reconciliation mechanisms.

102 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.

103 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

104 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.

105 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 data quality.

106 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.

107 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.

108 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.

109 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

110 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.

111 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.

Alignment of concepts: Non-harmonised concepts

112 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.

113 A common vocabulary, shared and applied consistently along the different reporting frameworks make the integrated concepts comparable and harmonised.

114 Data harmonisation does not require data granularity. Data concept alignment can be defined and implemented on any level of granularity, using only the common dictionary definitions or completing with calculation rules mapping data with different levels of granularity.

Overlap of concepts

115 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 and converge data, avoiding unnecessary differences that can be burdensome for institutions.

116 If regulators cannot converge, institutions can also compare concepts and know-regardless of how similar or different they are and facilitate the data preparation.

117 **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.

118 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).

119 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.

Different reporting formats

120 The dictionary should be agnostic to any technology or related reporting exchange formats, and should be 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.

Spread and quantity of reporting requirements along legislative text

121 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.

Transparency and returned value for reporting institutions

122 A common data dictionary is a fundamental tool to support transparency and 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.

123 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.

Annex 6: Reporting Frameworks – Current state of integration

European harmonised reporting

124 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 also include the data validation/transformation rules for data quality assurance.

125 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.

126 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.

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

128 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.

129 Work on IReF has progressed significantly in a joint effort of the ESCB. The implementation date is envisioned for 2024-2027. A questionnaire took place in 2021, asking for input from banks (i.e. deposit-taking institutions) on a series of aspects that will form the basis of the design of the statistical reporting and the IReF system.

National regular reporting

130 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.

131 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.

132 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 institutions).

133 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.

Non-harmonised reporting

134 In addition to the regular data, there are a number of additional ad hoc or recurring data requests from prudential, 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 **Error! Reference source not found., Error! Reference source not found.**

Integration gaps

- 135 The European-wide regulatory reporting of prudential, resolution and statistics are at different stages of semantic integration and syntactic implementation.
- 136 On prudential and resolution frameworks, all the different frameworks are semantically integrated and all the data definitions have been 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.
- 137 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 public and available in a database together with other EBA outcomes, like the standard taxonomies for digital data exchange.
- 138 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).
- 139 On statistics, none of the different European-wide frameworks is integrated and they all have different data dictionaries. The ESCB has started the IReF initiative and plans to implement it in 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).
- 140 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.
- 141 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.
- 142 Each authority has its own infrastructure and the only sharing are the XBRL taxonomies produced by the EBA for data exchange and the SDMX taxonomies produced by the ECB for data exchange.
- 143 This segmented approach is responsible for a less-efficient model with missing opportunities of semantic integration and duplication efforts on redundant data dictionary efforts.
- 144 Possible semantic overlapping may exist in between the European-wide reporting and the national reporting at prudential, resolution and statistical frameworks.

Annex 7: The data dictionary – The framework differences

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

146 A preliminary overview of the profile of data of the different frameworks and the current integration status are relevant elements to take into account to explain the data dictionary requirements and define the possible paths and ways to go for the future development of feasible solutions.

147 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 institutions systems.

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

149 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.

150 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.

151 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, which originates new reporting frameworks managed within different calendars and added at least on a yearly basis. It is of utmost importance that prudential and resolution reporting requirements stay fully aligned with the underlying regulations and accounting standards and no discrepancies or delays are admitted. 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.

Annex 8: Experiences on European-wide data dictionaries

152 On regulators' side the different efforts on data dictionaries show different implementation maturity and different scope of integration and application. In fact, while some data dictionaries are already implemented, others are still in study or in pilot phase and even admit they are all prepared to be extended in scope; wider application to the different areas is still not implemented at European-wide level.

153 The DPM data dictionary, has been used from 2013 in each new EBA regulatory framework and later extended to integrate SRB frameworks. It is supporting all the European-wide frameworks of prudential and resolution reporting. It has been used by all CAs and RAs to report prudential and resolution banking data and was recently extended from banks to investment funds reporting. It has been supporting the standard exchange formats for the second level of reporting – between all CAs and RAs to the ECB, the SRB and the EBA – and in an increasing number of countries the DPM is also used by institutions to send the reporting data at the first level of reporting – from institutions to CAs and RAs. As indicated in the IR survey, the DPM is being used by reporting institutions to understand the data requirements referred to in regulatory templates. Some CAs have been using the DPM to extend to their national specific prudential and statistical requirements.

154 The ECB has established its SDD which will be expanded to support the future IReF for statistical reporting in the banking sector (2024-2027).

155 The ECB has been working with institutions on the BIRD project, developing solutions for institutions' compliance preparation. On supervision and resolution reporting BIRD is using the DPM data dictionary definitions and statistical dictionary definitions as the target concepts to achieve with BIRD defined transformations. The goal is to define a common approach by defining both a generic database that institutions would populate, and a set of transformations that institutions would apply to calculate the regulatory values. The BIRD initiative is running for more than 5 years, providing a set of documentation on the input database and the associated transformations for use by reporting institutions and service providers on prudential reporting.

Annex 9: The going granular impacts on a common data dictionary

156 The common regulatory data dictionary must assume the need to cover different granularities on reporting. In some cases regulators can achieve a common report of a particular subset of data serving two or more frameworks; in other situations this will be not possible.

157 The way to describe the same data with different granularities is in essence the same as describing different data concepts – it only implies the use of a good, complete and consistent methodology for data definition and ensures that the more complex aggregated data and the simpler granular data can be consistently defined and are comparable.

158 The options for going granular on regulatory reporting imply the need to prepare metadata of transformations that could be implemented in any system or transformation engine. When going for an integration based on a common granular approach it is important to identify transformation requirements and all the related costs. In fact, an accurate analysis should be performed in each case, avoiding considering only the possible simplification obtained in data collection because important costs exist when maintaining and operating the related data transformations. If the transformations are assumed other than by the institutions, it is relevant to clarify that the cost will be bigger than when they are implemented within institutions' internal processes. In fact, this will imply extra data (anchor values) and the need to support heavy and complex feedback loops. The quality assurance and transformation maintenance will increase the need for a strong operational system and adequate Q&A support.

159 The different scenarios analysing the regulatory going-granular options discuss two fundamental points: i) data for which institutions should be responsible (granular data, calculated data or both) and who will execute the transformations (regulators, institutions or a common future CDCP). Regardless of the future decision on these two points, the regulatory data dictionary would be able to support different levels of granularity defining the data relationships by consistently using the same glossary terms and the appropriate validation and transformation rules.

160 The regulators' granularity approach would reduce the need for institutions to develop transformations at least for the specific data that will be requested at a more granular level. The general tendency for regulation to be required at a more granular level will be a factor that undermines the need for a common solution for compliance (input approach) as it would reduce the need to develop the transformations by institutions. This together with the fact that some transformations are not possible to do from institutions' systems will significantly

reduce the space of any developments for a common approach for compliance (input approach).

Annex 10: Data dictionary – Cost-benefit assessment

General cost and benefit aspects

161 This section investigates the costs and benefits of the implementation of the common regulatory data dictionary, including those identified by the EBA and additional costs and benefits reported by stakeholders in the feedback on the EBA Discussion paper. Overall, most respondents to the feedback received on the EBA Discussion paper consider the move to a unique data dictionary as highly costly and less than one quarter as moderately costly.

162 Costs include those associated with setting semantic data definitions, data standardisation (translation of the semantic definitions to a formal data dictionary), creation of infrastructural tools and other outcomes that form part of the lifecycle of regulatory reporting (collection, validation, transformation).

163 Although the introduction of the common data dictionary entails costs, most stakeholders expect high costs reductions by integrating the national regulatory reporting together with the harmonised reporting, with only a third of them considering the cost reduction as moderate or low. Similarly, more than half of the respondents expect high cost reductions from the integration of ad hoc regulatory reporting, with only three of them expecting small cost reductions.

164 The transition to a binding common regulatory data dictionary is unavoidable but is seen as very costly by authorities. Some costs are expected to be of a short-to-medium-term nature while others would also be applicable in the long term. A common data dictionary will be effective only if it is used by authorities in their systems and in the data collection from institutions. However, this would imply authorities changing their systems.

165 Other costs are related to the possible implementation of a separate data dictionary for the IReF initiative, which aims to consolidate the statistical requirements into a unique framework with different levels of granularity. The IReF, is expected to be implemented in the period 2024-2027 and some duplication efforts could arise on defining a separated data dictionary independently of the common semantic and syntactic regulatory dictionary.

Regulatory data dictionary: costs

166 For competent authorities, the costs of setting up the common regulatory data dictionary are related to integration at the semantic, syntactic and infrastructural levels. First are the costs related to the semantic integration of frameworks from the coordination efforts with different authorities in order to integrate definitions from all frameworks (prudential, resolution and statistics) and evaluate and decide about the feasibility of semantic integration. These costs would be largely reduced by taking into account all the integration already delivered in between the frameworks of supervision and resolution and by adopting an incremental and progressive approach for the new frameworks to integrate those from the statistical area. The semantic integration costs of each framework varies from case to case and depends very much on the type of integration that is possible and how much it would imply in terms of costs to achieve just common definitions or something more (i.e. achieving common granular report with an additional cost layer related to the maintenance of accurate transformation definitions and all that is involved, like quality maintenance of transformations and the need of more exigent Q&A mechanisms). Second are significant costs involved in the implementation of the syntactic data dictionary (model structure to support the formal and standardised data dictionary) that can be reduced by starting from the already existing experiences on European data dictionaries. Third are the costs of adapting the systems to a new common regulatory data dictionary (more costs and less visible benefits in the initial phases), are very much dependant on how authorities have been developing their systems and how much they have already used the same integrated data dictionaries in the first and second level of reporting.

167 From the feedback on the EBA Discussion paper, competent authorities proposed a staggered approach as the feasible path towards integration and to achieve a common data dictionary of regulatory data. Thus, a phased approach would integrate framework by framework into the common data dictionary and will permit the progressive achievement of the target, by increasing the scope of integration on a case-by-case basis. However, the feedback also pointed out that in order for the data dictionary to achieve its goals, it needs to ensure unique and comparable definitions of the reporting requirements.

168 For institutions, costs are mainly related to adapting to a unique and new data dictionary in the reporting processes. From the feedback on the EBA Discussion paper, most institutions that responded to this question use more than one data dictionary in their reporting processes and would like to have only one regulatory data dictionary with all the requirements from different authorities. Therefore, there are costs associated with human resources and internal systems to adopt a new regulatory data dictionary. In the initial phases, there are more costs and less visible benefits (i.e. for supervision and resolution).

169 Other costs identified by institutions in the feedback on the EBA discussion paper are related to adjustments to the regulatory reporting process, development costs related to the operating model, organisational changes to set up a data-driven organisation and to review the organisation of the teams in charge of prudential, resolution and statistical reports, and cost of reviewing the governance.

170 Moreover, institutions mentioned the need to integrate national accounting frameworks into the unique data dictionary to avoid the creation of additional compliance costs. In relation to the integration process, institutions pointed out the need to choose the proper functional and technical approach to avoid sunk costs.

Regulatory data dictionary: benefits

171 For competent authorities, benefits could be obtained from the creation of comprehensive information of all regulatory frameworks, their value and scope of application (permanent stocktake availability). With this, the data comparability can exist and uniform mechanisms to data exploration and dissemination can be implemented. Other benefits are likely to arise from the creation of a level playing field for analysis by implementing data processes for types of analysis (i.e. comparative analysis, time series, ad hoc analysis, predictive analysis, etc.). Moreover, there could be more data-sharing between authorities.

172 The most relevant benefit is that the binding adoption of the regulatory data dictionary will enable having common reporting processes instead of developing one or more in each country. In fact, the binding nature of the data dictionary for collection processes (first and second level of reporting) will reduce the costs of the system. There are benefits obtained from the fact that data is collected only once, thus reducing process redundancies across countries and the amount of ad hoc data collections.

173 From the feedback on the EBA Discussion paper, competent authorities pointed out that benefits could be higher if the common data dictionary includes clear definitions related to reporting requirements and transformation rules. To maximise the benefits, deep cooperation between the parties is necessary in order to achieve a unique definition throughout the dictionary. Moreover, if the common data dictionary covers the ad hoc reporting for supervision, the transparency of outstanding data requests would incentivise authorities to reduce additional demands and to drop overlapping requests, which could represent a benefit for both authorities and institutions. In fact, the benefits of integrating data definitions stemming from ad hoc reporting requirements might be important, as evidenced from the cost-of-compliance study.

174 Institutions would benefit from clearer and more structured reporting rules easier to understand among people with different roles, backgrounds and skills involved in the reporting process. The integration of all regulatory frameworks under a unique regulatory data dictionary enables standardisation and integration with other national and international standards and ad hoc requests that could reduce the burden on the processes of regulatory reporting for institutions. Moreover, it creates better data access that eases the preparation of data reports with interest to institutions and other stakeholders. Lastly, it creates a level playing field in the application of the requirements that ensures a common understanding of the data requested and transformations performed.

175 From the feedback on the EBA Discussion paper, institutions agreed that national requirements could be reduced if competent authorities extract data for national interest from EU reports, a single source of data is found in one place, there is higher efficiency, better and structured governance within the bank (due to aligned requests from the supervisor), the consistency of the control process is facilitated for each report and across reports, data quality assessment is improved, there is additional available free time for analysis needed for strategic decision-making, there is a decreased amount of data produced by reducing data overlaps between frameworks and multilingual issues are reduced.

176 Institutions agreed on the costs included in the EBA Discussion paper, but mentioned that the benefits would depend on the quality of the data dictionary.

Common approach for institutions' compliance: costs

177 The common approach for compliance (CAC) would define a common input database and a common set of transformations to be defined by institutions. This approach requires the definition of a common database in order to run the common transformations. This solution will allow institutions to maintain their current situation and use their current solutions in their reporting and compliance processes or move to the common solution for compliance. According to the feedback on the EBA Discussion paper, all respondents except two use more than one data dictionary and institutions didn't refer to the need for this common approach. However, some banking associations referred it as relevant to facilitate the compliance processes of institutions they represent.

1. The costs associated with this option are related to the fact that the need to update the input database and the correspondent transformations in each new evolution of regulatory reporting has to be agreed and decided upon by all the institutions. A cost-benefit assessment would need to be accurate and define clearly who will bear the costs of defining and maintaining the database and the system aligned. Service providers are also involved; costs related to ensuring the level playing field for all different service providers are likely to arise.

178 The implementation of the common input approach for compliance (CAC) implies huge costs, higher than the costs of implementing the common regulatory data dictionary. This is explained by the needs to define, manage and execute complex transformations. Among the milestones, institutions mentioned the difficulties to define calculated data (according to internal models) uniquely, the difficulties in aligning local practices (each national authority allows some exemptions which creates big differences), as well as the alignment of reporting frequency and deadlines.

Common approach for institutions' compliance: benefits

179 The benefits of implementing a common approach for compliance needs to be subject to further investigation, as the interest of institutions was not explicitly manifested in the

feedback on the EBA discussion paper. However, from their responses it can be expected that some institutions could have some restrictions on moving to this common input approach as they have to maintain their internal processes and data dictionaries to ensure a more overall integration between their business management systems and their compliance systems, and it would not be an option to separate both.

Annex 11: The granularity level of the reporting requirements

Definition

180 Different authorities need information on business concepts/phenomena that can be quantified³⁴ (measured or defined) and therefore further analysed. ‘Loans’, for example, can be understood as such a business concept/phenomenon.

181 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.

182 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.

183 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).

184 Reporting requirements have been defined at different levels of granularity reflecting the underlying regulation purpose and responding to different policy needs. Devising ordered

³⁴ 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).

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.

185 At the moment none of the reporting requirements are 'fully granular' (or 'atomic') since:

- Frequency is not higher than monthly (although on national level, some statistical data are requested on a daily or weekly basis).
- 'Granular' statistical frameworks are based on end-of-period stocks, not flows (i.e. individual transactions/payments are not reported).
- Code lists are often much less granular than in institutions' internal systems (especially when it comes to types of instruments, type of collateral, purpose).
- In addition, data models granularity is referred to as 'normalisation'. Statistical frameworks using entity relationship models (typically AnaCredit) use a very low level of normalisation (close to ontology/semantic model), whereas institutions' internal data warehouses consist of a large number of tables and relationships where data is normalised. Data de-normalisation requires numerous transformations, which can be complex especially when it comes to calculations about collateral. In contrast, transaction reporting required by the ESMA is close to atomicity when it comes to row granularity and frequency.) This means that institutions already resort to a lot of transformations and aggregations when reporting granular data for statistical frameworks such as SHS or AnaCredit. All other 'traditional' frameworks such as BSI, FINREP, can also be seen as 'granular' to some extent.

186 A definition of granularity should be elaborated and analysed in the context of the reporting frameworks under the integrated reporting scope since the topic is complex and the opposition between 'granular' datasets and 'aggregated' datasets is somewhat artificial. Thorough analysis of the right level of granularity for the three dimensions (rows – records representing individual transactions/contracts/customers and others, columns (attributes with their associated code lists for categorical data and time), as well as model normalisation is needed before making diverse assumptions about the feasibility to go granular.

Feedback received from the EBA Discussion Paper

187 **Tree options have been presented in the Discussion Paper together with an overview of potential costs and benefits.** In all three options a common data dictionary was considered to have already been in place. In Option 1, the current level of granularity in the prudential and resolution data requirements could be maintained in the future integrated system. For the statistical data collection, the granularity will be further increased, as decided following

the cost-benefit assessment launched in the IReF project. In Option 2, where the granularity of the data requirements might be reconsidered such that data could be required with a higher granularity, while currently more aggregated data requested might be reduced. The third option, Option 3 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. The discussion paper highlighted that preliminary assessments pointed to the fact that Option 3 might not be feasible.

188 The discussion paper was looking for additional feedback to the feasibility and cost effectiveness of increasing the granularity of the reporting requirements as well as additional insights on the possible challenges and solutions. The main points touched upon are presented below.

189 Regarding the implications of granular data reporting on the institution's compliance with the BCBS 239 principles, most of the feedback received did not see any impediments. Some of the feedback received highlighted the positive impact that more granular reporting would have on the ability of institutions to follow such principles. However, one feedback received highlighted that BCBS239 would no longer be optimally met in the case in which group and individual reporting would be produced in different applications and that compliance can hardly be segregated from metrics calculation and certification.

190 Regarding the options presented, all respondents considered that Option 3 would be unfeasible. Some respondents saw Option 3 as a long-term objective that could be aimed for if experience proved it might be possible. The main implication of such a conclusion would be that it is not possible and sufficient to define data requirements only at a certain high granularity and collect it. Various impediments have been highlighted that point either to legal or technical limitations. Few feedback received saw only Option 1 as being the cost-efficient solution moving forward or some considered a possible path, moving from Option 1 to Option 2.

191 Most of the feedback received however considered that the future integrated reporting system should consider Option 2 as a way forward. Option 2 considers that there may be cases where currently aggregated data reported may be substituted by granular data requirements, however some aggregated data would still need to be required. The cost efficiency of implementing Option 2 is however highly dependent on the design of the process. Among possible cost inefficiencies the following were mentioned: the lack of proportional design, a high amount of feedback loops and anchor values, if additional granularity lead to additional reporting requirements on top of what is currently reported, if aggregated values would still need to be produced and reported by institutions, a poor design of the granular collection, if granularity were not accompanied by harmonisation in other areas such as data quality standards, revision policies or remittance dates. In addition, the challenges to collect granular data and define standardised and harmonised transformations might be significant (see paragraph 133 in the main report). Nevertheless, despite these challenges, general comments received indicate that these topics should be further explored.

Considerations when going more granular

192 Deciding on the optimum level of granularity would require **a common understanding of what defines the granular characteristics of a data set**. A common language would facilitate the comparison of the current data collections, while decisions such as increasing the granularity level would be easier to be understood in terms of which changes are envisaged and how big these changes are. Such a common understanding will be achieved with the creation of a common data dictionary.

193 **Need for a proper design/methodology of granular data definition and collection** – defining the data at a granular level together with its collection would require careful thinking of its design, especially in the light of a staggered approach where granular collections would be gradually included in the data dictionary (need for flexible and scalable solutions). The design would refer both to technical aspects (e.g. business requirements and modelling considerations) but also to the underlying operational procedures.

194 **Application of proportionality in reporting requirements** should be reconsidered in case of a move towards more granularity, as additional aspects might play a role in the reporting burden. Apart from defining which granular collections should be applicable for smaller institutions, additional proportionality measures should be considered with respect to anchor values/feedback loops, transformations and validations.

195 **Need to consider for the transition** period to avoid an excessive period of coexistence between old and new reporting requirements (for those cases where granularity can be achieved) and avoid the lack of modularity in the transition approach (to allow for gradual changes).

196 To achieve integration under a more granular reporting, 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. 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. 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. This would require a close cooperation between standard setters at both European and national level.

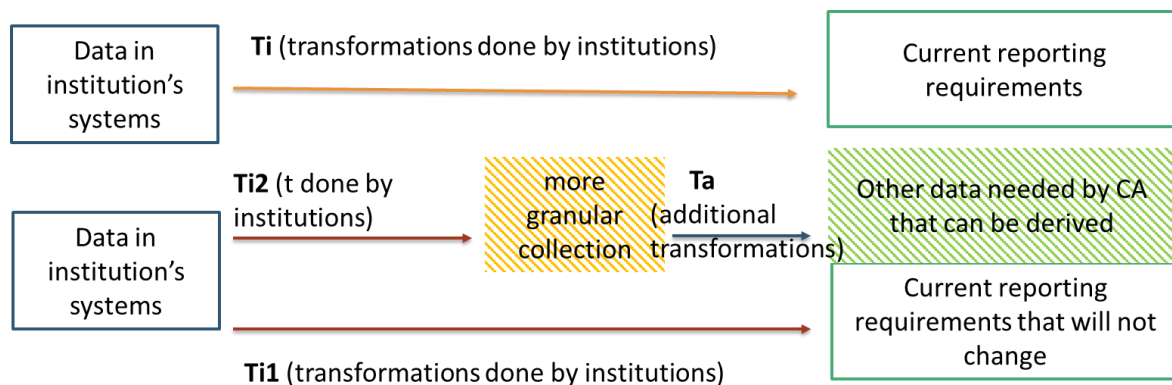
Transformations

197 Just as with the need to identify what kind of data can be requested at which level of granularity, it is equally important to distinguish between different types of transformations applying to these levels of granularity.

198 Currently each institution defines a set of transformations (Ti) from its internal systems to the given set of reporting requirements as currently depicted in the regulations.

199 With a possible move towards granularity driven by the integration of different datasets, one could define an additional set of transformations (Ta) that would link the more granularly defined data requirements to data needed for different purposes by competent authorities. Ti previously done by institutions would now be substituted by Ti1 and Ti2.

Figure 7: Granularity and data transformations



200 While it is considered that the Ti1 and Ti2 transformations will continue to be under the full responsibility of the institutions, the new Ta transformations would require additional thinking. While for statistical purposes/needs these additional transformations could be under the responsibility of authorities, in the case of supervisory and resolution needs, it has to be ensured that institutions remain responsible for all their data (reported and derived).

201 The feedback on the EBA DP showed that institutions and authorities alike would prefer to have **fixed definitions for the transformations that would leave no – or where not feasible, as little as possible – room for interpretation outside the one allowed in the legislation.** That means that Ta are a set of standardised transformations that would allow the computation of more aggregated data needs from a more granularly defined set of data.

202 There is a strong dependency between granular data and transformations, as the limitation to define standard transformations will show the limitations to how granular one can go and vice versa; the limitation to accessing granular data will limit the number of transformations that can be defined.

203 The process related to Ta can be split into: definition, data input, calculation and data output. Based on the feedback received, different preferences were expressed related to these processes and their governance. Both regulators and institutions are aware that there are costs of going granular. Depending on the design proposed, these costs might be too high to make for a change possible.

Scenario	Definition of transformations	Data input responsibility	Calculation/Execution of transformations	Data output responsibility
1	Jointly Authorities and Institutions	Institutions	Institutions	Institutions
2	Jointly Authorities and Institutions	Institutions	Authorities	Authorities
3	Jointly Authorities and Institutions	Institutions	Run centrally in a CDCP and shared with institutions and authorities	

204 It is generally agreed that **a joint effort of all the stakeholders involved (e.g. CA, regulators, reporting institutions) is needed when defining transformations.** Different experts from different areas would have to contribute to the reconciliation of data definitions and development of transformation rules. 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 needed only for specific reporting areas. Aligned with the Governance chapter, **industry would have a consultation role, while the transformations would be an integral part of the data dictionary.** Therefore, governance considerations for transformations would be part of the overall common data dictionary discussions. Proportionality aspects and representativeness of the industry should be ensured.

205 **Responsibility of the data input, both the content as well as the form (mapping from their internal data to the data required by the authorities)** would remain with the institution.

206 **Responsibility for calculating/executing the additional transformations and responsibility for the output** would depend on what kind of transformations are defined and the data output of these transformations. Different governance set-ups might be preferred or feasible. This would have a further impact on the cost efficiency of moving to additional granularity and would require further investigation especially from a legal perspective. Switching to increased granularity without reconsidering some of the processes would be very costly for institutions, (leading to an increase of complexity due to a new intermediate granular data model) and would make a switch to more granularity prohibitively costly.

207 Together with defining and applying transformations, additional processes might need to be considered such as a Q&A process to check the application of the transformations and processes to ensure data responsibility/accuracy (including feedback-loops and anchor values).

Annex 12: Granularity – Cost-benefit assessment

208 This section investigates the costs and benefits of higher granularity in the reporting requirements of statistical, prudential and resolution data in the context of an integrated reporting framework. The feedback received on the EBA Discussion paper showed the broad support to further explore an increase in granularity. Therefore, this section is focused on the cost-benefit assessment of the increase in granularity of overall requirements for both competent authorities and institutions. Further costs and benefits should be explored as part of a more concrete framework set-up for more granular reporting (including technical, operational, legal and governance aspects).

209 Both competent authorities and institutions would have associated **costs** to move towards a more granular data collection. Regarding **competent authorities**, the costs are related to the definition of transformations (jointly with institutions), the quality assurance process that might be more costly and collaboration with other authorities to integrate and reconcile data definitions across the legal frameworks of each area of regulation. Moreover, maintenance costs are expected to be high due to the evolving nature of the common regular reporting and the design of the set of calculations as new reporting requirements are added. In terms of coordination, there are costs from the coordinated action among different standard-setters at both European and national level to analyse the commonalities and differences in the data, the purpose for which it is collected under each regulation and the current process for collecting it.

210 Apart from the abovementioned costs, **from the feedback on the EBA Discussion paper**, competent authorities also identified costs related to additional staff needed, coordination efforts, risk of additional financial burdens for institutions due to cost allocations and costs related to the involvement of the industry in an advisory capacity.

211 The **costs that institutions may assume** are related to the collection of the data at a more granular level from various systems internally in a timely manner, the non-exemption of reporting aggregated data to ensure compliance with regulations and costs related to transformations. Regarding transformations, the costs are related to the definition of transformations and reconciliation of data definitions (jointly with competent authorities and regulators), responsibility of the data input, responsibility for calculating additional transformations and responsibility for the maintenance and enhancements of a common set of transformation rules. Moreover, institutions may also need additional human and IT resources for setting up, testing and maintaining the transformation, with the involvement of business experts from different areas.

212 Apart from those costs, **institutions also reported other costs in the feedback on the EBA Discussion paper**, such as homogenisation issues (from national GAAP to IFRS), costs of implementation (IT architecture, databases capacity, coordination efforts, etc.) that affect both institutions and authorities especially if it is a common set of transformation has to be shared by authorities and institutions, the inexistence of relief for institutions as calculation of aggregates still needs to be performed; in addition, requirements for granular data need to be fulfilled. Overall, institutions identified the timeline as the main cost driver for the implementation of the new reporting system (when the new dictionary is set up, it must be flexible enough to be implemented by all institutions and authorities, easy to update and adjust).

213 The **benefits for competent authorities** from increasing the level of granularity are expected to come from the enhanced flexibility in the usage of data through the possibility to transform it to respond to new policy needs, increased possibilities to process and analyse data with the purpose of supporting additional analysis, reduction of ad hoc data requests, higher comparability of the current data collections, better understanding for everybody of the data reported, enhanced transparency in the aggregation process through the definition of transformations that will link the data across different aggregation levels, less Q&A as these definitions are likely to reduce the Q&A about the data that should be reported, at least at the more aggregated levels. From the **feedback on the EBA Discussion paper**, competent authorities pointed out other benefits mainly from the absence of room for variations, the increased comparability, the binding nature, the standardisation of the process, a better understanding of supervisory considerations and information symmetry.

214 **Institutions would benefit** from enhanced stability in the reporting process; reduced new reporting requirements over time, as ad hoc requests may become less frequent; further cross-country harmonisation and standardisation of national reports with particular benefits to large groups that operate across borders within the EU; reduction of national reporting requirements (as going more granular in the data collection might further incorporate part of the current additional national requirements through national extensions, making them no longer necessary); creation of a level playing field in the application of the requirements (clearly and transparently defined transformations); simplification in the report preparation process (as the additional granularity strengthens the similarities between data stored in the institutions' systems and data reported); improvement in the quality of bank data at a more granular and at an aggregated level; clear and transparent transformations will ensure a level playing field for the application of regulatory requirements.

215 The **feedback on the EBA Discussion paper** showed that the **benefits** identified by institutions are aligned with the paragraph above, mentioning the reduced number of ad hoc requests, the additional stability of the requests, higher cross-country harmonisation, reduced duplications and reduced complexity of the reporting requirements. Moreover, institutions highlighted that the benefit comes mainly from a common dictionary shared among regulators which enables simplifying data within institutions. Another benefit could also be a common collection layer (from semantic to infrastructure level). On top, one of the

most important benefits would be the convergence between regulatory and internal monitoring uses.

Annex 13: Tasks of the decision-making bodies (European authorities, national authorities)

Data definition	<ul style="list-style-type: none"> – Develop/define regulatory reporting text and templates/aggregates to be endorsed by the relevant bodies. – Develop content of the single dictionary, which includes, among others: <ul style="list-style-type: none"> ○ vocabulary terms; ○ unique concepts, validation rules, reporting structures; ○ reporting calendar obligations. – Agree on the content of the common data dictionary.
Data collection	<p>Develop and decide on:</p> <ul style="list-style-type: none"> – a policy for data quality; – a resubmission policy; – master data on institutions; – reporting frequency, waivers and thresholds; – transmission formats; – reporting obligations calendars (who has to report what and by when). – reporting obligation rules (filling rules, etc.); – data flow channel and common data collection; – new data exchange standard files.
Data transformation	<p>Develop and decide on:</p> <ul style="list-style-type: none"> – transformation rules/validation rules; – decide plausibility checks.
Data exploration	<p>Develop and decide on data-sharing and data access.</p>

Annex 14: Tasks where coordination, collaboration and/or joint work is needed and where the JRC could have a role

<u>Data definition</u>	<ul style="list-style-type: none"> – Promote planning coordination. – Conduct integration impact analysis of reporting regulations. – Coordinate/advise on joint WG creation and outcomes. – Advise on the integrated reporting resulting from regulatory joint efforts: <ul style="list-style-type: none"> ○ common data concepts and validation rules; ○ common frequencies, institutions scope and master data. – Advise on data dictionary translation and exchange standards, which includes, among others: <ul style="list-style-type: none"> ○ vocabulary terms; ○ unique concepts; ○ validation rules; ○ transformation rules; ○ exchange mechanisms. – Ensure technical checks of joint WG on the different outcomes.
<u>Data collection</u>	<p>Provide advice in terms of integration issues for:</p> <ul style="list-style-type: none"> • a policy for data quality; • a resubmission policy; • reporting frequency, waivers and thresholds; • reporting obligations calendars (who has to report what and by when); • transmission formats and reporting obligation rules (filling rules, etc.); • dataflow channel and common data collection
<u>Data transformation</u>	<p>Advice in terms of integration issues for:</p> <ul style="list-style-type: none"> ○ transformations needed.
<u>Data exploration</u>	<p>Promote and advise on data-sharing and access to data.</p>

Data and metadata management	Advise on best practices in managing data and metadata, proposing possible efficiency gains in the reporting process including granularity exploration.
-------------------------------------	---

Annex 15: Best practices on governance ad hoc and national data request

Best practice guidance for authorities for an agile coordination mechanism (short/medium term)

216 These practices could help in streamlining and improving the internal governance processes at the authorities whereby new (regular or ad hoc) data requests can be requested only after checking that the data is not already reported and assessing the burden on supervised entities.

217 Creation of committees in each authority to perform a preliminary assessment 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 regulatory data dictionary or adapt the new data requests to the information currently available.

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

219 Each authority should have a register of the different data requests. The ECB has already established an SSM-wide data collection database which could be used as a model for other jurisdictions.

220 Any new data requests should be processed by using extensions of the common regulatory data dictionary (once it is built). 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 or other reasons valid in the view of the NCA.

221 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. The data dictionary team could provide technical support to authorities for preparing this information.

222 In the short term and in alignment with the cost-of-compliance recommendations the EBA could provide further guidance on best practices aligned also in the context of integrated reporting.

223 Once the JRC is set up and functioning, to follow-up on the possible data duplications could be considered, if any. Although the JRC may not impede data collection, it could have a role in following-up under which circumstances ad hoc and national data collections could not make use of the standard definitions.

Coordination mechanism through a possible CDCP (long term view)

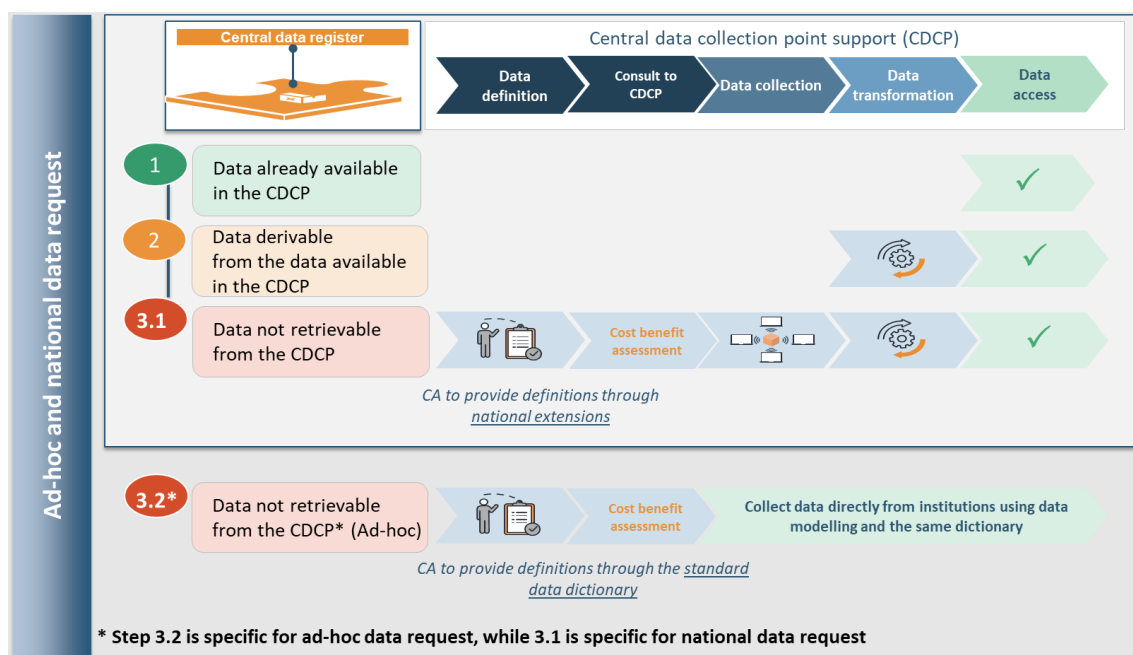
224 The long-term view of the agile coordination mechanism to process national and ad hoc data requests could consist of the following steps:

225 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 if 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.

226 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.

Figure 8: Coordination mechanism for data requests



227 The same previous pre-conditions will be applicable in this approach: agility, assistance to authorities, legal powers reside in authorities, exclusions and ex post inclusions.

228 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 presents a justification (very specific request, very short notice or short term, information collected for purposes other than statistical, prudential, resolution, etc.) to the CDCP.

Annex 16: On the push and pull mechanism

229 The EBA discussion paper further proposed for discussion a comparison between the current reporting model, the push approach, by sending the data from the institutions to the authorities versus a possible new approach, the pull model whereby it is the authorities that would pull the data from institutions' internal systems. Different aspects were considered between the two models, adding as another possible option the mixed approach (some data would be pulled while some other data will be pushed).

230 The traditional way of pushing the data by the institutions seems to be the preferred way across all respondents given the set-up put forward by the discussion paper (where the pull is seen as direct access to the institution's system and not preferable due to implementation costs, data responsibility, process change burden, operational risk and security challenges).

231 However, from the feedback received there seems to be a different understanding of a pull and push model and what a mixed model could look like. Significant feedback received see a mixed approach as one in which institutions would push the data to a place while authorities would pull it. The feedback received showed that there is some interest in exploring a different mechanism of accessing data in the following direction: i) either there would be a separate repository in the institutions system (latent data) that would fit certain types of data and from where authorities could potentially pull it or ii) data would be pushed by the institutions to some repository and pulled by authorities from there. However more details would be needed for a proper assessment, covering technological and concrete proposals for a governance and process setup of such a system.

232 The future IR system should be flexible enough to accommodate various ways to manage data and respond to technology innovation. Therefore, future in depth discussions on types of data and or/set up of a CDCP should account for various mechanisms to access the data.