EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR FINANCIAL STABILITY, FINANCIAL SERVICES AND CAPITAL MARKETS UNION

General Affairs
Economic analysis and evaluation

WORKSHOP ON BUILDING A COMMON DATA DICTIONARY IN EU FINANCIAL SERVICES

20 October 2023

DISCUSSION PAPER

DISCLAIMER: *This note has been prepared by the Commission Services (DG FISMA) and cannot be understood as representing the Commission's position on these issues and does not bind the Commission in any way.*

1. INTRODUCTION

On 20 October 2023, DG FISMA is holding a virtual workshop on building a common data dictionary in EU financial services. Building on the efforts of the European Supervisory Authorities (ESAs) and other relevant authorities to develop sectoral dictionaries in their domains, the longer-term aim is to have a common data dictionary covering all EU financial services.

As set out in the Commission's strategy on supervisory data ([1]), building such a dictionary will be key to ensure consistency of reporting requirements and achieve more data standardisation. Having a common way to define the meaning and possible values of the data to be reported will also make it easier to share and reuse the data for different purposes. In addition, the data dictionary can contribute to other longer-term goals such as making reporting requirements machine-readable and machine-executable.

The workshop aims to bring together experts from EU and national authorities as well as the financial services industry to present and discuss their expectations for a common data dictionary, including the main use cases and the design requirements. The workshop will also be an opportunity to get an overview of the state of play on the development of sectoral data dictionaries in banking, insurance/pensions and financial markets and to exchange views on how to advance these dictionaries and bring them together into a common dictionary.

([1]) COM/2021/798 final

This paper prepared by DG FISMA ([2]) seeks to provide a basis for discussion at the workshop. It also sets out targeted questions to gather input from experts from authorities and industry on what the common data dictionary should be. The objective is to bring different views together to reach a shared understanding.

The discussion paper is structured as follows: Section 2 gives an overview of past and ongoing initiatives to build a data dictionary. Section 3 presents different potential use cases for the dictionary, collected in various discussions with a wide range of stakeholders. Section 4 aims to set out what information the dictionary should contain to be able to support the identified use cases. It structures the information into different components and describes the interactions between them. The section is rather technical in nature, which is inevitable to achieve a sufficiently precise description and promote an informative discussion. Section 5 provides an overview of additional general requirements on the data dictionary. Section 6 concludes and invites stakeholders to provide feedback on the questions for discussion. Annex 1 is a brief glossary of key terms used in this paper ([3]). Annex 2 provides concrete illustrative examples of what information the different components of the dictionary would contain for a selected reporting requirement.

2. BACKGROUND: PREVIOUS AND ONGOING WORK

Significant work has already been undertaken to build data dictionaries and deliver more integrated data collection in different sectors, both at EU level by the ESAs and the European Central Bank (ECB), and at national level. Future work to develop data dictionaries, and build a common data dictionary, should build on this experience and the lessons learned.

While it is outside the scope of this discussion paper to review all the ongoing work, the following describes examples of such projects, focusing in particular on the banking sector. In its feasibility study on bank integrated reporting ([4]), the European Banking Authority (EBA) provided an elaborate assessment of building a data dictionary covering supervisory, resolution and statistical reporting in the banking sector, based on input from multiple stakeholders.

EBA represented different parts of the reporting process chain at different levels: conceptual (semantic level), formal and standardised formats (syntactic level), and the technological architecture (infrastructure level) as illustrated in Figure 1.

([2]) Earlier drafts were discussed with the Supervisory Reporting Roundtable, which is regularly hosted by DG FISMA to exchange and coordinate on reporting-related matters and comprises experts of the ESAs, the European Systemic Risk Board, the Single Resolutions Board and the European Central Bank.

([3]) As several terms used in this paper are used and understood differently in different contexts.

([4]) https://www.eba.europa.eu/eba%E2%80%99s-feasibility-study-integrated-reporting-system-provides-long-term-vision-increasing

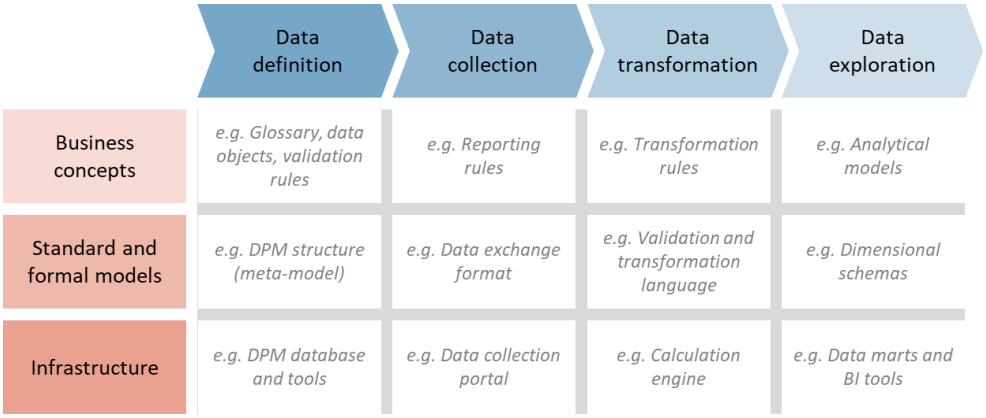

Figure 1 Reporting process chain and three levels of abstraction

EBA concluded that a data dictionary for prudential, statistical and resolution data collection was the central piece supporting the whole reporting process chain. According to EBA such a dictionary should be understood as a metadata repository covering business concepts represented in a standard format and supported by the appropriate infrastructure.

Furthermore, EBA identified the purpose of this data dictionary as describing all existing reporting requirements and transformations, aiming to avoid data and process redundancies to the highest possible extent, providing data clarity and comparability, and enabling data sharing. The feasibility study states: 'The common regulatory data dictionary [in banking] would support the integrated reporting system by including all definitions of all kinds of data requested by regulators in an articulated and consistent way, providing the description of the necessary interlinkages between data elements, whereby the data and transformations of this data dictionary are defined and maintained by the authorities within a formal and standard data dictionary prepared to facilitate the digital processing of the regulatory data.'

EIOPA has been working closely with EBA on the development of a data dictionary in the insurance sector. The two authorities share the data point modelling (DPM) approach and have recently updated it ([5]) to allow even closer alignment between them.

ESMA has launched its own initiative to develop a data dictionary in the financial markets sector (DATAD), which is currently closer to a repository of existing metadata as it relies now only on the metadata from its existing databases and the reporting messages. ESMA's objective is to get a comprehensive and comprehensible view of all the information it collects that is currently present in multiple databases. In addition, it also expects the dictionary to allow, in the future, identification of redundant reporting requirements and inter-dependencies between reporting requirements to better assess the impact of changes to these provisions.

In the context of the Integrated Reporting Framework initiative (IReF), and building on the Bank Integrated Reporting Dictionary ([6]) (BIRD), the ECB has set up a working group on an integrated reporting data dictionary (WG IRDD), which has also undertaken an assessment of developing a data dictionary covering the statistical and prudential reporting

([5]) EBA and EIOPA publish Data Point Modelling Standard 2.0 to foster collaboration and harmonisation in the field of supervisory reporting | European Banking Authority (europa.eu)
([6]) https://bird.ecb.europa.eu/

in banking. It has consulted national central banks to map the landscape and expectations for such a dictionary.

WG IRDD found that the term data dictionary was used with many different meanings and as synonym of several concepts, ranging from a simple collection of terms and their definitions to more complex structures. It attempted to provide a description of the common ground as '*a data dictionary being a convention/methodology on how to store metadata (semantic layer), a meta (data) model for the (structured) storage of metadata (syntactic layer), and a set of IT tools storing the metadata in the structure as given by the metadata model (infrastructure layer).'*

Further, WG IRDD concluded that the differences between existing data dictionaries lie in the form, structure, convention, and methodology according to which metadata is stored (using a general standard such as DPM, SDMX ([7]), or a tailored modification of such a standard). In addition, there are variations in the sets of IT tools built around the dictionaries.

The ECB concluded that a data dictionary, understood in a broad sense, can include many different features and structures and is use case specific. The ECB compiled a set of must-have features of the data dictionary for statistical and prudential reporting in the banking sector based on a survey of national authorities ([8]).

The ECB and EBA have set up an Informal Coordination Group ([9]) (ICG) that takes forward the work on bank integrated reporting, including the development of a data dictionary covering supervisory, resolution and statistical reporting in the banking sector. The ICG launched dedicated expert workstreams related to the development of such a data dictionary.

The banking industry has provided its views on the data dictionary for banking, generally supporting the creation of a single EU data dictionary as the cornerstone of an integrated and standardised EU framework for reporting in banking. According to industry views, the data dictionary should improve efficiency and consistency, and avoid duplication and unnecessary complexity of the reporting requirements. The data dictionary is seen by the banking industry as a pre-condition to starting the discussion on other important aspects of the integrated reporting project in the banking sector. However, while many efforts have been undertaken or are ongoing, currently none of the existing initiatives (IReF/BIRD, DPM, …) labelled as "data dictionaries" meets the requirements in full, not even within their respective domains (i.e., statistics, prudential, resolution).

Broadly speaking, both EBA and the ECB conclude that for a data dictionary to be implemented in practice one would have to define:

- what information (metadata) related to different parts of the reporting process chain the dictionary should contain – *semantic level*
- how that information should be represented in the dictionary – *syntactic level*
- what technology should be used to implement the dictionary – *infrastructure level*

([7]) SDMX – Statistical Data and Metadata eXchange | Welcome to the SDMX website
([8]) Not published.
([9]) Including the SRB and the Commission. Preparations are ongoing to set up a more formal structure, the Joint Bank Reporting Committee, that also includes national authorities and input from industry.

This discussion paper is limited to the discussion of what information should and should not be included in the common data dictionary – i.e., focus is on the semantic level. This is intentional as a precise definition of the content of the common data dictionary is a prerequisite for defining the other two levels.

3. USE CASES FOR THE COMMON DATA DICTIONARY

Building on the earlier work and further feedback obtained from authorities and other stakeholders, this section aims to define the purpose of the common data dictionary by elaborating on its main use cases. These are:

1. Interpretation of reporting requirements
2. Verification if certain data is already being collected
3. Defining a new or amending an existing reporting requirement
4. Facilitating the integration of reporting
5. Understanding the reported data

The following subsections give more detail on each of the use cases. At the end of this section, Figure 2 shows the main parts of the reporting chain related to each use case.

3.1. Interpretation of reporting requirements

Both authorities and reporting entities would use the common data dictionary to identify and interpret the reporting requirements (e.g., applicable to the reporting entity or in the remit of the authority). The common data dictionary would provide them with details of the reporting requirement necessary for business interpretation and implementation, including validation rules, and with references to the legal basis – to ensure a common understanding of the information that needs to be reported. The common data dictionary would also show how a reporting requirement has evolved over time (e.g., to let reporting entities identify changes needed to adapt processes to revised requirements).

The user would be able to look up one or a set of reporting requirements and view their complete description: who reports, to whom and what; specific sections of legal act(s) on which the requirement is based; precise information on substance and form of the information to be reported (definition, allowed values, format, standards to follow); data collection arrangements (transformation, validation, data flows and quality requirements), timeframes (first collection date, frequency, reference dates and submission deadlines), validity and changes over time. The information should be easy to find and understand by different types of users, from legal and business experts to technology experts responsible for the implementation of the requirements.

This use case is mostly relevant for reporting entities to obtain a comprehensive and unambiguous understanding of how to comply with a reporting requirement. This would help improve the quality of reporting, reduce the time of implementation and the reliance on the Q&A process for clarifications. It would also ease the transformation of the internal data of reporting entities into the required reports.

3.2. Verification if certain data is already being collected

An authority would use the common data dictionary to verify if certain data is already being collected to avoid introducing duplicated reporting requirements. It should be

possible to discover what the legal basis is for its collection, and which authority has the data (and potentially if and how it can be accessed).

Users seeking to verify the existence of a requirement to report desired data elements could find this information in the dictionary, or they could find existing similar data elements that (at least partially) satisfy the new reporting needs. The common data dictionary would also facilitate the assessment of whether the desired data elements can be obtained via a transformation of already collected ones (e.g., from more granular data).

3.3. Defining a new or amending an existing reporting requirement

An authority would use the common data dictionary when defining new or amending existing reporting requirements. It would define the new or updated reporting requirements using the concepts from the data dictionary, reusing existing concepts as much as possible and introducing new ones only if necessary. For a new reporting requirement, the dictionary should allow a sufficiently comprehensive description of the requirement(s) that would ensure a common understanding, as described in use case 3.1, in an unambiguous and structured way. For an amended reporting requirement, the dictionary should allow linking the new and the previous version of the requirement (i.e., versioning to understand how a reporting requirement develops over time). The dictionary should allow joint development and peer reviewing of new or modified reporting requirements among authorities.

Using the dictionary to define reporting requirements would improve their precision and consistency. The dictionary would encourage authorities to adopt a structured approach to design the requirements and support the use of standardised terms to describe them. It would also reveal any gaps in the specifications or inconsistencies with other requirements. By providing an unambiguous and structured representation of reporting requirements, it would facilitate the design and implementation of appropriate models and formats for exchange and storage of reported data.

3.4. Facilitating integration of reporting

An integrated reporting system aims at improving the reporting process and reducing burden for both authorities and reporting institutions. It covers the entire reporting process chain, including definition of requirements, data collection, data sharing and re-use, and implies enhanced coordination and collaboration between authorities. Its long-term objectives include ensuring a common set of uniquely defined concepts used to describe the reporting requirements (commonly referred to as the *'define once' principle*) and avoiding the duplicated reporting of the same data under multiple requirements (commonly referred to as '*report once' principle)*. There seems to be a wide agreement among stakeholders that a common data dictionary stands in the centre of the reporting process chain, and as such is a first step towards integrated reporting.

Authorities working together with reporting entities would use the data dictionary to pursue the define-once principle. Users could identify similar but not identical reporting requirements derived from different reporting frameworks. The differences would be reflected explicitly in the data dictionary together with the relationships among concepts.

Having the dictionary where concepts can be uniquely identified, precisely defined, and relationships between them represented explicitly would support the subsequent work on aligning the definitions and, where feasible, gradually integrating redundant concepts (this

process is commonly referred to as *semantic integration*). The results of the semantic integration would in turn be registered in the dictionary.

Integrated reporting, as outlined above, nonetheless goes beyond consistent definitions of concepts, description of the reporting requirements and collected data. It also covers other parts of the reporting process chain: common standards and models (e.g., data exchange formats, common identifiers), data sharing and reuse with appropriate procedural aspects (e.g., access rights, quality assurance), and infrastructure (e.g., a central data collection point or standard interfaces to access data collected by various authorities). The common data dictionary would not directly encompass these other parts but could facilitate putting in place technical and architectural elements of integration.

3.5. Understanding the reported data

Data users would be able to find in the common data dictionary information about datasets and individual data elements collected by different authorities as a result of various reporting requirements. For the information to be useful e.g., to perform data analysis, this would require a sufficiently comprehensive description of the collected data without any ambiguities.

When performing analysis, an authority could use the dictionary to discover and interpret available data and to determine the feasibility and ways of combining data reported under different frameworks, in different sectors and/or at different points in time ([10]). Data analysis often requires comparing, transforming and combining data, and so the dictionary could also allow capturing information about data derived in the process of the analysis (i.e., new concepts, data transformations, etc.) as opposed to only about the data that is actually reported.

Much of the descriptive information about the collected data would be identical to the description of the underlying reporting requirement (see related use cases 3.1 and 3.3). However, some information relevant for understanding the reported data would likely differ from the one needed to interpret reporting requirements ([11]).

Also capturing information about what analysis different authorities carry out and which data they use for it could contribute to sharing of expertise between them and would also provide valuable information about the impact that modifications to a particular reporting requirement may have on the work of other authorities. Access to this kind of information would have to be controlled, however, as disclosing the precise information about analyses authorities perform on the data (e.g., for AML purposes) could undermine the effectiveness of supervision.

([10]) This task is different than verifying if certain data is already being collected, which is described in use case 3.2. Data analysis will also likely be carried out by different sets of users than verification of whether certain data is already being collected to avoid introducing duplicate requirements.

([11]) For example, some of the information about the reporting requirement (such as the name of the authority to whom to submit the report, the trigger for reporting, etc.) is not usually relevant for understanding the collected data. On the other hand, information that is not necessarily available in the description of the reporting requirement, is added to the data throughout the reporting process chain (e.g. precise submission date of the report, quality checks performed and the results of the data quality assessment), and it is important for understanding the collected data to have that information described in the data dictionary.

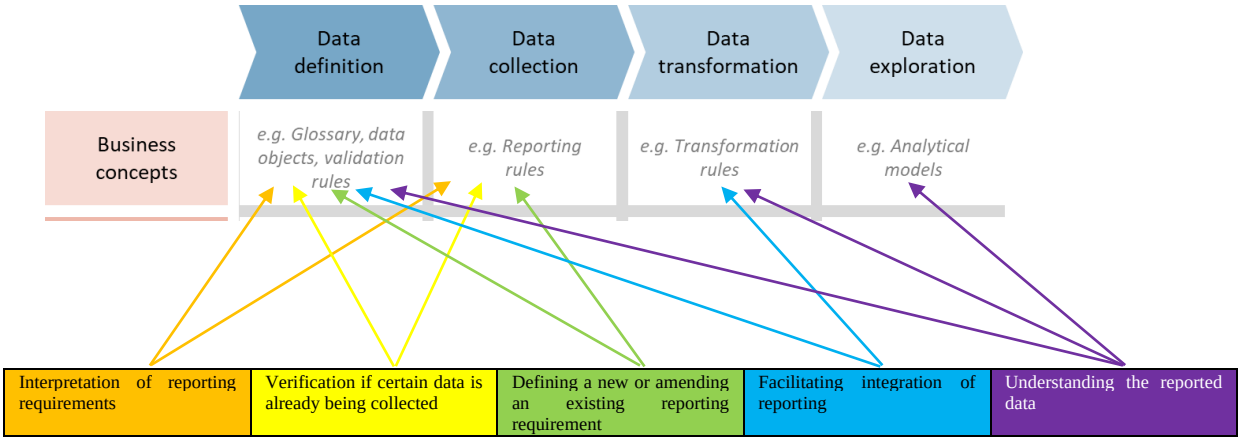

Figure 2 Mapping of use cases to the reporting chain

Questions:

1. **What is the main purpose of the common data dictionary? What is the relative importance of the listed use cases for your organisation?**
2. **What other use cases for the common data dictionary, if any, should be considered?**

4. POTENTIAL CONTENT OF THE COMMON DATA DICTIONARY

This section attempts to describe the information that would potentially need to be included in the common data dictionary to support the different use cases identified in the previous section. **All use cases require the common data dictionary to convey information about the meaning of the data to be reported, analysed, or shared.** This information can be divided according to its character into several components.

1. Glossary of concepts
2. Repository of data elements
3. Repository of reporting requirements
4. Data catalogue
5. Relationships, assertions, transformations component
6. Registration and administration information component

Some of these components may go beyond what many understand to be the core content of a data dictionary ([12]).

([12]) The components may also go beyond what is in the data management domain called a data model, which is a visual representation of the relationships between different data elements in a system.

4.1. Components of the common data dictionary

The potential components of the dictionary would each contain information of different character, as described below.

Glossary of concepts - The concept glossary would contain the description of the terms and definitions of a set of concepts (including possible multiple terms and/or definitions for the same concept) – with references to the relevant legal acts or other authoritative sources (e.g., international standards) from which the terms and/or definitions originate. The concepts in the glossary would represent units of knowledge and would be used to describe other pieces of information in the dictionary, as described below. A regulatory concept glossary would be essential for all the identified use cases.

Repository of data elements – this component would contain a complete, consistent, and structured description of the meaning and the possible values of all basic units of data – *data elements* – using the concepts from the concept glossary. The data element descriptions would be self-sufficient and not relying on information included in the reporting requirement(s) under which the data is collected or in the dataset(s) in which the data elements might be included by the authority/ies that collect or process them. This would permit re-using the data elements (as one reuses LEGO bricks) in other components of the common data dictionary, as explained below. A repository of data elements would be essential for all the identified use cases.

Repository of reporting requirements – this component would contain a comprehensive and structured representation of all relevant *reporting requirements*. Such a representation would have to include:

(1) *Reporting population* – who are the reporting entities subject to this requirement;
(2) *Content of the report* – what information should be reported – assembled by referring to one or more data elements from the repository of data elements component;
(3) *Recipient(s)* – to whom the information should be reported;
(4) *Timing* – when the information should be reported (e.g., frequency, submission date);
(5) *Conditions* – what, if any, additional conditions to be met for the reporting requirement to apply; and
(6) *Legal basis* – reference to the underlying legal acts (referring to the provision(s) in the original legal document laying out the reporting obligation).

While relevant for all use cases, a repository of reporting requirements appears to be essential in particular for the use cases "Interpretation of reporting requirements", "Defining new or amending existing reporting requirements", and "Facilitating integration of reporting".

Data catalogue – Unlike the repository of the reporting requirements, the data catalogue would provide a description of the data reported to the authorities and not the reporting requirements that lead to the data being reported. The authorities usually organise the data they receive into datasets and often add supplementary information to those datasets (see use case 3.5 above). Unlike the repository of reporting requirements, the data catalogue would not contain the information on the 'who', 'to whom', 'when' and 'under what conditions' to report.

The data catalogue would have some commonality with the repository of reporting requirements in that it would refer to the repository of data elements for describing individual elements of collected data. This would ensure that data elements are described in exactly the same way throughout the reporting process chain. The data catalogue would also contain information linking the collected data to the underlying reporting requirement in the reporting requirements repository.

In addition, the data catalogue would include the information on how the individual data elements are organised in the datasets. It could also include information needed to use the data, such as where and by whom the data is accessible (e.g., reported to which authority, including if the data has been shared, access rights, the technical access route), examples of how it is being used, warnings on pitfalls one should avoid when using it, and where the data quality information can be found and how it should be understood.

A data catalogue appears to be essential for the use cases "Understanding the reported data" and "Verification if certain data is already being collected" and potentially also "Facilitating integration of reporting" to support the report once principle.

Relationships / assertions / transformations – This component would contain information that represents the various associations between the different pieces of information in the dictionary.

Relationships would be essential for the interpretation of the information in the data dictionary. There exist various types of relationships between concepts and objects in the real world that the dictionary should be able to represent. Some important ones are:

- Classification relationships which allow building taxonomies of concepts from the glossary (e.g., loan is a debt instrument).
- Composition relationships – which connect concepts representing the whole and its parts (e.g., assets are part of a balance sheet, or data element gross carrying amount of trade receivables is part of FINREP template F05.01)

Assertions ([13]) would represent statements about one or more pieces of information that must be true. Data validation rules are an important type of assertions.

Transformations would provide information (e.g., a formula) on how to derive values of certain data elements from values of other data elements (potentially reported under different frameworks, in different sectors and/or at different points in time).

This component appears to be essential for all the use cases as it provides the information on the relations between the items in the dictionary and therefore allows conveying the meaning of the information.

Administration / registration – The administration component would contain information on initial registration and subsequent administration of the lifecycle of all the items in the dictionary. This component would support joint management of the content of the dictionary by multiple authorities by holding information on the roles and rights of different users in accessing information in the dictionary, adding new items in the dictionary (registration), and modification and lifecycle management of existing items in

([13]) Transformations and assertions would be included in the relationships component because they could be viewed as special types of relationships that are distinct in that they must contain a formula or statement linking other items in the dictionary.

the dictionary (administration). In particular, it would contain information on which items were registered by which authority and provide the versioning functionality that is needed to capture and trace the lifecycle status (administrative status, period of validity, etc.) of various items in the dictionary (concepts, data elements, reporting requirements, relationships) over time. This component appears to be essential in particular for the use cases "Defining new or amending existing reporting requirements", and "Facilitating integration of reporting".

4.2. Interplay between the components

The components of the common data dictionary introduced in the previous section are not independent parts but complement and build on one another. Their potential interplay is schematically illustrated in Figure 3.

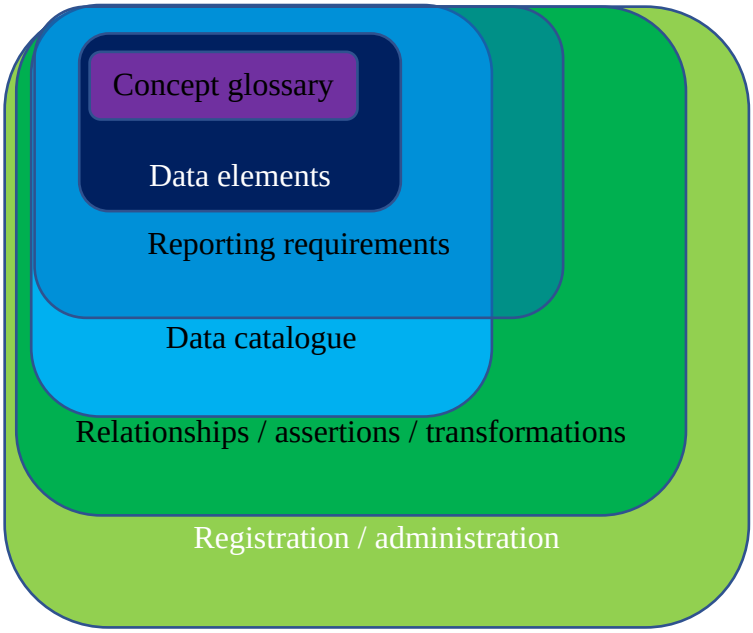

Figure 3 Components of the common data dictionary

During the implementation of the common data dictionary the components could either be all implemented at once or be progressively added to the dictionary starting from the core ones. It would mean the dictionary would initially support only some of the use cases, or only partially and progressively support the use cases in parallel with achieving more complete coverage.

The concept glossary, with a rather simple structure, could form the foundation of the dictionary by providing the vocabulary for other components and holding elementary information for each of the concepts – the term(s) used to represent them, their definition(s) and the authoritative sources (legislative or not) from which they originate. Most of the concepts would likely not be directly describing the reported data elements but would be related to the characteristics needed to describe them (e.g., categories of financial institutions, types of financial instruments, accounting concepts) in the repository of data elements.

The repository of data elements could be viewed as the workhorse component of the dictionary because it would focus on describing the individual pieces of data being reported using the concepts from the concept glossary. Its structure would be more complex to allow

a sufficiently comprehensive description of the meaning and possible values of each data element. The data elements included in this repository would be referred to by other components, most importantly the repository of reporting requirements and the data catalogue, ensuring consistent description of data elements across the whole reporting process chain.

The reporting requirements component would contain the information required to describe the specifics of those requirements. For the information on what should be reported, this component would use the data elements from the data elements repository while adding other relevant information, as explained in section 4.1., to fully specify the reporting requirement. In addition, it could include the information on the layout of the reports (tables, templates, etc.), often referred to as rendering information. This component would be the entry point to the dictionary for users who want to interpret the reporting requirements.

The data catalogue component of the dictionary would hold information describing the datasets collected by the authorities. It would draw on the repository of data elements to describe individual elements of the datasets and add additional information needed to sufficiently describe the datasets.

The relationships / transformations / assertions component would contain the information that links several items from previous components, as described in section 4.1. It would be a step change in the expressive power of the dictionary, but also in its complexity.

Finally, the administration / registration component would allow keeping track of changes of the information in all the components of the dictionary and support the joint governance between multiple authorities.

4.3. Tools

Although not part of the common data dictionary, add-on tools (browsers, visualisations, API, etc.) would be essential for interacting with its various components. They would provide an interface to the content of the data dictionary, so that it can practically support the different use cases.

Appropriate tools would be essential for the development and maintenance of the data dictionary by authorities in a collaborative way, as well as for making the data dictionary content available to the wider public. To make use of the rich content of the data dictionary, tools could be developed to present the content in a tailored and user-friendly way for authorities and reporting entities alike (depending on different user needs and access rights). Visual maps could provide an overview of the different reporting requirements and relationships, which then could be compared to conduct analysis across reporting frameworks and sectors.

Other tools could identify and help analyse similar concepts and definitions. Such tools may include search and filter functionality. Warnings/alerts could help identify possible data gaps, overlaps, inconsistencies, etc.

Tools may also help generate content of the dictionary e.g., producing all the corresponding data elements in a reporting requirement in an efficient way (potentially using AI/natural language processing), as well as leverage versioning information in the registration component to present the status of the data dictionary at any given point in time (a “time-machine” function).

Questions:

3. **Are the presented components of the data dictionary all necessary? Which components are missing?**
4. **Are the presented components defined correctly?**
5. **Is there any way to prioritise the implementation of the components (e.g. staged approach)? If there is, in what order?**

5. GENERAL REQUIREMENTS FOR THE DATA DICTIONARY

The content and structure of the dictionary should meet a number of other general requirements, including that the dictionary should be:

- **comprehensive:** the data dictionary should preferably cover all collected data (prudential, transaction, statistical, etc.), whatever their granularity;
- **focused on data comparability**: the data dictionary should describe data elements consistently to achieve comparability of data;
- **ready for digital processing**: the data dictionary should enable the digital processing and exchange of its content as well as of the reported data;
- **technology-agnostic:** the data dictionary should be technology-agnostic and compatible with any data exchange standard;
- **ready for human use**: the data dictionary should be easy to use and understandable by people for analysis and collaborative work;
- **covering all regulatory data chain processes**: the data dictionary should cover data collection, validation and transformation, analysis and dissemination to support interoperability of different processes;
- **supporting joint maintenance by multiple authorities**: a common data dictionary should be developed and maintained jointly by authorities overseeing the EU financial system. Its structure and content should support the governance processes of such joint development and maintenance.

Questions:

6. **Are the general requirements on the data dictionary in this section complete? Are they sufficiently precise to serve as guidance for the development of the dictionary? If not, how can we advance to achieve sufficient precision?**

6. CONCLUDING REMARKS

This discussion paper provides a basis for discussion and for gathering input from experts from EU and national authorities as well as the financial services industry on what the common data dictionary in EU financial services should be, with the aim to reach a shared understanding. It presents potential use cases for the common data dictionary based on previous work and discussions with stakeholders. In addition, it outlines the potential content of the dictionary to support the identified use cases and organises this content into modular, but interacting, components. Finally, it summarises the general requirements on the dictionary, expressed in previous work of different authorities.

Building a common data dictionary is a highly complex and demanding task, requiring expertise in several domains and significant resources. This paper only discusses the objectives of a common dictionary, how a common data dictionary would be used and what information it should contain. Agreeing on what to build is a key first step needed to advance on implementation. A whole set of other issues related to how such a common dictionary should be developed will require further discussion and technical assessment. Although not discussed in this paper, feedback on the approach to building a common data dictionary is also welcome (see question 7).

DG FISMA aims to support the work and facilitate the cross-sectoral coordination. The practical implementation of the common data dictionary, however, requires the technical expertise of other authorities, including the ESAs who are working on dictionaries in their sectors. A common dictionary can only be achieved by different authorities working together and leveraging the expertise of the industry.

Stakeholders' contributions are welcome on all aspects presented in this paper and on the questions included in the respective sections in particular – either during the workshop or as written feedback afterwards.

Questions:

7. **How far should we go in building a common dictionary that applies across sectors? What approach should be taken to develop it and what role should different stakeholders play? To what extent should the approach be cross-sectoral as opposed to the current staged approach that starts with sectoral dictionaries?**

Contact: FISMA-SUPERVISORY-REPORTING-REQUIREMENTS@ec.europa.eu

ANNEX 1 – GLOSSARY

assertion – sentence or statement which is assumed to be true (*Example: Carrying amount is greater or equal to zero.*)

concept – unit of knowledge created by a unique combination of characteristics (a concept is independent of its representation)

data – re-interpretable representation of information in a formalised manner suitable for communication, interpretation, or processing

term – representation of a concept by a linguistic expression (e.g., the concept of credit institution can be represented by terms: 'entidad de crédito', 'Kreditinstitut', 'credit institution', …)

definition – representation of a concept by a descriptive statement which serves to differentiate it from related concepts

data element – unit of data that is considered indivisible (*Example: the data element 'age of a person', with values consisting of all combinations of 3 decimal digits*)

data model – a visual representation of the relationships between different data elements in a system. It acts as a blueprint for organizing and structuring data to ensure consistency, accuracy, and accessibility.

dataset – identifiable collection of data available for access or download in one or more formats

metadata – data that defines and describes other data

relationship (or relation) – sense in which concepts may be connected, via constituent roles (*Example: causality is a relationship with two constituent roles: cause and effect*)

taxonomy – a hierarchical arrangement of concepts in which groups of concepts are classified as subtypes of more abstract concepts

transformation – a description of a data manipulation operation that may be performed on data

ANNEX 2 – ILLUSTRATIVE EXAMPLE – FINREP

This annex presents an illustrative example of the information that would be included in the common data dictionary (as set out in section 4) for a specific reporting requirement under a selected reporting framework. The example is intended to provide a concrete case for discussion, which could be potentially generalised and refined to specify the requirements for the dictionary.

The example – chosen from FINREP reporting in banking – introduces the reporting framework and the relevant legislative texts that contain the information about the reporting requirement and the data element at hand. It then provides the information that would be captured in each of the components of the dictionary for it to support the different use cases.

Legal framework

This example uses a specific FINREP reporting requirement and a specific data element reported in row 0030, column 0005 in the template F05.01 of FINREP, part of the prudential reporting framework under the Capital Requirements Regulation (CRR) in banking sector. The information to sufficiently describe the reporting requirements and the data element to be able to interpret what its value represents is provided in several provisions across several pieces of legislation:

REGULATION (EU) No 575/2013 (CRR) ([14])

REGULATION (EC) NO 1606/2002 (IAS) ([15])

COMMISSION IMPLEMENTING REGULATION (EU) 2021/451 ([16]) (ITS)

- Annex III Reporting financial information according to IFRS
- Annex V Instructions for reporting of financial information

Information to be captured in the data dictionary

The data dictionary should bring together the scattered information from legislation and provide a sufficient description in a consolidated manner that is easy to interpret and adheres to other requirements laid out in section 4. Manually consolidated information related to the reporting requirement and data element selected for this example is given below. The information is organised by components. As done throughout this discussion paper, only the content of the information is presented and no particular structure or technological implementation is suggested.

Component 1: Glossary of concepts

([14]) http://data.europa.eu/eli/reg/2013/575/oj
([15]) http://data.europa.eu/eli/reg/2002/1606/oj
([16]) http://data.europa.eu/eli/reg_impl/2021/451/2022-03-03

Sample concepts illustrating the relevant content of the glossary.

Term/designation	Definition		Context	ID
	Text	Source		
institution	a credit institution or an investment firm	CRR, Article 4(3)	CRD/CRR	1
credit institution	an undertaking the business of which is to take deposits or other repayable funds from the public and to grant credits for its own account	CRR, Article 4(1)	CRD/CRR	2
methods for prudential consolidation	The institutions that are required to comply with the requirements referred to in Section 1 on the basis of their consolidated situation shall carry out a full consolidation of all institutions and financial institutions that are its subsidiaries or, where relevant, the subsidiaries of the same parent financial holding company or mixed parent financial holding company…	CRR, Part 1, Title II, Chapter 2, Section 2	CRD/CRR	3
Reporting reference dates	Institutions shall submit information to competent authorities as this information stands on the following reporting reference dates: (a) monthly reporting: on the last day of each month; (b) quarterly reporting: 31 March, 30 June, 30 September and 31 December; (c) semi-annual reporting: 30 June and 31 December; (d) annual reporting: 31 December.	ITS, Article 2 (1)	CRD/CRR	4
FINREP reporting frequency	The information referred to in paragraph 1 shall be submitted as follows: (a) the information specified in Annex III, Part 1, with a quarterly frequency; (b) the information specified in Annex III, Part 3, with a semi-annual frequency; (c) the information specified in Annex III, Part 4, with the exception of the information specified in template 47, with an annual frequency; (d) the information specified in Annex III, Part 2, template 20, with a quarterly frequency where the institution exceeds the threshold laid down in Article 5(5), the second subparagraph; …	ITS, Article 11 (2)	CRD/CRR	5
–	institution that prepares its consolidated accounts in conformity with the international accounting standards adopted in accordance with the	CRR, Article 99 (2)	CRD/CRR	6

Term/designation	Definition		Context	ID
	Text	Source		
	procedure laid down in Article 6(2) of IAS			
other than held for trading, trading assets or held for sale assets	The following accounting portfolios based on IFRS shall be used for financial assets: (a) 'Financial assets held for trading'; (b) 'Non-trading financial assets mandatorily at fair value through profit or loss'; (c) 'Financial assets designated at fair value through profit or loss'; (d) 'Financial assets at fair value through other comprehensive income'; (e) 'Financial assets at amortised cost'.	ITS, Annex V, Part 1.15	CRD/CRR	7
accounting portfolio	financial instruments aggregated by valuation rules…	ITS, Annex V, Part 1.13	CRD/CRR	8
European Banking Authority EBA	A European Supervisory Authority established by the Regulation (EU) 1093/2010	Regulation (EU) 1093/2010, Article 1	ESA regulations	9

Component 2: Repository of data elements

Data element designation: none

Data element definition: Gross carrying amount (as defined in ITS Annex V Part 1.34) of trade receivables (defined in ITS Annex V Part 2.85) included in the 'other than held for trading, trading assets or held for sale assets' (defined in ITS Annex V Part 1.15) accounting portfolio (defined in ITS Annex V Part 1.13).

Data element value: Numeric value, currency, scaling, precision.

Data element ID: 101

Component 3: Repository of reporting requirements

Reporting requirement designation: FINREP reporting requirement

Reporting requirement definition: Institutions (defined in Article 4(3) of CRR) subject to Article 4 of Regulation (EC) No 1606/2002 and credit institutions (defined in Article 4(1) of CRR) other than those referred to in Article 4 of that Regulation that prepare their consolidated accounts in conformity with the international accounting standards adopted in accordance with the procedure laid down in Article 6(2) of IAS, shall report financial information (as defined in Annex III of ITS) using prudential scope of consolidation (defined in ITS Annex V Part 1.12) as of the reporting reference date (defined in ITS Article 2 paragraph 1) with required reporting frequency (defined in ITS Article 11 paragraph 2).

Structure of the corresponding reporting requirement information in the reporting requirements component of the dictionary:

Reporting population: credit institutions and institutions that are publicly traded companies

Content of the report: financial information is a collection of templates (defined in Annex III, Part 1 of ITS) using methods of prudential consolidation (defined in Section 2 of Chapter 2 of Title II of Part 1 of CRR) as of the reporting reference date (defined in ITS Article 2 paragraph 1)

Recipients: competent authorities (defined in Article 4(40) of CRR)

Timing: FINREP reporting frequency (defined in ITS Article 11 paragraph 2)

Conditions: reporting entity prepares its consolidated accounts in conformity with the international accounting standards adopted in accordance with the procedure laid down in Article 6(2) of IAS**.**

Legal basis: REGULATION (EU) No 575/2013, Article 99 (2)

Reporting requirement ID: 1001

Component 4: Data catalogue

Reported datasets description: required datasets from FINREP reporting requirement

Underlying reporting requirement: FINREP reporting requirement (ID 1001)

Dataset owner authority: EBA (ID 9)

Collection of data elements from the data elements repository identified by their ID including the data element with ID 101. Potentially other information about data quality assessment performed on the data sets, how to access the datasets, the data model used for their storage.

Component 5: Relationships, assertions, transformations

The following are some relevant relationships between concepts (from the glossary but also from other components of the dictionary) that should be captured by the relationships component of the dictionary.

- ‘classification’ relationship, i.e. ‘is a(n)’ relationship:
    - ‘other than held for trading, trading assets or held for sale assets’ is an ‘accounting portfolio’
- ‘composition’ relationship, i.e. ‘is a part of’:
    - Data element ‘ID 101’ is a part of template ‘F05.01’
    - Template ‘F05.01’ is a part of 'financial information’

Component 6: Registration and administration information

The registration and administration component of the dictionary would hold the following information:

FINREP reporting requirement (ID 1001), created on date dd/mm/yyy, by EBA, is effective as from date dd/mm/yyy, and replaces the reporting requirement ID XXXX.

Such administration information would also be present for all the other concepts (from the glossary but also data elements, datasets, relationships, …).

Questions:

8. **Does the above example represent the information that should be included in the common data dictionary for a given reporting requirement and data element? If not, what should be added or removed? Is the information split appropriately across different potential components of the dictionary?**