A detailed look at the future of data standards for regulatory reporting

01   Introduction

In "The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger," author Marc Levinson explores how a seemingly simple innovation - the shipping container - revolutionized global trade and transformed the world economy. By standardizing the way goods were packed, shipped, and handled, this unassuming metal box unleashed unparalleled efficiency, cost savings, and widespread economic growth. Today, a new revolution is underway - one that is poised to make a similar impact on the financial world. As regulators seek greater transparency and accountability in reporting, data standardization emerges as the key to unlocking the transformative potential of regulatory reporting. Just as the shipping container brought order and efficiency to a chaotic industry, data standardization has the power to streamline and revolutionize the way financial information is collected, analysed, and shared.

Indeed, the last few years have seen an exponential growth in the amount and variety of data for regulatory purposes created by financial institutions. Without a coherent structure of data in corporate repositories, a lot of time and resources are lost cleaning and interpreting that data instead of using it. Hundreds of thousands of data points must be reported in an increasingly complex regulatory reporting system. It is now clear that supervisors and the industry are re-thinking the reporting model in light of increasing compliance requirements. This is evidenced by the numerous data standardisation initiatives underway (cf. Annexe). Furthermore, our research into the intricacies of Push vs Pull highlighted the pre-requisite of a granular data standard.

But data standardisation is a vague term, and a clear description of the objective is important. Data can come at many levels of granularity, and data models can come at many levels of abstraction for many different use cases. A data standard is merely an agreed format to exchange information such that there is as little ambiguity as possible in interpreting received data. It should also be clarified that open standards do not mean open data.

Given the large-scale systems of financial institutions, and increasing need for interoperability, standardisation can have wide ranging benefits. Particularly in the regulatory data space where comparability is a key criterion.

In this paper, we will attempt to present data standards at each level of the regulatory reporting process and a general blueprint for successful data standardisation, gathered from the expert advice and experience of data standards leaders in the financial and other sectors - hopefully overcoming the confusion caused by the many layers of data, transformations and processing systems involved.

02 The key components of a data standard

A. The dictionary

The key to a successful data standard lies in clear and precise definitions. It is essential to develop a robust data dictionary, especially when dealing with a standard that applies to different regions and countries. The data dictionary encompasses several key elements:

1. A methodology (taxonomy): This component outlines the approach and classification system used within the standard. It includes codes, identifiers (such as master data), and ensures consistent data representation. This needs to be objective and principle-based. If one thinks of the Linnaean Taxonomy it adopts an approach to classify mammals as "animals that suckle their young" and sticks to it, even when one turns out to be a whale.

2. A glossary: A comprehensive glossary is essential for defining and clarifying the terminology used in the standard. It encompasses a classification system and provides explanations for terms. It also includes codes and identifiers that contribute to maintaining data consistency.

3. Data structure definitions: These definitions are responsible for establishing a clear and well-defined framework for organizing and formatting data elements within the standard. They specify the relationships, format, and organization of the data, encompassing semantics, taxonomy, and relational tables.

By establishing a comprehensive and well-defined dictionary, data standardization projects can overcome challenges and increase their chances of success. An example of a data dictionary can be found in the European Banking Authority's (EBA) and European Insurance and Occupational Pensions Authority's (EIOPA) Data Point Model (DPM), which provides a standardized framework for reporting financial data in the banking and insurance sectors.

B. The exchange syntax

A defined syntax or format for data exchange, ensuring compatibility and interoperability between systems. This includes specifying the protocols, standards, and conventions for transmitting data between different entities.

C. Examples

Practical, real-world examples and how they would be used. Generation of examples by the standards organisation ensures that the standard works and is not too abstract or too difficult to implement. Examples help identify edge cases and considerations for scalability or usability and can also serve as precedents which can be referenced for future edge cases.

D. Other considerations

When considering the principles for a unique classification system, several factors should be taken into account:

1. Completeness: Ensuring that every product or entity can be classified within

one of the subclasses defined in the standard, leaving no room for ambiguity or gaps.

2. Unambiguity: Guaranteeing that each product or entity can only be classified in one specific subclass, avoiding overlapping or conflicting classifications.

3. Extensibility: Designing the core model in a way that allows for future extensions and modifications without necessitating changes to the core structure. This enables the standard to adapt and evolve over time.

4. Objectivity: Definitions and classifications should be achieved, as much as possible, through observable and precise facts rather than subjective definitions (e.g. "is highly liquid"). Imprecise facts inevitably lead to misclassifications and should be avoided. In some cases, this may not be possible but efforts should be made to add precision to the fact (e.g. "has a daily trading volume of at least $1mn")
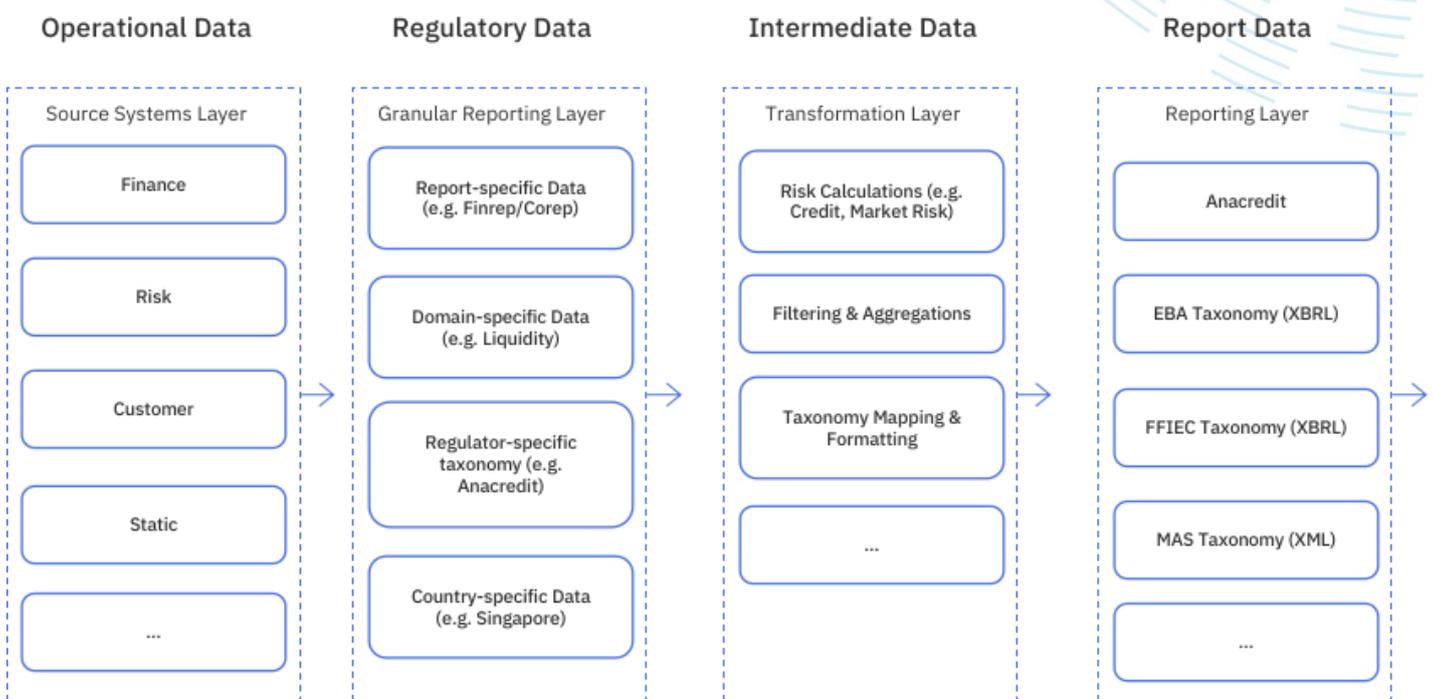
By adhering to these principles and considering the various purposes the standard aims to serve, a robust and versatile classification system can be developed, enhancing the effectiveness and applicability of the data standard.

03   Where are data standards needed?

In order to effectively discuss regulatory data standards effectively, we outline below the framework as the stages of data in the regulatory reporting process.

The 4 stages of regulatory reporting data

Operational Data	Regulatory Data	Intermediate Data	Report Data
Source Systems Layer	Granular Reporting Layer	Transformation Layer	Reporting Layer
Finance	Report-specific Data (e.g. Finrep/Corep)	Risk Calculations (e.g. Credit, Market Risk)	Anacredit
Risk	Domain-specific Data (e.g. Liquidity)	Filtering & Aggregations	EBA Taxonomy (XBRL)
Customer	Regulator-specific taxonomy (e.g. Anacredit)	Taxonomy Mapping & Formatting	FFIEC Taxonomy (XBRL)
Static	Country-specific Data (e.g. Singapore)	...	MAS Taxonomy (XML)
...			...

While these four stages are necessarily inherent in the process, they are not always well-distinguished within firms. Stages are sometimes merged, indistinguishable or inexistent. This leads to issues in the regulatory reporting process and confusion in the applicability of data standards regarding the stage of the process.

1. Operational Data is defined as the data from source systems which typically employ vendor-specific data models or product-specific standards (such as FPML or FIX). Data at this layer cannot (and should not) be fully harmonized as business requirements will differ between organizations,

systems and firms. Smaller firms will not have the same systems as larger firms or deal in the same products.

2. This data is then transformed in a Regulatory Data layer (often in a number of distinct data warehouses), where data can be later consolidated, normalized and enriched to transform it into the necessary regulatory semantics for various regulatory needs. A lack of standards at this Regulatory Data layer causes problems of reconciliation and other mismatches further down the road. This is where the regulatory reporting journey begins in earnest. Whereas Operational Data is used to serve customers, Regulatory Data is used to serve another customer, the Regulator. Regulatory Data is still highly granular however and remains very close to the operational data of the firm, with an almost one-to-one correspondence.

3. The Intermediate Data layer is where the complex calculations and regulatory logic takes place. This layer represents the process by which firms arrive at their final reportable values. Given the wide range of reporting obligations across firms, this layer is also difficult to standardize. The systems and intermediate data for credit risk reporting will be wildly different between a smaller, regional bank and that of a G-SIB reporting at many consolidation levels across many different jurisdictions. Data at this stage is not always stored as intermediate outputs for practical reasons but it should still conform to a defined schema so as to be reproducible, auditable and interrogable if needed.

4. The final reporting layer consists of producing the Reporting Data for submissions. This layer is often quite well-defined by regulators in a taxonomy-style dictionary supported with a rigorously defined submission format (e.g. XBRL, XML). Rather than a lack of standards, the most common complaint from firms was the existence of too many standards, each with their own implementation nuances and timelines. Greater harmonization amongst regulatory bodies was seen as the greatest benefit.

As data moves through these various stages, it undergoes transformations, calculations, and enrichment processes. Each layer of aggregation introduces its own set of complexities and challenges, which need to be addressed for effective data standardization.

For a concrete example, consider the reporting of "Retail" customer deposits in the Liquidity Coverage Ratio:

Operational Data: An operational system may record the amount held in an account at the end of the day, any pending transactions to come out/in, fees and interest that the account has incurred.

Regulatory Data: Regulatory calculations for Retail deposits reference a 'balance' or 'gross carrying amount' which must be deduced from the operational data above.

Intermediate Data: The computation for Retail involves first characterizing whether this classifies as a retail exposure for the purposes of credit risk as well as summing all a customer's balances across their deposits to see if it falls below 1mn EUR. So, this may yield intermediate data such as "exposure class" and "total balance EUR".

Reporting Data: Finally, based on the outcomes of the computation, the Reporting Data for "Retail" will contribute to the sum of all retail deposits reported in the appropriate field for the relevant taxonomy.

By considering this flow of data, we can see how data begins in a granular state and progressively gets aggregated to achieve the final reporting data items. From the example above, we can note that Operational and

Regulatory Data both describe data at a granular level while Intermediate Data may be at a slightly more aggregated level (1 total balance for many deposits) and Reporting Data will most definitely be at a much more aggregated level where a single figure may represent millions of deposits.

It should be noted, for "reports that look more like data" (such as Anacredit), this framework still holds true, with the caveat that the intermediate data stage might be unneeded when the Reporting Data is much more granular in nature.

The landscape of data standardization across the regulatory reporting process is characterized by numerous projects and initiatives that address different aspects of the problem. Below we have highlighted a number of data standards and their applicability across this spectrum in the regulatory reporting process.

Institutions should carefully consider the standards they are applying, their granularity and their and standardizing granularity levels, data mappings, and consolidation methods, organizations can ensure the accuracy, comparability, and integrity of data throughout the entire aggregation process. This enables reliable reporting and analysis, allowing stakeholders to make informed decisions based on consistent and trustworthy information.

Prudential Regulatory Reporting - Data Stages				
Data Standard	Operational (granular)	Regulatory (granular)	Intermediate (semi-aggregated)	Reporting (aggregated)
LEI	X	X	X	X
ISO (various)	X	X	X	X
ISIN	X	X	X	X
FIBO	X			
SWIFT MT	X			
ISDA CDM	X			
FIRE		X	X	
FIX	X			
EIOPA Taxonomy				X
EBA DPM Taxonomy				X
BIRD		X	X	
IREF				X
FPML				X
BoE Taxonomy				X
FFEIC Taxonomy				X
Anacredit				X

04 What are the key characteristics of a good standard?

It is clear from the table above that the financial industry does not suffer from a shortage of standards, but it does suffer from a lack of adoption or poor utilisation of existing standards with each new domain or jurisdiction creating a new standard. So, to better understand the characteristics of good standards, we interviewed numerous professionals who have initiated, contributed to and maintained standards, both from the above list, but also from other regulated industries such as healthcare, the environment and transportation in order to arrive at some common desirable characteristics of "good" standards.

A. Open-Source Access and Contribution

A standard should be open source in the modern sense of the word. It should be freely accessible on a public website, both by individuals and programmatically (machine readable via APIs). This minimizes hurdles to access, use and implement the standard while also maximizing the opportunity for contributions, enhancements and critiques to make the standard better. An open and inclusive community fosters collaboration, knowledge-sharing, and continuous improvement.

While it may be scary for some standards maintainers to provide this much transparency in a public domain, ultimately, it has been proven to lead to higher quality, more widely used and more robust standards. Open source also ensures portability. While GitHub is popular, it is important to acknowledge its status as a private company with its own interests. An open-source approach allows for a quick migration (by anyone) to another hosting platform.

Open-source goes beyond the standard itself, but also implies that the standard should not depend or require any proprietary standards, tools or

information. Open standards often serve as foundations for other open standards. The internet for example is built on layers of open-source standards such as HTML, CSS, Javascript, IEEE 802.11 WLAN.

B. Easy to use

A data standard should prioritize simplicity and clarity to ensure it is easily understandable and workable for both the public and private sectors. According to the Financial Stability Board, the technical approach used should not create barriers to comprehension. Minimizing technical complexities and dependencies is crucial for user-friendly data standard implementation. Avoiding third-party software and licenses reduces costs and compatibility issues, while leveraging open technologies and simple interfaces like text files empowers organizations without constraints. The example of GTFS (General Transit Feed Specification) illustrates the significance of simplicity in encouraging widespread adoption and implementation. By keeping the specification concise, GTFS became an internationally recognized and voluntarily adopted standard. The choice of the CSV format aimed to lower the participation barrier for smaller agencies with limited resources.

When developing a data standard, prioritize usability over theory. Striving for a perfect schema may not be practical due to real-world data limitations. Instead, focus on creating a pragmatic and user-friendly standard for financial institutions. Balance comprehensiveness with practicality to avoid burdening day-to-day operations. Prioritize core requirements and allow flexibility for optional elements to promote industry-wide acceptance.

Simplicity also plays a vital role in engaging the developer community in the standardization process, as seen in community-driven open-source projects like the FIRE schemas published on GitHub. Clear and uncomplicated data formats enable developers to easily understand and utilize the data, fostering the creation of innovative applications and services. This, in turn, cultivates a vibrant developer community around the standard.

Furthermore, simplicity promotes interoperability by facilitating the exchange and processing of data across different software systems. With a straightforward specification like GTFS, developers and transit agencies can seamlessly integrate data with various mapping platforms, trip planners, and other transit-related tools.

C. Easy to understand

Developing an effective data standard requires a clear and comprehensive narrative that articulates its purpose, target audience, vision for its usage and maintenance, and the specific types of data it aims to capture. This alignment fosters support, engagement, relevance, and adoption. Defining guidelines and protocols for usage and maintenance ensures sustainability and collaboration.

To ensure effective implementation, minimizing the need for interpretation is crucial when developing data standards. Precise definitions and reliance on undeniable external information help prevent misunderstandings and distinguish between facts and opinions. Clear rules benefit all stakeholders by increasing confidence in data accuracy and saving valuable time.

A standard should avoid introducing new terminology. A standard should be designed with a business language orientation, facilitating meaningful discussions with industry stakeholders. Understanding and focusing on the operating models of banks allows the standard to capture industry nuances accurately. By collectively agreeing on definitions and terminology that align with operational realities, the standard becomes more relevant and applicable. This collaborative effort enhances seamless data exchange, compatibility, and promotes interoperability across entities.

D. Practical

Providing documentation with examples that are clear, concise, and accessible to a wide audience ensures that users can navigate and

implement the standard effectively. To achieve this, it is crucial to include examples and case studies that further enhance understanding and showcase the practical value of the standard. By illustrating real-world scenarios and demonstrating how the standard can be applied, users can grasp its benefits more readily.

A good data standard must also be aligned with the interests of the data consumer, with a clear scope, problem that it is solving and practical considerations. Taking an industry-led approach allows a standard to align with business needs and existing industry practices, avoiding imposition by regulatory bodies. "The availability of more resources within the industry contributes to the success of this approach, with funding and governance structured to provide incentives," as mentioned by a technology leader at a US bank. Moreover, industry-led standards are more likely to gain widespread adoption, reducing fragmentation and confusion within the financial sector. The influence of peer pressure within the industry, demonstrated by FINRA in the US, can act as a powerful driving force.

A key characteristic of good data standards is their ability to adapt and evolve in response to changing business needs. It is crucial to develop a strategy for legacy data standards that takes into account the evolving landscape and executes necessary updates accordingly. By proactively planning for periodic revision cycles, you create a framework that allows for regular review and ensures that data standards remain up-to-date. This iterative process allows for adjustments and enhancements based on emerging trends, technological advancements, regulatory changes, and evolving business requirements.

In the realm of data standards for the financial industry, the pursuit of a global approach is of utmost importance. Therefore, this involvement should extend beyond national borders to include contributions from a global user community, such as civic hackers, other cities, and larger vendors. Collaborating with a wide audience of global stakeholders ensures scalability and maintains the standard's neutrality and global applicability.

By establishing a global standard and providing a robust public infrastructure, we can create an environment that promotes trust, reliability, and seamless interactions among stakeholders. This approach acknowledges the importance of harmonizing efforts on a global scale and leveraging technology to streamline processes, ensuring a consistent and efficient exchange of data in the realm of regulatory governance.

E. Governed by principles

Through the course of our interviews, we came across many different governance frameworks. Committee-driven, founder-led, hierarchical, democratic, competitive or compromised-based. Most successful projects, however, fell into two categories: Committee-driven or founder-led. When one considers the world's most successful standards, there are committee-driven projects such as ISO or W3C as well as founder-led projects such as Linux, Python. While these may seem like radically different approaches, they share one key aspect: A strong adherence to a set of principles. Whether these principles are set and enforced by a committee, or they are set and enforced by an individual does not seem to matter. The importance is to have a clear set of principles. Through regular enforcement of the principles, the project will set precedents for future contributions such that the individual or committee members can be replaced over time while still adhering to the same contribution principles.

It goes without saying, via the open-source requirement above, that the success of a standardization project relies on the involvement of diverse stakeholders from the industry, including experts and representatives from various segments. This inclusive approach ensures that the standard addresses the needs of all parties involved and promotes accurate and unbiased definitions (ISO, 201812). To ensure widespread adoption and utilization of the standard, it is crucial to engage the entire user base, including publishers, software developers, and end-users, throughout the development process. This collaborative effort guarantees that the standard effectively meets the requirements of all stakeholders. Many regulatory

standards exclude software vendors from contributing or commenting on standards. This is a mistake as the vendors are typically the ones implementing standards as can be seen from a standard such as XBRL, where nearly every firm uses a dedicated software provider. It is true that vendors may be acting in their own best interest, but again, an open-source approach will expose those vendors to criticism from other vendors and ultimately principles should be designed that will automatically reject a vendor-biased contribution.

Much like vendors, banks, too, are private sector organisations and contributors from firms should also be considered to be acting in the best interests of their firm. Therefore firm-specific and firm-biased contributions should also be ruled out using effective principles.

Principles are often hard to define upfront and state clearly, which is why many successful standards began life as a small, focused team, characterized by minimal turnover, who over time established a set of documented principles for contributions. Effective communication, active participation, and efficient decision-making are best facilitated by a small team. Collaboration enhances understanding of use case requirements, leading to a tailored standard. A small team enables rapid iteration and refinement, benefiting from streamlined feedback loops and agile development. Ensure team members possess diverse perspectives and relevant expertise. Former regulators we interviewed emphasized the necessity of effective leaders to steer action towards better and more efficient solutions.

F. Widespread adoption

The goal of any data standard is widespread adoption, and the level of adoption has a compounding effect where greater adoption fuels greater adoption. Depending on the domain and level of resources at the start, this may take many years. Getting on this path can be even more difficult when one of the above key characteristics are lacking or ignored.

Throughout the interviews, it became clear that adoption of voluntary standards proved more beneficial in reducing costs and serving as foundations for further standards and technologies, as opposed to imposed standards. Examples of voluntary standards in finance (e.g. FPML, LEI, ISO standards like dates, currencies, countries) have led to endless technologies built on top whereas mandatory standards (e.g. XBRL or SDMX) have remained niche use cases that have not seen adoption outside their sole purpose or as foundations for other technologies. It should also be recognised that without mandatory standards, regulatory harmony is not possible and therefore, voluntary widespread adoption may not be the goal of every standard. A central bank economist aptly stated that "a bad standard is often better than no standard." Finally, voluntarily adopted standards that are then made mandatory once reaching a critical mass are often the most effective as through many existing live implementations, the community, skills and related tools are all widely available and understood.

05 Conclusion

Standards are not always desirable or necessary everywhere but are absolutely critical elsewhere. Therefore, to achieve effective standardization in the regulatory reporting process, it is recommended to pursue a dual approach consisting of two layers of standards at the Regulatory Data and Reporting Data layer:

1. A single Regulatory Data standard

The first layer involves a granular Regulatory Data standard for data collected by firms, which should be developed by the industry with regulators providing feedback and guidance. This approach allows the industry to take the lead in defining the data elements and structures that are most relevant to their operations, ensuring these data items can be

practicality sourced, maintained and delivered in alignment with industry needs and capabilities. By involving regulators in the process, their expertise can be leveraged to ensure compliance and address any potential concerns about how this data may meet the needs of the ultimate Reporting Data layer. Additionally, regulatory involvement is beneficial to clarify any semantic terminology or edge cases that inevitably arise with complex financial products and regulations. A successful example of such a standard is FIRE, which has been developed by the industry with support from the European Commission. The FIRE standard is now actively used by dozens of firms around the world to represent their Regulatory Data and transform it into Reporting Data.

2. A single Reporting Data standard

The second layer entails a standard for Reporting Data, which is the responsibility of regulators to define with industry providing feedback and guidance. A member of the Bank of England's Transforming Data Collection team stated "that this is the data the Bank needs, therefore it is their right and responsibility to define it." A standard at this level, such as the EBA's Data Point Model, focuses on the specific data requirements for reporting purposes. It is important to keep these two levels of standardization distinct to maintain clarity and avoid unnecessary complexity. This allows each standard to be refined independently, addressing the unique needs and considerations of data collection and reporting, respectively. As discussed, many, many standards exist at this level for reporting purposes but greater harmony is needed amongst them to avoid overlaps and confusion. The IReF project, from the European Central Bank, is a strategic effort to harmonize the existing Reporting Data standards for statistical reports within the EU and is expected to go live in 2027. While this harmonization has been welcomed by firms, many have expressed the feeling that the harmonization has not gone far enough and more cross-border cooperation of regulatory bodies to standardize forms and requirements would be welcome.

By adopting this dual-level approach, the industry can have greater control and flexibility over the granular data they need to collect and maintain while ensuring alignment with regulatory requirements. Meanwhile, regulators can also maintain control of the data needed for effective supervision with feedback from firms and third parties. This collaboration between industry and regulators will foster a more efficient and effective standardization process, leading to improved data quality, streamlined reporting, and enhanced regulatory compliance.

Endgame

Promoting robust standards at the two ends of the regulatory data journey will foster growth and innovation throughout the remainder of the process. The development of mature standards like HTML, CSS and Javascript led to many web-based applications. Over time, intermediate web frameworks, open source libraries, components, open APIs and design systems have emerged to fill the gaps in between that have reduced costs and improved efficiency of web development. Similarly, one can envisage a future where the development of these two streams of Regulatory Data and Reporting Data standards in an open and collaborative fashion will ultimately lead to a growth in related tools and technologies (hopefully also free and open source) producing solutions around the Intermediate Data, regulatory calculations, analytics and more. Without the adoption of these standards, the industry will struggle to enable higher-level automation through the use of modern AI and machine learning tools.

Global standards for regulatory reporting offer numerous benefits. They streamline processes, improve data quality and accuracy, enhance transparency, and strengthen regulatory oversight. It fosters greater comparability and consistency in reporting, enabling more effective risk management, monitoring, and analysis. Moreover, it promotes efficiency in data collection and analysis, reducing costs and efforts associated with reconciling disparate data formats and structures.

Therefore, achieving global standards for Regulatory and Reporting Data is not just a desirable goal but a necessary one for the proper and efficient management of the financial industry. Standardization should not be treated as a luxury that can only be pursued during ideal financial situations or when banks have surplus resources to allocate towards it. On the contrary, it is crucial to prioritize standardization efforts using the available means as soon as possible.

NB: It should be noted that the term regulatory reporting in this paper refers to prudential and statistical regulatory reports and not trade/transaction reporting.

Special Thanks

This whitepaper is the product of 18 interviews conducted during June 2023. We would like to extend our special thanks to the members of our Regulatory Council, Bill Coen, Bill Dudley and Lyndon Nelson and the many other individuals interviewed within the financial industry and public sector institutions that have contributed their thoughts and insights into this topic.

About Suade Whitepapers

Suade is a London-based RegTech organisation that is a leading regulatory reporting software provider for banks and financial institutions of all sizes. Suade Whitepapers aim to address a key theme or topic within the RegTech industry from a neutral, research-driven lens by leveraging the wisdom of our community. In order to achieve a balanced view, the Suade team typically conducts off-the-record interviews with practitioners, senior managers from the financial industry, public sector and other professionals experienced in the topic to inform the publication. The goal is to publish a considerate, educational, technical publication to drive further discussion via roundtables, webinars and other events towards achieving a more stable and transparent financial system.

If you would like to participate in associated events or future whitepapers, please get in touch at: research@suade.org.