Compliance Centric Data Quality Management – The Banking and Financial Industry Perspective

Full Paper

Faisal Khan
School of Information Technology and Electrical Engineering
The University of Queensland
Queensland, Australia
Email: faisal.khan@uq.edu.au

Marta Indulska
UQ Business School
The University of Queensland
Queensland, Australia
Email: m.indulska@business.uq.edu.au

Shazia Sadiq
School of Information Technology and Electrical Engineering
The University of Queensland
Queensland, Australia
Email: shazia@itee.uq.edu.au

Abstract

The banking and financial industry is subject to a large number of regulations, including several recent data quality regulations, which makes it difficult for organisations to judge the effort and cost involved in compliance to multiple, often overlapping, regulatory obligations. The aim of this paper is to identify and analyse data quality related regulations, and map them against a common set of data quality dimensions to expose overlaps and inconsistencies to inform compliance efforts. In our study we identify seven global data quality regulations/frameworks applicable to the banking and financial industry and conduct a systematic analysis of data quality stipulations within. Our study explores the breadth and depth of coverage of data quality dimensions in the regulations, and identifies inherent overlaps and inconsistencies. We argue that understanding of data quality requirements within and across the regulations is an essential first step towards the design of new approaches for compliance centric data quality management.

Keywords Data Quality, Banking regulations, Compliance, Risk

1 INTRODUCTION

The shifting, and often vanishing, sectoral and geographical boundaries are having a profound change on business. While these changes offer a myriad of business opportunities, they also render businesses liable to deal with multiple regulators (Currie et al. 2018). The effort required to address regulatory requirements, and the cost of non-compliance, can make it difficult to harvest the benefits of the emerging opportunities. At the same time, the data-driven global economy has influenced organisations to change the way they operate. Due to heavy reliance on data for most organisatonal activities, data has become one of the most important organisational assets (Chen et al. 2012). Together with advancement in computational capacity and intellegent systems, organisations are increasingly motivated to become and remain more 'data-ready' Sadiq (2013). As organisations create and aquire increasing volumes of data, problems associated with the quality of data are also increasing. Data Quality (DQ) has, over the years, been defined in various ways. One widely accepted way to define it is as: "the fitness of the data for the intended purpose" per Juran and Goodfrey (1999). Such a definition encompasses a number of so called DQ 'dimensions', such as accuracy, availability and accessibility, completeness, consistency, currency, reliability and credibility, validity and others (Jayawardene et al. 2015).

While many businesses have made efforts to achieve good data quality to realise some of these business opportunities, regulatory compliance has not been the focus of data quality management efforts because few regulations were in place until recently. Regulations related to data quality have, however, now been enacted in several industries globally, with challenging timelines and requirements. The banking and financial industry, due to sensitive nature of data it deals with, has been the target of a number of regulators as well as other government agencies (Currie et al. 2018).

Incidents resulting from poor quality of data can harm organisations considerably (Redman 1998). For example, the ANZ bank in 2014 had to refund $70 million to its 235,000 home loan customers. The root cause was found in overcharged interest as agreed discounts or offset accounts had not being linked to customer mortgage accounts. ASIC played its supervisory and regulatory role and in addition to the amounts paid to the customers, ANZ spent an additional $4 million to improve their data and systems related capabitlies (ABC 2014). In a similar instance, which took place in 2013, Bank of Queensland was forced to refund its customers $12M in overcharged interest. Such errors had been attributed to a combination of manual and automated processes, including data aggregation practices (ASIC 2013). A comparable recent example is that of Westpac, who was forced to repay tens of millions of dollars after incorrectly charging 40,000 customers. Westpac confirmed that the matter was related to their data related capabilities (7News 2019). It is evident from these examples that DQ management is becoming increasingly important for the banking and financial industry. The growing impact of DQ on public interests has motivated governments and financial industry regulators globally to enact regulations related to data, including DQ, such as The Solvency II (EU 2009), Principles for effective risk data aggregation and risk reporting (BIS 2013), U.S Dodd Frank Act (USC 2010) and Managing Data Risk-Prudential Practice Guide (APRA 2013).

However, current DQ management approaches generally consist of a set of guidelines and techniques that, given a set of stated requirements elicited from data users within a given application context, define a rational process to assess and improve the quality of data (Shewhart and Deming 1986). DQ requirements are thus typically user centric, and DQ management processes can therefore be largely segregated from compliance obligations stemming from regulations related to data. Accordingly, the aim of this paper is to explore the opportunity and potential for creating compliance centric DQ management approaches, i.e. those targeted on meeting prioritised compliance objectives, as opposed broad approaches that aim at addressing all user driven DQ requirements with less prioritisation of compliance, which may lead to failure in some cases and slow progress in others.

As a first step towards a compliance centric data quality management approach, there is a need to analyse DQ related regulations in a way that respective compliance obligations can be compared, which is what we aim to do in this paper in the context of the banking and financial industry. To do so, we identify and analyse DQ regulations focused on the banking and financial industry, with additional analysis of sector-neutral DQ frameworks, to uncover the commonalities and variations. We use a common but comprehensive set of definitions of DQ dimensions, which is based on a systematic review of DQ literature (Jayawardene et al. 2015), against which DQ regulations and frameworks, selected for this study, were analysed.

In the remainder of this paper we first provide the necessary background on DQ dimensions and relevant regulations and frameworks. We then outline our research approach, followed by analysis of the selected regulations and frameworks against the set of DQ dimensions (Jayawardene et al. 2015). We conclude the paper with a summary of outcomes of this study and discussion on next steps towards compliance centric data quality management.

2 BACKGROUND AND RELATED WORK

DQ has been defined by different experts in multiple ways, often abundant with commonalities and deviations. This is perhaps not surprising as researchers and practitioners with particular interest in different areas and industries look at data quality dimensions differently, and rate their importance differently, based on the impact in their industry or research interest Garvin (1987) and Sadiq (2013). While a simple and high-level "fitness for use" definition exists, as described by Juran and Godfrey (1999), decades of data quality research have proliferated various understandings of data quality through its underlying dimensions.

However, a recent systematic consolidation of the definitions and dimensions (Jayawardene et al. 2015) indicates that data quality is primarily a reflection of: accuracy, availability and accessibility, completeness, consistency, currency, reliability and credibility, usability and interpretability and validity of the data. Each of these dimensions in turn has a number of recurring patterns indicated through the type of metrics and implementations. For example, validity consists of business rules compliance, meta-data compliance, standards and regulatory compliance, and statistical validity. The full list of DQ dimensions, their constituent patterns and corresponding definitions are shown in Table1.

Dimension	Pattern	Description
Completeness	Completeness of mandatory attributes	The attributes which are mandatory for a complete representation of a real world entity must contain values and cannot be null
	Completeness of optional attributes	Optional attributes should not contain invalid null values
	Completeness of records	Every real world entity instance that is relevant for the organization can be found in the data
	Data volume	The volume of data is neither deficient nor overwhelming to perform an intended task
Availability & Accessibility	Continuity of Data Access	The technology infrastructure should not prohibit the speed and continuity of access to the data for the users
	Data maintainability	Data should be accessible to perform necessary updates and maintenance operations in its entire lifecycle
	Data awareness	The data users should be aware of all available data and its location
	Ease of data access	The data should be easily accessible in a form that is suitable for its intended use
	Data Punctuality	Data should be available at the time of its intended use
	Data access control	The access to the data should be controlled to ensure it is secure against damage or unauthorized access
Currency	Data timeliness	Data which refers to time, should be available for use within an acceptable time relative to its time of creation
	Data freshness	Data which is subjected to changes over the time should be fresh and up-to-date with respect to its intended use
Accuracy	Accuracy to reference source	Data should agree with an identified source
	Accuracy to reality	Data should truly reflect the real world
	Precision	Attribute values should be accurate as per linguistics and granularity

Dimension	Pattern	Description
Validity	Business rules compliance	Data must comply with business rules
	Meta-data compliance	Data should comply with its meta-data
	Standards and Regulatory compliance	All data processing activities should comply with the policies, procedures, standards, industry benchmark practices and all regulatory requirements that the organization is bound by
	Statistical validity	Computed data must be statistically valid
Reliability	Source Quality	Data used is from trusted and credible sources
	Objectivity	Data are unbiased and impartial
	Traceability	The lineage of the data is verifiable
Consistency	Uniqueness	The data is uniquely identifiable
	Non-redundancy	The data is recorded in exactly one place
	Semantic consistency	Data is semantically consistent
	Value consistency	Data values are consistent and do not provide conflicting or heterogeneous instances
	Format consistency	Data formats are consistently used
	Referential integrity	Data relationships are represented through referential integrity rules
Usability and Interpretability	Usefulness and relevance	The data is useful and relevant for the task at hand
	Understandability	The data is understandable
	Appropriate Presentation	The data presentation is aligned with its use
	Interpretability	Data should be interpretable
	Information value	The value that is delivered by quality information should be effectively evaluated and continuously monitored in the organizational context

Table 1. DQ dimensions, patterns and corresponding definitions (Jayawardene et al. 2015)

In the context of the banking and financial sector, evidence suggests that the industry is responsive when it comes to enactment and enforcement of DQ regulations. For example, a number of DQ related stipulations are present in regulations such as 'Principles of effective risk data aggregation' (BIS 2013) and 'Solvency II' (EU 2009). The motivation of regulatory bodies to introduced mandatory compliance requirements for managing the quality of financial data is not surprising given significant financial investments, the multi-billion dollar exposure in the financial sector and the earlier discussed impact on public interests if errors arise. It is pertinent to mention that the Bank for International Settlements, which serves as a bank for central banks, in 2013 issued above mentioned regulation (BIS 2013) for the banking and financial industry in the wake of the 2007-08 banking crisis. The move came as a result of indications that most of the failed banks had ineffective capabilities for processing and aggregating their data, including risk related data. Indeed, in recent years, we have observed that banking and financial industry institutions are subject to an increasing number of regulations related to data. This situation implies that one organisation may be required to comply with multiple, possibly overlapping, regulations (Currie et al. 2018). However, at this time, there is a lack of a shared understanding of data quality dimensions and metrics of interest in this context. This situation, in turns, makes it difficult to ascertain the overlap and differences within and across regulations, creating a challenge for organisations planning their compliance activities to better respond to multiple regulatory requirements and to identify what controls might be missing.

Specifically, there are three regulations and one regulatory guide that feature prominently in this context, namely: the 'Solvency II' (EU 2009) legislative program, introduced by the European Union in

2009 with a gradual adoption timeline stretched over several years, which establishes specific data quality compliance requirements for insurance companies in the European Union, 'Principles for effective risk data aggregation and risk reporting' (BIS 2013), issued by the Bank for International Settlements in 2013, which aims to improve banks' ability to aggregate risk data and recommends a robust data framework to anticipate problems and 'U.S Dodd Frank Act' (USC 2010) enacted in 2010, which aims to protect customers by improving accountability and transparency. It should be noted that there was a partial roll back of Dodd Frank Act in May 2018 (NYT 2018). However, the data and data quality related content remains part of this law. In addition, as mentioned above, a guide from a regulator to advise on compliance namely 'Managing Data Risk-Prudential Practice Guide' (APRA 2013) - was issued in 2013 by APRA (Australian Prudential Regulation Authority). It aims to assist regulated entities in managing data risks, including risks to quality of data.

Furthermore, many sectors and governments globally have issued guidelines and frameworks for DQ, which are not mandatory to comply with, but relate to the growing requirements of standardisation in managing and reporting of the data. These include 1- 'Australian Bureau of Statistics 'Data Quality Framework' (ABS 2009), issued in 2009, which defines standards for assessing and reporting on the quality of statistical information. 2- 'Statistics Canada Quality Guidelines' (SC 2009) issued in 2009, which aims to ensure that information is relevant and of high quality. 3- 'European Statistics Code of Practice' (ES 2007) issued in 2007, which aims to introduce, systematise and improve DQ management. Several other frameworks and regulations, such as Bank of England's Data Quality Framework, for example, have addressed requirements and/or guidelines related to DQ management in the financial sector. Keeping in view the history of financial sector regulatory regimes, we anticipate that some of these voluntary guidelines and frameworks will become mandatory regulations and laws in the future.

3 APPROACH

Our aim in this paper is to provide a mapping between the data quality domain and banking regulations and frameworks that relate to DQ, and subsequently to use the mapping to create an objective and canonical view of their inter-relationships. To ensure the repeatability of our work, we have developed a rigorous, four step approach to undertake the mapping, as explained in the following subsections.

3.1 Identification of a comprehensive DQ reference source

To conduct our research we needed a comprehensive set of data quality dimensions that could be used as a basis of the comparison with regualtions and frameworks. Accordingly, we reviewed data quality research with the view to identifying such a source. Over the last two decades, researchers and practitioners have suggested several valuable classifications of data quality characteristics such as Loshin (2001) and, Redman (1998). However, over the course of time, many of the definitions for different data quality dimensions, or underlyng concepts, have overlapped, and some definitions have developed conflicting interpretations. Thus, we observed that DQ definitions have regressed to a level of disparity that does not support a shared understanding of the core knowledge of the discipline. For this reason we focused our analysis on prior research that offered consolidations of DQ approaches and dimensions. Through this process we identified a relatively recent study (Jayawardene et al. 2015), which was informed by a large number of seminal and also recent studies in the area of data quality, such as the works of Wang and Strong (1996), Kahn et al. (2002) Price and Shanks (2004) and Sadiq (2013). These studies cover various definitions of data quality dimensions and patterns as well as various academic and practical approaches to achieve, maintain and improve data quality. For this reason, we consider (Jayawardene et al. 2015) to offer a comprehensive consolidation at the level of required granularity for our analysis – i.e. it identifies thirty three (33) specific DQ patterns that address eight (8) DQ dimensions as presented in Table-1 above. This reference source is generic in terms of its coverage and can be applied in any industry, including banking and finance. Accordingly, we adopt the work of (Jayawardene et al. 2015) as the baseline of DQ dimensions.

3.2 Selection of relevant DQ regulations and frameworks

We analysed financial industry regulations to ensure their intent is mainly to address DQ capabilities of the financial institutions. As discussed in section 2, our focus in this study is on regulations for the banking and financial industry specifically, however, for generalisability purposes, we also examined data quality frameworks issued by other government bodies. We initially analysed overarching banking regulations such as the Basel series, however, we found them to be broad in terms of their coverage, which encompasses market risks to the global banking and financial industry. Nevertheless, one of the key regulations of our study, i.e. 'Principles of effective risk data aggreggation' (BIS 2013), is a subset of the Basel framework (it was issued by Basel Committee on Banking Supervision). Given our selection criteria i.e. the intent of the document is to ensure quality of data being gathered, generated, stored, processed and used by the target audience of the document, we identified four regulations: 1-'Principles of effective risk data aggreggation' (BIS 2013), 2-'Solvency II' (EU 2009), 3-'Dodd Frank Act' (USC 2010), 4-'Prudential Practice Guide' (APRA 2013), and three non-regulatory guidelines and frameworks: 5-'Statistics Canada Quality Guidelines' (SC 2009) 6-'European Statistics Code of Practice' (ES 2007), and 7-'Data Quality Framework' (ABS 2009) to construct our study upon.

The above selected regulations and frameworks provide, if not a complete then a significant footprint of the finanical and banking industry's compliance obligations with respect to data (quality) management. Full texts of the above-mentioned documents were reviewed to ascertain whether there is relevant and sufficient content pertaining to data quality to include the relevant regulation or framework in our study.

3.3 Coding

After selecting the seven sources and checking them for relevance, we mapped their content to the 33 DQ patterns from our reference source (Jayawardene et al. 2015 – see summary in Table-1). We carried out the mapping by analysing every clause, requirement or definition stipulated in the selected resources and considering the most accurately corresponding data quality pattern from our reference source. The patterns were assigned a sequential number for identification and referencing purposes. To maintain integrity and authenticity of the study, two researchers, independently of each other, mapped the data quality dimensions and requirements stipulated in the regulations, guidelines and frameworks to the patterns listed in our reference source. While performing the mapping, the lead researcher considered the base documents as well as associated guidance and supplemental resources (such as, for example, the Committee of European Insurance and Occupational Pension Supervisors (CEIOPS)'s 'Advice for Level 2 Implementing Measures' on Solvency II (Van Hulle 2011) and similar guides for other regulations produced by independent industry organisations or individuals) to accurately capture the meaning and essence of the stipulated requirements.

To ensure consistency of mapping throughout the study, scoring criteria were established to measure the level of commonality (see Table 2). The scoring criteria were designed to capture not just alignment to the DQ patterns, but also potential issues with lack of consistency of terms - for instance if a particular DQ requirement is mentioned in the regulation or framework but its label is different than that of our selected reference source then we show the coverage with corresponding shade of grey. Thus the darker the corresponding cell is, the more adequate the coverage of that particular DQ pattern in a regulation. If labels mentioned in the reference source and the relevant regulation also match, this represents the highest level of commonality and is in addition represented by a white square.

Scoring Criterion	**Graphical Representation**
Pattern is neither mentioned directly nor covered indirectly.	
Pattern is not explicitly mentioned however there is some evidence of consideration.	
Pattern is mentioned but lacks details required to fully understand its intent and purpose.	
Pattern is mentioned however definition and detail differ from those stipulated in the selected reference source.	
Pattern is mentioned with adequate detail, and the definition matches with the one stipulated in the selected reference source.	

Table 2. Representation of Pattern Coverage in Regulations and Frameworks

At the conclusion of the coding process, we calculated Cohen's Kappa to evaluate the level of agreement between the results of the two independent coders. The level of agreement was 89%, indicating a high level of agreement, as explained by Wood (2007). Further, to consolidate the coding of the two researchers we followed the process of validation used by Abdullah et al. (2012) and Ledford et al. (2013). That is, we conducted two further independent iterations by two additional researchers to consolidate the mappings and reach a final agreement. The results shown in Table 3 represent the final agreed results.

4 RESULTS AND ANALYSIS

Table 3 provides a summary of the analysis of DQ coverage in the seven selected sources against the thirty three (33) DQ patterns and the eight (8) associated DQ dimensions, following the legend provided in Table 2. In addition to the summary, we uncovered insights on the coverage and consistency of data quality requirements within the regulations and frameworks. Due to space limitations, we only provide some example insights below.

DQ Dimension	DQ Pattern	Regulations and Frameworks						
		1	2	3	4	5	6	7
		Coverage						
Completeness	Completeness of mandatory attributes							
	Completeness of optional attributes							
	Completeness of records							
	Data volume							
Availability & Accessibility	Continuity of data access							
	Data maintenance							
	Data awareness							
	Ease of data access							
	Data punctuality							
	Data access control							
Currency	Data timeliness							
	Data freshness							
Accuracy	Accuracy to reference source							
	Accuracy to reality							
	Precision							
Validity	Business rules compliance							
	Meta-data compliance							
	Standards and regulatory compliance							
	Statistical validity							
Reliability	Source quality							
	Objectivity							
	Traceability							
Consistency	Uniqueness							
	Redundancy							
	Semantic consistency							
	Value consistency							
	Format consistency							
	Referential integrity							
Usability and Interpretability	Usefulness and relevance							
	Understandability							
	Appropriate presentation							
	Interpretability							
	Information value							

Table 3. DQ coverage scores of regulations and frameworks analysed in the study

1-Principles for effective risk data aggregation – BIS (Switzerland)
2-Solvency II (European Union)
3-Dodd Frank Act(US)
4-Prudential Practice Guide – Managing Data Risk (Australia)
5-Statistics Canada Quality Guidelines
6-European Statistics Code of Practice
7-ABS Data Quality Framework (Australia)

For ***completeness*** we found that sector-neutral frameworks have not covered this dimension well, and whereas completeness is mentioned in all regulations in one way or the other, the focus has remained on *data volume* and *completion of records* patterns, while patterns such as *completion of attributes* have not been mentioned. For example, 'Principles of effective risk data aggregation' (BIS 2013) covers completeness of records by stating *"Bank should be able to capture and aggregate all material risk data across the banking group"*. ***Availability and accessibility*** has wide coverage, with *continuity of data access*, *ease of data access*, and *data punctuality* being most common. For example, the 'Data Quality Framework' (ABS 2009) mentions accessibility, by stating: "*Accessibility is a key component of quality as it relates directly to the capacity of users to identify the availability of relevant information, and then to access it in a convenient and suitable manner"*. The ***currency*** dimension is covered in 5 out of the 7 documents analysed. Although the description varies significantly e.g. 'European Statistics Code of Practice' (ES 2007) states: *"Timeliness of information reflects the length of time between its availability and the event or phenomenon it describes*." whereas Principles for effective risk data aggregation (BIS 2013) states: *"Timeliness – A bank should be able to generate aggregate and up-to-date risk data in a timely manner"*. **Accuracy** is mentioned in all of the analysed documents although with significant variations. Some examples include 'Solvency II' (EU 2009): *"Data is considered to be accurate if it is free from material mistakes, errors and omissions*." and the 'Data Quality Framework' (ABS 2009): *"Accuracy refers to the degree to which the data correctly describe the phenomenon they were designed to measure"*, a definition which matches well the *accuracy to reality* pattern. For the **validity** dimension we found that the focus of regulations is on the compliance related patterns of validity, while the focus of frameworks is on *statistical validity* pattern. **Reliability** was not sufficiently addressed in any of the resources analysed in our study. One document makes mention of reliability while discussing timeliness, but not as a requirement itself. The **consistency** dimension is better addressed in regulations than in frameworks, which use the term coherence that, to some extent, aligns with this dimension. For instance, the Data Quality Framework (ABS 2009) defines coherence by stating: *"For managing coherence, collection agencies should use standard frameworks, concepts, variables and classifications, where such are available, to ensure the target of measurement is consistent over time and across different collections"*. ***Usability and interpretability*** has received good overage with all of the five patterns mentioned in at least one document. For example the 'Data Quality Framework' (ABS 2009) defines it as: *"Interpretability refers to the availability of information to help provide insight into the data"*. We note that the 'Dodd Frank Act' (USC 2010) briefly mentions accuracy and availability but provides no information on what these terms mean, nor mentions any of the other DQ dimensions, hence we were not able to associate it with a DQ pattern. For this reason, Table 3 does not show any DQ coverage pertaining to the 'Dodd Frank Act' source.

Our analysis aims to uncover commonalities and possible redundancies in DQ compliance efforts across multiple regulations. Let us take the case of a specific organisation: Societe Generale to demonstrate how our results can be used. Societe Generale provides its banking as well as insurance services in diverse geographical locations such as Europe, USA, and Australia. To do so, it has to comply with regulations of the banking and the insurance industry in Europe, USA and Australia. Our analysis indicates that by first aligning with the 'Principles for effective risk data aggregation' (BIS 2013), which provides the widest coverage of DQ dimensions, Societe Generale could highlight most DQ related gaps. Through the identified overlaps with other regulations they could also see what aspects of other regulations they potentially comply with, and what other gaps need to be addressed to ensure compliance. Such knowledge on the priority order of implementation may help direct organisational resources in way that reduces the overall compliance effort.

In addition, our study also highlights the need to provide clear and explicit definitions for DQ requirements as well as the need for a shared sector-wide understanding of DQ requirements and dimensions. We found a lack of explicit definitions in a number of regulations we analysed. For instance, the 'Dodd Frank Act' (USC 2010) requires regulated entities to ensure accuracy and integrity of information that is to be made available to stakeholders. However, these terms are not explicitly defined in the act. Given the proliferation of various definitions of data quality dimensions over the years, this situation may lead to different interpretations of accuracy and integrity by different entities. Moreover, the act does not cover completeness, consistency, validity, or currency aspects of DQ – arguably important to other regulations and frameworks - see Table 3. Similarly, the 'Data Quality Framework' (ABS 2009) does not address completeness or any synonymous dimension. The closest reference in the framework is to that of coherence, which attempts to address completeness in an alternate way by stating: "Coherence is an important component of quality as it provides an indication of whether the dataset can be usefully compared with other sources to enable data compilation and comparison." While we found that the selected regulations and frameworks differ in their explicit requirements, their explanatory notes provide some of the missing details. However, explanatory notes (issued by the issuer of the original regulation) are not always available (e.g. as is the case for the 'Dodd Frank Act' (USC 2010)), therefore there is an inherent risk that compliance efforts at different organisations may have variable results for the same regulation.

Further, the disparity in the use of the same DQ dimension and/or explanation leads to a lack of shared understanding, which in turn can result in duplication of work, inconsistent results of compliance from different organisations, and inability of the community to share best practice. For example, 'Dodd Frank Act' (USC 2010) defines accuracy as: "...the degree to which the data correctly describe the phenomenon they were designed to measure." Meanwhile, 'Solvency II' (EU 2009) defines accuracy as: "Data ... is free from material mistakes, errors and omissions." As another example, 'Prudential Practice Guide' (APRA 2013) limits the scope of timeliness by defining it as "the degree to which data is up-to-date", while 'Principles for effective risk data aggregation' (BIS 2013) extends its scope by stating: "A bank should be able to generate aggregate and up-to-date risk data in a timely manner while also meeting the principles relating to accuracy and integrity, completeness and adaptability." Accordingly, 'Principles for effective risk data aggregation' (BIS 2013) appears to stress that data will only be considered timely if it meets other quality requirements simultaneously.

More broadly, on the basis of our research we argue that for enacting regulations associated with data, due to continuous developments and rapid changes in this area, a broader consultation encompassing industry as well as academia is required to develop effective regulations. A relevant example in this regard is Australian Prudential Regulation Authority, which took over a year to complete several rounds of consultation with stakeholders before finally transitioning their information security guide to a mandatory compliance standard (namely CPS-234). An analysis of the consultation details reveals that industry was well-engaged in the process however there is no evidence of research being considered in the consultation process (APRA 2019). In summary, we argue that clear, explicit and community-agreed and shared definitions of data quality will help to improve the uptake of DQ frameworks and regulations and assist with developing compliance centric DQ management approaches to reduce the burden of compliance relating to data quality aspects.

5 CONCLUSION AND OUTLOOK

In this paper we presented the analysis of selected DQ regulations and frameworks that are relevant to the banking and finance sector against a comprehensive reference set of DQ dimensions. We found that while a few basic data quality dimensions are common across several regulations and frameworks, variances exist that increase the chances of a significant disparity in compliance behavior, or even non-compliance. Given the rise of DQ regulations, organisations need to augment their data quality management efforts towards meeting compliance obligations. The results of our study provide an essential first step towards the design of compliance centric DQ management approaches. Our study provides a summary mapping, and an indication of the degree of similarity, to help organisations navigate their compliance to regulations and frameworks against their own data quality efforts. The results of our study can also assist organisations to identify DQ regulations, which, when adhered to first, can help in complying with other regulations as well, thus reducing compliance effort and improving the efficiency of compliance processes.

We acknowledge that while we made every effort to comprehensively cover depth and breadth of the topic, our study is not without limitations. While our findings provide the essential foundation to augment the compliance-centric approaches to DQ management, they may have limited generalisability for industries other than banking and finance. We also acknowledge that we are not legal experts, while the study examines regulations and laws, hence some differences of interpretation may exist.

Further research is needed to investigate and design DQ management approaches that are compliance centric. An appropriate approach could be action research (Stringer 2013), which emphasises practical and solution-oriented inquiry to solve real-life problems by engaging relevant stakeholders, gathering most relevant data, taking planned actions and managing sustainable change. We further argue that there is a need for regulators to issue additional guidance, or improve the regulation documentation, to promote clarity and shared understanding. Such guidance or updates will help better define data quality requirements so that a uniform intent and purpose of the regulations is conveyed to regulated entities, hence making the process of compliance easier while also improving the overall data quality of the organisations.

6 REFERENCES

Abdullah, N. S., S. Sadiq and M. Indulska. 2012. "A compliance management ontology: Developing shared understanding through models". International Conference on Advanced Information Systems Engineering, Springer.

Australian Broadcast Corporation (ABC). 2014. “ANZ to refund $70m” https://www.abc.net.au/news/2014-01-29/anz-refund-70-million-overchare-interest-repayments-mortgage/5225928 Retrieved on 24 May. 2018

Australian Bureau of Statistics (ABS). 2009. “Data Quality Framework”. Canberra.

Australian Prudential Regulation Authority (APRA). 2013. “Prudential Practice Guide, CPG 235 Managing Data Risk. APRA. Sydney.

Australian Prudential Regulation Authority (APRA). 2019. “Consultation on Prudential Practice Guide CPG 234 Information Security” https://www.apra.gov.au/information-security-requirements-for-all-apra-regulated-entities

Australian Securities and Investments Commission (ASIC). 2013. “BoQ to refund customers after system error” https://asic.gov.au/about-asic/news-centre/find-a-media-release/2013-releases/13-070mr-boq-to-refund-customers-after-system-error/ Retrieved on 23 Mar. 2018

Bank for International Settlements (BIS). 2013. “Principles for effective risk data aggregation and risk reporting” Basel.

Chen H, Chiang RH and Storey VC. 2012. “Business intelligence and analytics: From big data to big impact,” MIS Quarterly (1:36) December (4).

Currie, Wendy L., Daniel P. Gozman, and Jonathan JM Seddon (2018). "Dialectic tensions in the financial markets: a longitudinal study of pre-and post-crisis regulatory technology." Journal of Information Technology 33.4: 304-325.

European Union (EU). 2009. “Directive 2009/138/EC of The European Parliament And Of The Council (Solvency II)”. Brussels.

European Statistics (ES). 2007. "Handbook on data quality assessment methods and tools." European Commission, Brussels.

Garvin, D. 1987. "Competing on the eight dimensions of quality." Harvard Business Review: 101-109.

Jayawardene, V., S. Sadiq and M. Indulska. 2015. “An analysis of data quality dimensions”.

Juran, J. and A. B. Godfrey 1999. "Quality handbook." Republished McGraw-Hill: 173-178.

Kahn, B.K., Strong, D.M. and Wang, R.Y., 2002. Information quality benchmarks: product and service performance. Communications of the ACM, 45(4), pp.184-192.

Ledford, C. J., D. A. Seehusen, M. M. Villagran, L. A. Cafferty and M. A. Childress 2013. "Resident scholarship expectations and experiences: sources of uncertainty as barriers to success." Journal of graduate medical education 5(4): 564-569.

Loshin, D. 2001. “Enterprise knowledge management: The data quality approach”, Morgan Kaufmann.

New York Times (NYT). 2018. “Congress passes Dodd Frank rollback for smaller banks”. https://www.nytimes.com/2018/05/22/business/congress-passes-dodd-frank-rollback-for-smaller-banks.html Retrieved on 21 Oct.2018.

Price, R. and G. Shanks. 2004. A semiotic information quality framework. Proceedings of the International Conference on Decision Support Systems DSS04.

Redman, T. C. 1998. "The impact of poor data quality on the typical enterprise." Communications of the ACM 41(2): 79-82.

Sadiq, S. 2013. “Handbook of data quality: Research and practice”, Springer.

Seven News (7News). 2019. “Westpac bank to return home loan customers” https://7news.com.au/business/banking/westpac-bank-to-refund-home-loan-customers-after-processing-error-c-205524 Retrieved on 15 Jul. 2019

Shewhart, W. A. and W. E. Deming. 1986. Statistical method from the viewpoint of quality control, Courier Corporation.

Statistics Canada (SC). 2009. “Quality Guidelines”. Ottawa.

Stringer, E. T. (2013). “Action research”. Sage publications.

United States Congress (USC). 2010. The Dodd–Frank Wall Street Reform and Consumer Protection Act 2010. Washington, D.C.

Van Hulle, K., 2011. "Solvency II: state of play and perspectives". Zeitschrift für die gesamte Versicherungswissenschaft, 100(2), pp.177-192.

Wang, R. Y. and D. M. Strong. 1996. "Beyond accuracy: What data quality means to data consumers." Journal of management information systems 12(4): 5-33.

Wood, J. M. 2007. "Understanding and Computing Cohen's Kappa: A Tutorial." https://www.scribd.com/document/82858112/Understanding-and-Computing-Cohen-s-Kappa-A-Tutorial Retrieved on 20. Aug. 2018