Thesis Consent Form

This thesis may be consulted for the purposes of research or private study provided that due acknowledgement is made where appropriate and that permission is obtained before any material from the thesis is published. Students who do not wish their work to be available for reasons such as pending patents, copyright agreements, or future publication should seek advice from the Graduate Centre as to restricted use or embargo.

Author of thesis	Nicolette Helena Voskuilen
Title of thesis	Data Quality for Regulatory Compliance
Name of degree	Master of Engineering in Environmental Engineering
Date Submitted	27/04/2022

Print Format (Tick the boxes that apply)	
✔	I agree that the University of Auckland Library may make a copy of this thesis available for the collection of another library on request from that library.
✔	I agree to this thesis being copied for supply to any person in accordance with the provisions of Section 56 of the Copyright Act 1994.

Digital Format - PhD theses

I certify that a digital copy of my thesis deposited with the University is the same as the final print version of my thesis. Except in the circumstances set out below, no emendation of content has occurred and I recognise that minor variations in formatting may occur as a result of the conversion to digital format.

Access to my thesis may be limited for a period of time specified by me at the time of deposit. I understand that if my thesis is available online for public access it can be used for criticism, review, news reporting, research and private study.

Digital Format - Masters theses

I certify that a digital copy of my thesis deposited with the University is the same as the final print version of my thesis. Except in the circumstances set out below, no emendation of content has occurred and I recognise that minor variations in formatting may occur as a result of the conversion to digital format.

Access will normally only be available to authenticated members of the University of Auckland, but I may choose to allow public access under special circumstances. I understand that if my thesis is available online for public access it can be used for criticism, review, news reporting, research and private study.

Copyright (Digital Format Theses) (Tick **ONE box only**)	
✔	I confirm that my thesis does not contain material for which the copyright belongs to a third party, ***(or)*** that the amounts copied fall within the limits permitted under the Copyright Act 1994.
☐	I confirm that for all third party copyright material in my thesis, I have obtained written permission to use the material and attach copies of each permission, ***(or)*** I have removed the material from the digital copy of the thesis, fully referenced the deleted materials and, where possible, provided links to electronic sources of the material.

Signature

Date 27/04/2022

Comments on access conditions

Faculty Student Centre / Graduate Centre only: Digital copy deposited ☐ Signature Date

Data Quality for Regulatory Compliance

Nicolette Helena Voskuilen

A thesis submitted in fulfilment of the requirements for the degree of

Master of Engineering in Environmental Engineering,

the University of Auckland, 2022.

This thesis is for examination purposes only and is confidential to the examination process.

Abstract

This thesis explored the application of data quality assessment methods on regulatory standards concerned with environmental monitoring, and how these methods can contribute to assessing the extent to which a regulation requires compliance datasets to be interoperable and reusable.

As the volume of environmental monitoring data being collected is growing, increasing importance is placed on the quality of that data. While drinking water, wastewater and environmental waters are often regulated separately, they are intrinsically linked, and compliance data from one area may provide insight to another. By defining the data quality of these datasets, their interoperability is increased, allowing utility to extend beyond compliance assessments. The inclusion of greenhouse gas emission monitoring shows that this interoperability can exist between more than one industry type, and that is not limited to regulators, but also conveys benefits to other data users such as suppliers and researchers.

The method used was a qualitative assessment of regulatory standards, presented as a quantitative summary for analysis and trend identification. Data quality indicators applicable to environmental monitoring standards were identified, defined, and grouped. These indicators were assembled in a framework with a simple scoring system. Regulatory standards from the drinking water, wastewater, environmental water, and greenhouse gas emission industries were selected from New Zealand, Australia, and the European Union. These standards were assessed with this framework, allowing a summary to be compiled showing which data quality indicators were present in each standard.

A comparison of the distribution of data quality indicators across the assessed standards showed that the framework was effective in predicting the level of data quality required of compliance datasets for these standards. It allowed the assessor to understand how datasets from different standards could be interoperable, and how they could not. It also showed that the framework would be a useful tool in the creation of a new regulatory standard, as it provided a way of benchmarking existing standards at a national and international level.

The changing regulatory environment of the Three Waters industry in New Zealand provided a case study to show how an existing standard can be compared with a proposed draft, and what data quality indicators would be required in subsequent standards in order to allow the interoperability and reusability of compliance datasets.

Keywords: data quality, framework assessment, three waters, regulatory compliance, drinking water standards, interoperability, reusability.

Acknowledgements

This thesis was written with the guidance and support of my industry supervisor, Mark Homenuke, and my academic supervisor, Professor Naresh Singhal. Without your patience and kindness this would not have been possible. Thank you for giving me time and space to learn, and for listening to my many comments about the drinking water standards.

I would also like to thank Pete Johnson and Simon Porter. Both of you thought I could do something more, and you gave me the support and encouragement to believe in myself.

Special thanks to my family and friends for being there for me during this process. Your help – in the form of walks, coffees, cat pictures, and long discussions during various lockdowns – made all the difference.

Finally, thank you to my partner, Simon. I could not have done this without you. Your calmness kept me grounded. Thank you for your love, patience, and endless support.

Table of Contents

List of Tables

List of Figures

Glossary

ADWG	Australian Drinking Water Guidelines
AU Carbon	Carbon Credits (Carbon Farming Initiative – Domestic, Commercial and Industrial Wastewater) Methodology Determination 2015
AU NSW DW	Drinking Water Monitoring Program December 2005 (updated October 2011)
AU QLD EnvW	Monitoring and Sampling Manual 2018: Environmental Protection (Water) Policy 2009
BOD_5	Biological Oxygen Demand
C.t	The concentration of a disinfectant in mg/L multiplied by exposure or contact time in minutes
CCA	Climate Change Response Act 2002 (New Zealand)
cfu/mL	Colony forming units per 1 millilitre of sample
CMP	Compliance Monitoring Period
DAL	Division of Analytical Laboratories
DQ	Data Quality
DQD	Data Quality Dimension
DWA	Drinking Water Assessor
DWQAR	Drinking Water Quality Assurance Rules (December 2021 draft, New Zealand)
DWSNZ	Drinking-water Standards for New Zealand 2005 (Revised 2018)
DZ	Distribution Zone
E. coli	*Escherichia coli*
ECan	Environment Canterbury
EU	European Union
EU Carbon	Regulation (EU) 2018/1999 of the European Parliament and of the Council of 11 December 2018 on the Governance of the Energy Union and Climate Action, amending Regulations (EC) No 663/2009 and (EC) No 715/2009 of the European Parliament and of the Council, Directives 94/22/EC, 98/70/EC, 2009/31/EC, 2009/73/EC, 2010/31/EU, 2012/27/EU and 2013/30/EU of the European Parliament and of the Council, Council Directives 2009/119/EC and (EU) 2015/652 and repealing Regulation (EU) No 525/2013 of the European Parliament and of the Council
EU DW	Directive (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020 on the quality of water intended for human consumption
EU GW	Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy; and Directive 2006/118/EC of the European Parliament and of

	the Council of 12 December 2006 on the protection of groundwater against pollution and deterioration
EU WW	Council Directive of 21 May 1991 concerning urban wastewater treatment (91/271/EEC), amended[2] in 1998, 2003, 2008, and 2013
FAC	Free Available Chlorine
FAIR	Findability, Accessibility, Interoperability, Reusability
GHG	Greenhouse Gas
GPS	Global Positioning System
IANZ	International Accreditation New Zealand
IPCC	Intergovernmental Panel on Climate Change
KPI	Key Performance Indicator
LOR	Limits of Reporting
M2M	Machine to Machine
MAV	Maximum Allowable Value
mg/L	Milligrams per litre
MoH	Ministry of Health in New Zealand
MPN/100mL	Most probable number per 100 millilitres of sample
N/a	Not applicable
NATA	The National Association of Testing Laboratories
NGER	National Greenhouse and Energy Reporting (Measurement) Determination 2008 (Australia)
NSW	New South Wales, Australia
NTU	Nephelometric turbidity units
NZ Carbon	Climate Change (Stationary Energy and Industrial Processes) Regulations 2009
NZ DW TA	Drinking Water Quality Assurance Rules 20 December 2021 (Draft)
NZ EnvW	Resource Management (Measurement and Reporting of Water Takes) Regulations 2010
NZ WW Ash	ECAN Consent CRC030999.1
NZ WW Chch	ECAN Consent CRC051724
NZMS 260	New Zealand Map Series 260
OGD	Open Government Data
QA/QC	Quality Assurance and Quality Control
RC	Resource Consent
RMA	Resource Management Act 1991 (New Zealand)
SI	International System of Units

Three Waters	Drinking water, wastewater, and stormwater
USEPA	United States Environmental Protection Agency
WTP	Water Treatment Plant
WWTP	Wastewater Treatment Plant

Data Quality for Regulatory Compliance

1 Introduction

1.1 Overview

Organisations that produce environmental monitoring data often have many data sources and a variety of internal and external reporting methods which can result in a high volume of datasets of varying levels of quality. This volume of data may obscure high-quality data, preventing it from reaching its intended audience. The implications of this can range from adverse effects on operational efficiency, to incorrect reporting of internal key performance indicators, and to inhibiting the generation of datasets required to demonstrate regulatory compliance. Consequently, a challenge faced by many organisations is how to ensure the data being recorded is of a high level of quality, and how to enable its utilisation by different units within the organisation that can benefit from it.

Here the concepts of data governance and data quality become pertinent. Data governance requires the actions of data users, data collectors, and organisational management to ensure that data entering the organisation meets its required data quality standards [1], [2]. If an organisation was unsure how to define its data quality standards and wanted to find guidance on how to do so, one method would be to establish the data quality requirements of external report recipients. For example, if data is required to demonstrate compliance with a set of regulations, then in addition to determining what data should be reported on (such as parameters and their allowable limits), the regulations may also dictate the data quality requirements – as the quality of the data submitted for assessment may impact the organisation's degree of compliance.

In order to improve the data quality of compliance data being collected by a supplier, the data quality requirements of the relevant regulatory standards must first be evaluated. For example, an organisation operating in the Three Waters industry would need to understand the following:

1. What are the data quality requirements of the water standards that they need to achieve compliance with?
2. What is the level of data quality of the compliance data being collected?
3. Based on this comparison, what internal processes require what improvements in order to meet the data quality requirements of the applicable standards?

While water suppliers (organisations that operate water and/or wastewater treatment plants) would have a use for this method of comparison, it would also be beneficial to other organisation types; in particular, the regulatory body that compiled the standards. Studies have shown that an inability to share data between governmental agencies limits interagency learning and coordinated project work [3], and that where a standardised approach to drinking water data collection is lacking, the ability of the government to compare data from different sources is limited, preventing the monitoring of national trends [4]. Taking the regulatory environment of New Zealand, this method would allow a comparison of the current environmental monitoring standards at a national level, in order to see which data quality requirements are present and which are not. This would facilitate the comparison of data from governmental agencies, and would assist national trend monitoring of environmental data. If this comparison was extended to international standards, it would allow New Zealand standards to be benchmarked with those from other countries, and would provide examples of how data quality requirements are communicated in other regulatory environments.

Other entities that would benefit from such a method of comparison would be service providers working alongside the industry, and researchers. Service providers would be able to apply the framework to the relevant standard(s) in order to understand the regulatory requirements of the supplier; researchers interested in including the compliance datasets in other studies would be able to understand the data quality of the dataset, increasing its interoperability and allowing it to be analysed alongside other datasets.

Therefore there is a tangible benefit to the development of a way to compare the data quality requirements of various environmental standards – not just to suppliers as the producers of the data, but also to service providers working with the suppliers, researchers wishing to reuse the compliance datasets, and the regulatory body that maintains the standards.

In order to define the data quality requirements of a regulatory standard, indicators of data quality must be identified. Data quality indicators are aspects of a standard that, when present, increase the data quality of a dataset. This is achieved by increasing the level of confidence that the data received by the regulator is accurate, reliable, and auditable. By summarising the data quality indicators that are present in a standard, the overall data quality requirements of that standard can be described.

Data quality (DQ) is an emerging area of research; studies have focused on how to define DQ [5], [6], and also how to define aspects and indicators of DQ, along with the development of DQ assessment methods [7]–[9]. Common themes of DQ research include the tenet that data quality is contextual [5], [6], [10], and that any proposed standards or definitions require a consensus [6]. Reflecting this, DQ assessment methods can be developed iteratively [11], so that feedback on the performance of the method is included in its ongoing development and refinement.

One type of DQ assessment method is the utilisation of a framework [12]–[14]. These frameworks are typically comprised of a set of DQ indicators – an aspect of a data point that conveys information about its level of quality. These indicators are context-specific: using an example from the drinking-water industry, a DQ indicator of an *E. coli*[1] sample result would be the laboratory analysis method that was used. DQ indicators are then aligned with DQ dimensions (DQDs), which are more general concepts of quality; one DQD may have a series of DQ indicators associated with it. For the *E. coli* example, this would be aligned with a DQD such as "believability", where knowing what test method was used conveys more believability in the quality of that result, and so the quality of the data point is greater than an *E. coli* sample result where the test method is not known. DQDs in turn may be further grouped into hierarchical classifications [15]. This can be a useful way of translating data quality requirements into a business process assessment such as Key Performance Indicators (KPIs[2]). In the context of the *E. coli* example, the DQD of 'believability' may be part of a hierarchy called "Trust". By assessing the performance of the DQDs under that hierarchy using a simple yes/no assessment for each DQ indicator, a score can be assigned to the hierarchy (or KPI). Scoring systems for KPIs are often used as they are a straightforward way to identify where improvements in an organisation's operational processes are required, and also to evaluate when such improvements have been made. As a result, future *E. coli* sample results may be reported with the test method used, giving the recipient of the data greater confidence in the result, and also allowing the result to be compared with other datasets that used the same test method.

An example of the application of a framework is as follows:

> A dataset (such as *E. coli* sample results) is assessed by determining if the established DQ indicators are present in the dataset. This allows the framework to illustrate the performance of the dataset against the DQ dimensions and their higher-level classifications, or KPIs. By applying this framework to more than one dataset, a comparison of the data quality of each dataset can be made, allowing the strengths and weaknesses of each to be identified. In practice, if a regulator was assessing the data quality of datasets submitted for compliance, they would be able to establish which datasets included the test methods used to obtain the results, and which did not.

Complementary to the emerging research on data quality are the FAIR guiding principles for good data management and stewardship [16]. The FAIR acronym stands for Findability, Accessibility, Interoperability, and Reusability; should a dataset meet these qualities, it is said to be FAIR. Proposed in 2016, the Principles were designed as aspirations rather than prescriptive standards [17]. As a result they have grown in popularity and have precipitated a discussion on what makes data achieve the FAIR principles [17]–[19]. The FAIR principles are similar in nature to the hierarchical classifications of data quality. As such, assessment frameworks to determine the "FAIRness" of a dataset have also been proposed [20]; as with the DQD-Indicator method, FAIR assessment methods also focus on evaluating the data quality of a dataset by assessing the data points themselves [20]. Depending on the size of

---

[1] *Eschericha coli*, a species of bacteria that is used as an indicator of water quality in drinking water treatment and distribution processes.

[2] Key Performance Indicators (KPIs) are a commonly-used business metric. They have a wide variety of applications such as measuring changes over time and tracking the progress of projects.

the dataset, this can be a very manual task [9], and while it is possible for some data quality indicators to be designed for assessment by autonomous systems [21], the large amount of processing time required to render a DQ indicator as computable means that only a handful of indicators have been considered in this manner [22].

Of interest to this thesis is the application of data quality assessments to environmental monitoring standards relevant to the Three Waters industry. In addition to defining the data quality indicators of compliance datasets, the benefit of being able to compare the number of DQ indicators present in these compliance standards means that the datasets can be more easily compared across industries, increasing their interoperability.

DQ research and the FAIR principles provide a basis for understanding what data quality indicators of compliance datasets look like, however the manual assessment method of those datasets is not transferable. This is because rather than one research dataset being assessed, a regulator would have to assess every compliance dataset submitted to them – a particularly time-consuming task, and one that does not promote interoperability; for example, a drinking water regulator in New Zealand could only assess compliance datasets sent to them, and not those submitted to any Australian drinking water regulators. Therefore, the application of DQ assessment methods to compliance standards may be approached in a different manner. Instead of assessing the datasets themselves, the assessment may be carried out on the DQ indicators of these standards. Because the context of the compliance report datasets remains constant, in that its purpose is to achieve compliance with the standard, the indicators of data quality do not change between datasets; rather, they remain constant.

In order to comply with an environmental monitoring standard, the dataset must meet the requirements of that standard. While this refers to the sample results being within the acceptable limits, it also includes the aspects of the standard that ensure the results are reliable – i.e., that the quality of the data is acceptable. These aspects are the data quality indicators, and by understanding which data quality indicators are present in the standard, it allows the data quality of any dataset submitted for a compliance assessment with that standard to be estimated as well. This means that by assessing the data quality indicators of one standard, the expected data quality of multiple datasets can be determined.

If the data quality of a dataset is understood, it is more likely to be reused [23]. As well, assessment of datasets with known data quality increases confidence that the results being analysed are directly comparable; or even if they are not, the extent of the uncertainty is defined.

Therefore the benefit of assessing the data quality indicators of standards is that it is likely to facilitate data sharing and dataset comparisons, increasing the accessibility of compliance datasets, resulting in enhanced interoperability and reusability – two of the FAIR principles.

In addition, by determining which data quality indicators are present or absent, the overall data quality requirements of the standard can be understood. This allows the data quality requirements of the standards themselves to be compared, benefiting researchers, suppliers, and regulators.

1.2 Research Structure

This thesis will review data quality assessment methods (including frameworks, definitions, indicators, attributes, and the FAIR principles) to determine data quality indicators that are applicable to environmental monitoring standards that require compliance reporting.

Once identified, the data quality indicators will be summarised in a framework which will be applied to various environmental monitoring standards in the water and carbon industries. The framework will be used to determine the data quality of compliance datasets submitted in pursuit of compliance with those standards. The summary of data quality indicators present in each standard describes its data quality requirements; the framework can therefore also be used as a method of comparing the role of data quality in different standards.

This framework is expected to be beneficial to:

- Data creators (organisations that produce the compliance datasets such as water suppliers)
- Data curators (those that use environmental monitoring data to compile compliance reports)
- Data users (regulators and auditors)
- Data researchers (present in all of these groups but may also be external to the compliance assessment process, for example university researchers).

The framework will enable organisations to compare their internal data quality indicators with those required by regulatory standards, which may assist in achieving compliance. For water industry organisations that wish to calculate GHG emissions, it will allow them to critically assess what data quality requirements they currently meet, and what changes would be required to address those they do not, resulting in compliance monitoring processes that generate data that can be used for more than one compliance reporting process.

By understanding the data quality requirements of regulatory standards, an assessment can be made on the extent to which they enable compliance data to achieve the FAIR principles; similarly, it may allow a regulator in the process of writing new standards to identify what additional data quality indicators could be included to champion the FAIR principles. This could result in increased utility of the compliance datasets – instead of being used solely for static compliance evaluation, they could be used for national trend monitoring, international benchmarking, or in research.

The application of the framework will determine its efficacy as a tool to compare the data quality requirements of regulations across industry types, as well as between countries. This will be explored in a case study focused on the drinking water standards in New Zealand, as Taumata Arowai, the new drinking water regulator, is currently[3] writing new standards to replace the existing standards. The most recently published draft version of the proposed new standards will be assessed against the existing standards, and the DQ requirements of each will be compared.

1.3 Research Objectives

This thesis aims to identify and define the data quality indicators of regulatory standards that apply to drinking water, wastewater, environmental water, and greenhouse gas emission calculations. These indicators will be compiled in a framework that allows the data quality requirements of a set of regulatory standards to be critically compared with another. The legislation included in this review is from New Zealand, Australia, and the European Union.

The application of the framework will:

- Allow organisations to better understand their current data quality status
- Identify what the data quality requirements are for an organisation's existing compliance reporting processes, and any additional ones they wish to implement
- Determine what system and/or process improvements are required for an organisation to meet new or existing data quality requirements
- Allow regulators to compare the data quality indicators present in an existing standard with its proposed superseding standard
- Identify common data quality indicators of standards from different industries, allowing better interoperability of compliance datasets
- Create a benchmark for the data quality indicators of one standard to be compared with others – either within or across industry types, and at a national or international level.

The main objectives for this research are as follows:

- Identify regulatory standards relating to the water industry and the calculation of greenhouse gas emissions that require environmental monitoring to be completed in order to achieve compliance. Countries included are New Zealand, Australia, and the European Union
- Identify and define data quality indicators of regulatory standards that require compliance monitoring and reporting
- Build a framework that allows the data quality indicators of the identified regulatory standards to be assessed, in a manner that:
    - Uses the presence or absence of the data quality indicators to summarise the data quality requirements
    - Determines the expected data quality of a compliance report

[3] At the time of writing this document.

- Allows a regulator to compare the data quality requirements of different standards (both across industries and between countries)
- Assists an organisation operating in one or more of these industries to understand what data quality requirements it is required to meet.

1.4 Research Question

The primary research question is as follows:

Can data quality evaluation techniques and assessment methodologies be used to develop a framework to allow the critical assessment of the data quality indicators and requirements of environmental monitoring standards, leading to the determination of the quality of datasets submitted in pursuit of compliance with these standards?

The secondary questions are:

- What are the data quality indicators of regulatory compliance standards?
- Do data quality indicators of regulatory compliance standards allow the assessment of the data quality requirements of these standards?
- Can data quality indicators be used as a point of comparison for regulatory standards?

2 Literature Review

The literature review includes the research area of data quality, including the FAIR principles and data quality assessment methods. As environmental monitoring regulations are of interest due to a changing regulatory landscape and a large amount of utility for various stakeholders in the Three Waters industry, regulations for drinking water, wastewater, environmental waters, and greenhouse gas emissions from Australia, New Zealand, and the European Union are identified as being of interest to assessment by a data quality framework.

2.1 Principles of Data Quality

2.1.1 Data Quality Definitions

Data quality may be considered as an aspect of quality assurance and quality control (QA/QC) [24]. While QA/QC is a well-established discipline, the study of data quality is in an emergent phase. This is in part due to the rapidly changing nature of the information technology industry and the rise of 'big data' – as global data growth increases [15], there is more information available for analysis than ever before. As an analysis is only as good as the quality of the information it looks at, so data quality has now become a desirable attribute of information analysis [15].

In order to determine the data quality attributes of a dataset used in an analysis, a definition of data quality is required. Principal findings from studies and workshops [5], [6], [10], [15], [25] have found that:

- Data are a product, and as a product, have quality – similar to other products.
- Data quality needs to be measured for it to be able to be improved.
- Data quality is contextual – the application of the data influences the measure of its quality.

While standard QA/QC practices may be employed to evaluate product quality levels, this method is not as easily applicable to data quality. This is because the definition of data quality is often referred to as being "fitness for use by data consumers" [5], [26], [27], and therefore changes as the data consumer changes. This makes data distinct from other products as a measure of its quality is dependent on the context in which the data exists, leading to the question, what is a metric of data quality? One approach that has been used to measure data quality is to divide it into dimensions (Data Quality Dimensions or DQDs). A DQD is defined as a set of data quality indicators that represent a single aspect or construct of data quality [5], [15], [28]. For example, the DQD 'completeness' can be defined as "the extent to which data are of sufficient breadth, depth, and scope for the task at hand" [22]. As the DQDs have more tangible definitions than DQ as a whole, DQDs are easier to measure.

Data quality dimensions can be further broken down into smaller pieces termed data quality indicators and data quality elements [15]. These two terms – indicators and elements – are used interchangeably; in this document, the term 'indicator' will be used primarily. Data quality indicators, elements, and dimensions will be collectively termed attributes. DQDs can also be grouped at a more abstract level – termed hierarchical classifications – which typically reflect high-level attributes considered to represent data quality in the context of the dataset [15]. Figure 2.1 illustrates data quality attributes in a hierarchical formation.

By utilising hierarchical classifications, the data quality indicators can be evaluated in a manner that highlights which aspects of the hierarchy has strengths and weaknesses, and what they are. In the literature, hierarchies reflect desirable attributes such as intrinsic data quality, contextual data quality, representational data quality, and accessibility data quality [5]. While this approach was developed with a focus on research data, it has real-world application: an organisation seeking to better understand its data quality status could survey its data users to establish their data quality indicators, and a data quality assessor or project manager could fill the role of the researcher and classify these indicators as DQDs, assigning them to hierarchies that align with established Key Performance Indicators (KPIs). This allows an organisation's data quality status to be evaluated in a manner that aligns with existing reporting practices, and also allows improvements to be clearly communicated.

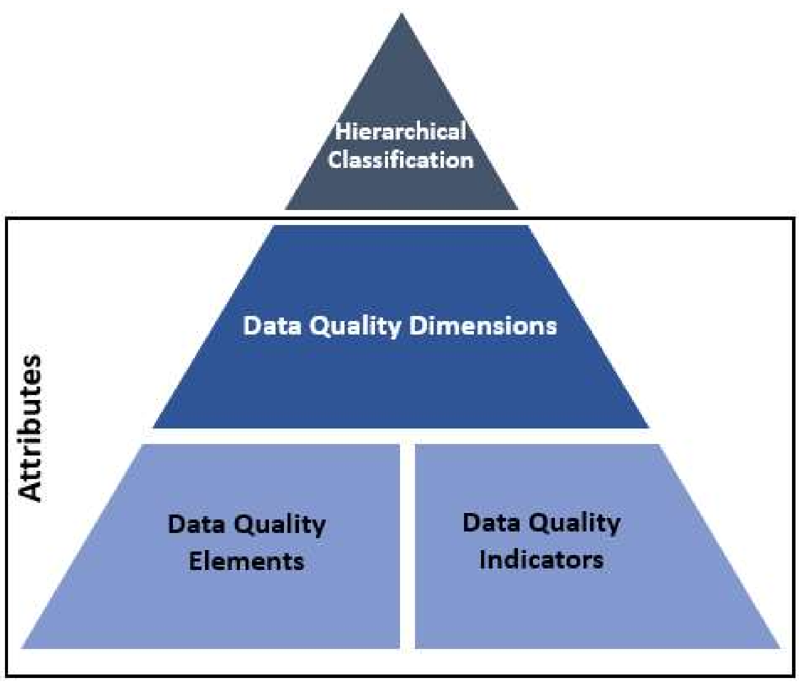


Figure 2.1 A schematic of data quality attributes arranged in a hierarchical formation

Therefore, by identifying data quality attributes relevant to the context and application of the dataset, they can be used to assess its data quality [12]. Three methods of identifying DQ attributes have been described: intuitive, theoretical, and empirical [5]:

- The intuitive method requires the researcher to determine which DQ attributes are important, given the context and use of the dataset.
- The theoretical approach focuses on how data may become deficient during the data manufacturing process.
- In empirical studies, data quality indicators are determined by data consumers or data users, as opposed to the researchers [5], [9], [15]. Once data users identify what they consider to be data quality indicators, the researcher is then able to group these into DQDs.

Of these three methods, the empirical approach aligns the most closely with the DQ definition of 'fitness for use by data consumers'. Previous studies have looked at how to apply the empirical method of DQ assessment to research data. One method in particular requires those in charge of data infrastructure (data curators) to evaluate the data [9]. The downsides of this method were noted as being only as reliable as the curators, and that it could be quite a manual task if every data point is required to be evaluated [9]. In addition, if the dataset had more than one end user or application, the DQ assessment may vary, due to the contextual nature of DQ. This method also requires the curator to understand both the context in which the data was collected, and the context in which it is being used or analysed. In a research setting, this could result in the researcher taking the role of curator, and therefore assessing the quality of their own data – i.e., the data creator assesses the data quality.

2.1.2 The FAIR Principles

In 2016, with the goal of clarifying good data management and stewardship practices, the FAIR Guiding Principles were proposed [16]. The Principles – Findability, Accessibility, Interoperability, and Reuse – were designed for application in the context of digital research objects [16], [17], but have since expanded in popularity and have been adopted by a range of industries, including the healthcare industry [29], [30].

Findable	Data are described, registered, or indexed in a manner that is clear and searchable, allowing the data to be found [16], [23], [31]
Accessible	Access to the data should be clearly defined, including any authorisation procedures [16], [23]
Interoperable	The data language used is shared and ideally structured using a common standard [16], [23], in order to facilitate integration with other data [31]

Reusable	Data are well-described so that they can be reused in new settings [31]. Data have an accessible data usage license to allow reuse with respect to ownership of the data [16]

Table 2.1 A summary of the FAIR Principles proposed by Wilkinson et al [16]

As the FAIR principles were designed to encourage the discovery and re-use of existing datasets with new datasets, thus allowing datasets to retain their value [32], they therefore facilitate the improvement of data quality [30], and encourage the secondary use of data [33]. As such, the Principles have been adopted as key data management techniques; however, as they do not take the form of a standard or specification [16], in practice, their application is varied.

The FAIR principles consider the role of metadata to be as important as the data in a dataset. Metadata here refers to information about the data, such as definitions, descriptions, methods used, dataset attributes, and any assumptions made [23]. By placing this emphasis on metadata, the FAIR principles support an increased focus on the role of data stewardship, or data curation.

In order to curate datasets with sufficient metadata to achieve the FAIR principles, data standards are required [32], [34]–[36]. While formal data standard development is complex, in the context of the FAIR principles, the term 'standards' more often refers to agreed-upon conventions and guidelines [36]. Data standards support the FAIR principles as follows:

- Findable: standard descriptors allow datasets to be discovered
- Accessible: standard data storage formats allow datasets to be accessed more readily
- Interoperable: datasets that share standard metadata can be compared either directly or by standard transformation methods
- Reusable: similar to the interoperable principle, if the metadata standards used by a dataset are known, it increases the likelihood that the dataset will be reused in future research, as the quality and context of the data is able to be understood. In addition, transparent data usage licenses address data ownership rights; if the usage license is clear, reuse of the dataset is more likely.

As a result, FAIR datasets can be found, accessed, compared, and reused autonomously by machines [23], [30], in addition to humans.

2.1.3 Assessment methods

While data quality assessment methods are often designed with research datasets in mind, the FAIR principles can be applied to a wide range of applications. Therefore, in designing the framework for this thesis, a combination of the assessment methods of each is explored.

A literature review of the role of data quality in the health sector found that most papers considered DQ to be a multidimensional construct [26]. While DQ indicators are typically specific to the dataset [9], by categorising these as DQ dimensions, datasets can be compared based on their level of data quality. The most commonly used DQDs were accuracy, completeness, consistency, correctness, and timeliness [26].

Using DQDs to define data quality is useful, but because DQ is contextual, the weighting of the DQDs may change with every application of the dataset. In addition, there is no established vocabulary or consensus framework for DQDs, which can lead to duplication of DQDs and confusion around their definitions [26]. This means that the way one dataset is assessed is unlikely to be repeatable, as the definitions of the DQDs may change between applications, over time, or both. However there is merit in their creation; the process of identifying data quality indicators means that data users are an important part of the DQ assessment procedure. In addition, categorising DQDs at a hierarchical level is a useful method for understanding DQ at an organisational level, as it lends itself to a high-level, overview-style analysis.

Applying this method to assess the data quality of datasets produced in order to achieve compliance with regulatory standards would follow these steps:

- Data quality indicators specific to the compliance dataset are identified and defined, and compiled as a framework
- These indicators are aligned with data quality dimensions identified as being pertinent to regulatory compliance
- The datasets are assessed against the framework to determine which DQDs are present
- Categorising the DQDs into hierarchies allows comparison of the datasets at a high level

This method is useful in that it allows data users to determine the indicators of data quality, and that the framework allows the same indicators to be assessed for more than one dataset. When combined with the FAIR principles, a methodology can be designed that describes which data quality indicators will increase the 'FAIRness' of the dataset.

Because the FAIR principles are aspirational [35], FAIR assessment methods can be more focused on how the data quality of a dataset can be improved [17], rather than just assessing its current level of DQ. This is of interest to this thesis as it would allow existing standards to be compared with new proposed draft standards in a manner that describes how to prioritise data quality, and what the outcome would be. For example, if neither the existing standard or the proposed new standard include a requirement for the GPS coordinates of a sample location to be part of the compliance report, the data quality for that indicator is the same for both datasets. However, if the indicator is aligned with one of the FAIR principles – in this case interoperability – then the assessment of the dataset also shows the benefit of including the indicator in the new standard.

An example of this is the FAIR Data Maturity Model [35] which uses indicators based on the original descriptions by Wilkinson *et al* [16] as the foundation of a data evaluation methodology. By utilising these indicators, data users have a standardised starting point from which to build a methodology using their own questions or metrics [35], that will illustrate how the FAIRness of their dataset can be improved [35].

Both method types – those that focus on the FAIR principles to show how a dataset can be improved, and those that evaluate the data quality dimensions that are present – utilise the concept of data quality indicators as the way to evaluate a dataset. And while the focus of many FAIRness assessments is on research datasets, a study by Eder & Jedinger investigated using the FAIR principles to evaluate election studies [37]. This study found that by using the FAIR principles as guidelines, data quality indicators specific to their analysis could be identified. By using a framework to score these indicators for each election study, a comparison of the FAIRness of the studies could be made [37]. This demonstrates that the FAIR principles have a broad application, and by ensuring that the data quality indicators are well-defined, a data quality assessment could be made not just on a dataset (for example, the results of an election), but on an object that assesses a series of datasets (for example, election results studies). Therefore, by identifying data quality indicators that can be used to assess a regulatory standard that requires a dataset to be submitted as a compliance report, the data quality of all such datasets can be assessed. In addition, this would allow the regulatory standard to be evaluated to determine how it helps or hinders the FAIRness of these datasets, and how changes to data quality indicators could improve this.

2.2 Regulations

2.2.1 Three Waters Regulatory Standards

For organisations in the three waters industry, and particularly those that supply drinking water, external compliance reporting can be an important requirement and may underpin many internal processes. If the regulatory standard outlines the data quality requirements of the compliance reporting process, this may influence the way a supplier collects and stores not just compliance data but operational data as well. However if a regulatory standard does not have strict reporting requirements around data quality, the regulator assessing compliance data may not have as much faith in the veracity of the data. In addition, a lack of standardised reporting requirements at a national level means it is more complex to compare datasets, meaning benchmarking exercises are more difficult [38].

In Australia, a central guideline document (the Australian Drinking Water Guidelines or ADWG) [39] provides supplementary information to a series of drinking water regulations that apply at the state/territory level. As there is no set of drinking water regulations that applies to Australia as a whole, compliance reporting for one set of regulatory standards may have different requirements to another. For example, in the Northern Territory, the Water Act refers suppliers to the ADWG in lieu of a set of standards, however no further compliance requirements are outlined. Comparatively, in New South Wales (NSW), the NSW Health Drinking Water Monitoring Program outlines the monitoring requirements for drinking water samples taken by NSW drinking water suppliers, and utilises the AWDG as a model of best practice [40].

The NSW Health Drinking Water Monitoring Program provides water suppliers with labels that state the test type required ([40], page 15). The results are sent directly from the testing laboratory to the regulator (NSW Health), removing the requirement of the supplier to compile data and submit it. This also allows the regulator to be certain that the results they receive are not subject to any transcription errors and can be relied upon when making compliance assessments.

The AWDG provides technical guidance on water treatment processes, but are not mandatory standards ([39], page 2). As a result, suppliers in NSW collect data in a different manner and with a different level of data quality compared with those in other states. This type of situation has been identified as one that can lead to knowledge gaps, and a potential disparity in the quality of the water being supplied [4]. In addition, a lack of a standardised national system means that both the access to compliance data and the ability to compare it can be difficult [4]. Conversely, if a national repository for drinking water compliance data exists, it allows regulators to assess supplier data in a broader setting [3]. In addition, if this repository is also accessible by the water suppliers themselves, it allows them to benchmark their own performance against other systems, potentially leading to better system performance.

By using a framework to assess the data quality requirements of different drinking water standards in Australia, the standards can be compared with each other, to understand which data quality indicators are common across the standards. However if the data quality indicators are similar, it makes potential improvements harder to achieve, as there is no existing example to base the change on. When compared with a regulatory standard from a similar industry with more data quality indicators present, it allows the assessor to understand which DQ indicators are required for a regulatory standard to require a higher level of data quality. In addition, if there is the potential to reuse the data between the industries, such a comparison identifies to what extent that can be achieved [38].

An example of a regulatory standard that contains a number of data quality indicators is the Environmental Protection (Water) Policy 2009 for Queensland [41]. This Policy outlines the data quality indicators for environmental monitoring data in detail, including the method of recording GPS coordinates of the sample location, the sample transport requirements and measures to ensure these were met, the chain of custody documentation used by the testing laboratory, and the method of results transfer from the laboratory to the supplier and the regulator. Because environmental monitoring – in particular, the water quality of drinking water sources such as rivers and groundwater – has tangible crossover with drinking water monitoring, understanding the level of data quality of these datasets increases their value to the other industry.

While in Australia the drinking water regulations vary between states, in New Zealand, the Drinking Water Standards for New Zealand 2005 (2018 revision) (DWSNZ) are the regulations that apply to the country as a whole. The DWSNZ detail the parameters that suppliers must test for, including the test method used, the acceptable values, and the compliance monitoring periods (CMPs) that apply [42]. While these standards apply to all drinking water suppliers in New Zealand, the national repository of this data is only accessible in totality by the regulator – suppliers are not able to benchmark against each other until a national report summarising the data is published [43]. The data presented in these reports is typically a high-level summary of compliance reporting and so has little utility for suppliers wishing to troubleshoot their operational reporting data in comparison with other suppliers around the country.

Similar to the NSW Health Drinking Water Monitoring Program, the DWSNZ include requirements for laboratory accreditation and test methods, but they do not specify the method of transmitting data from the supplier (or the testing laboratory) to the regulator. A supplementary document (Supplementary Criteria for Accreditation: MoH Register of Water Testing Laboratories, [44]) includes the reporting requirements a laboratory must meet if it is to become accredited (the DWSNZ specifies that only accredited laboratories may be used for compliance testing); however, data transfer methods are not detailed. Therefore when compliance data is sent from the supplier to the central repository[4], the regulator is able to see some aspects of data quality surrounding each test result (including, but not limited to, the test method used, the accredited laboratory used, the identification of the sampler, and other sample data such as the time and location of the sample), but with the understanding that this data was compiled by the supplier and so was not sent directly from the testing laboratory.

The design of a framework to identify and define the current DQ requirements of the DWSNZ is particularly pertinent, as the DWSNZ are expected to be superseded by a new standard. On 15 November 2021 the Water Services Act 2021 ([45] came into force, establishing the new Crown entity Taumata Arowai as the water services regulator and removing the role from the Ministry of Health (MoH). The DWSNZ written by the MoH are still in force while the new standard is in development by Taumata Arowai. On 20 December 2021, Taumata Arowai published a draft version of the new standard – the Drinking Water Quality Assurance Rules (DWQAR, Draft, [46]). This document, along with the draft water quality standards [47], is the proposed replacement to the DWSNZ (ref). This proposed draft will be assessed alongside the current DWSNZ.[5]

---

[4] Prior to November 2021, this central repository was Drinking-Water Online (DWO), an online portal for suppliers to upload their data so that Drinking Water Assessors for the Ministry of Health could assess it. Taumata Arowai became the regulator on 15 November 2021; from this date, suppliers were no longer required to submit data to DWO, but they were required to maintain records as required by the DWSNZ.

[5] As the draft DWQAR published on 20 December 2021 were assessed as part of this document, changes made by Taumata Arowai to these rules at any time after 20 December 2021 have not been included in the assessment. The author acknowledges that upon publication of this thesis, changes to the draft DWQAR are likely to have occurred; however, the assessment should be reviewed based on the 20 December 2021 version of this document.

Once the new drinking water standards are established, Taumata Arowai is expected to focus on updating wastewater regulations. Currently the framework for wastewater regulations is found in the Resource Management Act (1991) (the RMA), which requires local councils to manage wastewater treatment through the provision and management of Resource Consents (RCs) [48]. Because Taumata Arowai is a centralised regulator, it is expected the wastewater regulations will become part of the Water Services Bill, removing the requirement of local councils to act as regulators for wastewater treatment suppliers.

While neither New Zealand or Australia have a set of national regulatory standards concerning wastewater monitoring requirements in the manner of the DWSNZ, specific resource consents issued to organisations that process wastewater treatment in New Zealand often contain monitoring requirements. As such, by applying the assessment framework to these RCs, an understanding of existing data quality requirements of wastewater monitoring can be built. The value of this exercise is that it provides the assessor with the means to compare current wastewater monitoring requirements around New Zealand with the monitoring requirements of the drinking water standards. Therefore, should Taumata Arowai commence the design of a national set of wastewater regulations, the specific data quality indicators to ensure interoperability with existing drinking water datasets will be visible.

The benefit of assessing New Zealand and Australian regulatory standards concerned with Three Waters monitoring using a data quality framework is that it will allow a comparison of the data quality requirements of each standard to be made, along with how they enable compliance datasets to achieve FAIRness. This will assist any organisation that is working with more than one set of standards by facilitating comparison of existing and potential data quality indicators. It will also allow benchmarking between the regulatory standards from Australia and New Zealand. Using examples like the Environmental Protection (Water) Policy 2009 for Queensland, which has a high number of data quality indicators present, provides examples of what DQ indicators might look like if they are absent in other standards. This can provide utility for water suppliers that may need to produce compliance reports for more than one treatment process. It can also act as a guide to assist in the creation of new standards, in order to ensure they adequately capture data quality requirements to increase the interoperability and reusability of compliance datasets.

While the Australian and New Zealand regulatory standards are written at a state or national level, directives concerning environmental monitoring requirements issued by the European Union (EU) are written to guide the Member States in writing country-specific regulatory standards. For example, the EU Drinking Water Directive 2020 outlines the requirements that EU Member States must take to ensure the quality of drinking water supplied to consumers [49]. This includes what parameters a Member State requires a supplier to provide data on and requires a Member State to set acceptable values for these parameters, which may not be less stringent than what is included in the EU Directive (Annex 1, [49]). The Directive also requires Member States to publish information on the supply of drinking water for the use of consumers (Article 17, [49]). It also specifies what datasets are required to be set up for access by EU entities, including the European Commission, and notes the presentation method required (Article 18, [49]). By specifying this, it allows the creation of a standardised system whereby compliance reporting datasets can be easily found and comparisons between Member States can be made. Similar Directives have been written for groundwater protection (Directive 2006/118/EC, [50]) and wastewater treatment (Directive 91/271/EEC, [51]).

By assessing these EU Directives with the data quality framework, it allows the Member States to identify what DQ indicators they are required to include in their national-level regulatory standards. As well, it facilitates the comparison of the data quality requirements of existing Directives and standards alongside those that are in the process of being written to determine that all required DQ indicators are being captured.

As the Member States are assessed against the Directives, interoperability and reusability of the compliance datasets of the Member States is an important requirement. By including the data quality indicators that facilitate interoperability and reusability in the Directives, subsequent compliance datasets from all Member States can be expected to reflect these requirements. Therefore viewing the data quality indicators present in the EU directives alongside regulatory standards from New Zealand and Australia can provide insight into which DQ indicators increase interoperability and reusability.

2.2.2 Greenhouse Gas Emission Reporting Regulatory Standards

While there is considerable overlap in the compliance monitoring requirements for the Three Waters industry, due to the increasing importance of including carbon accounting in industrial processes, greenhouse gas (GHG) emissions monitoring is an emerging factor in Three Waters monitoring processes [52]. Wastewater treatment plants (WWTPs) are naturally a focus of emissions reporting because methane (a GHG) is emitted as a waste product [52]. Indications from Water NZ[6] are that emissions reporting is expected to become a central part of Three Waters best practice [52]. In New Zealand, while the Climate Change Response Act (2002) sets out operational requirements for a range of different industry types, regulations covering the methodology for monitoring GHG emissions from WWTPs have not yet been published. Should this change, it would benefit organisations operating WWTPs to understand the data quality requirements of the monitoring data they currently capture, and how it compares with additional GHG monitoring requirements.

Regulations covering the methodology and monitoring requirements of WWTP GHG emissions have been published at a national level in Australia. The Carbon Credits (Carbon Farming Initiative – Domestic, Commercial and Industrial Wastewater) Methodology Determination 2015 [53] can be assessed by the data quality framework and compared with existing GHG emission methodologies for New Zealand, in order to understand the potential DQ indicators for future WWTP emissions monitoring requirements. When combined with the data quality assessment of resource consents, suppliers may

[6] An independent organisation that acts as the industry body for the Three Waters sector in New Zealand.

gain an understanding of what monitoring practices and data quality indicators are likely to be included in future regulatory standards. This would be further supported by a framework assessment of the EU Directive concerning GHG emissions monitoring (Regulation (EU) 2018/1999, [54]) alongside those Directives focused on Three Waters monitoring.

2.2.3 Data Quality Assessment Methods for Regulatory Standards

Data quality assessment methods and FAIR assessment methods are typically applied to datasets, and are quantitative in nature [38], [55]–[58]. It is unusual to find a qualitative data quality assessment method, as this would typically not be applied to a dataset. The data quality framework developed by Micic *et al* [13] was designed to be applied to heterogeneous data, and so could consider the level of data quality for varied data types, such as continuous monitoring data and qualitative or descriptive data, although no formal assessment was implemented.

One study applied a data quality management approach as a method of assessing data collected by an organisation, showing that for an organisation to understand data quality, a holistic view of as data as an asset is required [59]. The method identified qualitative components of data quality likely to be present in organisations and used these to assess case studies. The outcome of this study allowed a comparison of how data quality was represented in the organisations studied [59].

Riebel (1999, [60]) acknowledged the role of compliance requirements in driving operational processes, and presented a system to improve the quality and reliability of environmental data used for compliance monitoring and process control. This paper created a checklist for internal use to increase organisational DQ in processes involved in compliance reporting. Indicators of DQ were also qualitative, such as documentation relating to sampling and testing methods, but did not refer to specific compliance legislative requirements [60].

While these studies were concerned with the role of data quality in organisations, Miasayedava *et al* [61] used environmental monitoring data collected as part of the EU's water framework directive to build an automated environmental flows system – an example of reusing open government data (OGD) for a research project. By using OGD, issues with data quality from other data sources were reduced, as environmental monitoring data can be unstructured and of varying data quality [38]. While OGD is aspirational, and embodies the FAIR principles, in order to allow the benefits of access to these data, the metadata should be of high quality and should be maintained [62]. This is supported by the study made by Khan *et al* [63], where New Zealand national-level datasets were used to investigate environmental health indicator linkages with drinking water data. While there is a benefit to using datasets beyond their original intent, one of the issues raised by the study was that it relied on "the definitions applied by the organisations collecting the data" [63]. Therefore there is a need to define the data quality indicators of regulatory standards that generate compliance datasets; as more OGD initiatives emerge, the potential for reuse of these datasets increases. If the data quality is defined at the level of the regulatory standard, subsequent users of the datasets will have a point of reference to understand the quality of the data being reused.

This is supported by Kerr (2000, [64]) who identified the requirement for a data quality framework as a method to inform the development of a data quality strategy for the Ministry of Health in New Zealand [64]. By integrating the framework in the processes of the organisation, business-driven and technical data quality considerations can be combined in order to improve the overall data quality of an organisation [65].

While qualitative data quality assessments have been made, they are typically applied to organisational situations, not to regulatory standards. Therefore when developing the data quality assessment framework for this thesis, a novel approach was taken: data quality indicators relevant to regulatory standards were identified whereby if the indicator was present in the standard, it conveyed a greater level of data quality on the resulting compliance dataset.

In designing a framework to assess regulatory standards, the required level of specificity of the DQ indicators was considered. When assessing a dataset, DQ indicators often assess the metadata associated with a data point. For example, metadata elements of an *E. coli* test may include the analytical method used, the time and date the sample was taken, and the temperature of the sample when it was received at the testing laboratory. However, when assessing a regulatory standard, these metadata elements can only be applied to the *E. coli* requirements of the standard, and not the standard as a whole.

Another example of a specific metadata indicator is the UV validation standards that are allowed by a drinking water standard. By including this metadata element in the framework, the outcome of this assessment would allow the comparison of the data quality of UV disinfection data for water treatment plants. For standards not concerned with water treatment, the metadata element would be given a rating of N/a. However; as with the *E. coli* example, it only assesses one part of the standard; whereas the goal of the framework is to understand the extent to which the standard in its entirety prioritises data quality.

Therefore, DQ indicators were required to assess whether the standard addressed the indicator in a manner that increased the DQ of the dataset. For example, instead of determining if the UV validation method was specified, the indicator would instead ask if allowable validation methods were included in the standard. If a reference to a validation method was included in the standard, this increased the data quality of the resultant compliance data set. This approach better reflected the intent to assess the data quality of regulatory standards at a high level.

By taking a more qualitative approach to the development of the DQ indicators, the role of the DQ dimensions was changed. Where DQ indicators in the literature were concerned with specific metadata, DQ dimensions were used as a way to group these indicators in a qualitative manner. By identifying qualitative DQ indicators, DQ dimensions are not required.

2.3 Application of Data Quality Assessment Methods

If the data quality requirements of a regulatory standard can be determined, the likely data quality of subsequent compliance datasets can be derived. In order to determine the data quality requirements of a standard, first the data quality indicators specific to regulatory standards must be defined. A framework is then built to determine if these data quality indicators are present or absent in a standard. The resulting assessment of the standard summarises its data quality requirements. In this way, the framework provides two levels of analysis: firstly, it describes the expected data quality of a compliance dataset, allowing datasets to be more easily compared; and secondly, it allows the data quality requirements of different standards to be compared.

By first applying the framework to the current DWSNZ and the proposed new drinking water standards[7], the data quality indicators present in both can be compared. By widening the regulatory standards being assessed to environmental waters and wastewater resource consents, the common data quality indicators can be identified, allowing a greater understanding of the potential interoperability and reusability of the compliance datasets.

Including environmental monitoring standards from Australia allows an external benchmark to be made; the assessment of the EU Directives furthers this, and illustrates how data quality indicators can prioritise interoperability and reusability, providing utility to New Zealand's nascent regulator as a model of how standards could align with each other across the water industry. Finally, assessing the data quality indicators of three waters standards against GHG emission standards shows how the data quality requirements of two industry types can be compared.

[7] The draft published by Taumata Arowai in December 2021

3 Methodology

3.1 Overview

While the empirical data quality assessment methods outlined in Chapter 2 are complex when applied to research datasets, they may be suitable to apply to datasets generated in the pursuit of regulatory compliance. In assessing compliance datasets, the data curator is also the data user, as both roles are filled by the regulatory body determining if the supplier has achieved compliance. In addition, the individual data points of a compliance dataset are not required to be assessed; instead, the data quality attributes of the regulations themselves can be analysed to determine the expected quality of the datasets submitted to the regulator. And, since the context – regulatory compliance – is static, the hierarchical classification of data quality attributes from different standards can be compared. This will be explored in a case study in Chapter 5, where the projected data quality outputs of a regulation that is currently being written (the draft Drinking Water Quality Assurance Rules published by Taumata Arowai on 20 December 2021 [46]), will be compared to the data quality outputs of the current standard that it will replace (the DWSNZ 2005, 2018 revision [42]).

The method of assessing the data quality of regulatory standards will draw on aspects of the empirical DQ methodology, and methods used to assess the FAIRness of datasets. The resulting framework will allow a regulatory standard to be assessed in a manner that summarises the extent of its data quality attributes, as well as allowing it to be compared to other standards. The aspirational nature of the FAIR principles will allow the framework user to determine how the data quality of a regulatory standard could be increased; the comparison with other standards will allow give examples of how to achieve this.

The development of this method follows these steps:

1. Determine DQ groups (analogous to hierarchical classifications)
2. Identify DQ indicators of environmental monitoring compliance data
    i. Identify indicators from the literature
    ii. Use the DWSNZ (2018)[8] as a reference standard
3. Categorise these DQ indicators into the groups determined in 2(i)
4. Use a binary (yes/no) scoring system for each indicator
    i. For some standards, indicators may not be included, for example continuous monitoring requirements may not be present. Such instances will result in a score of not applicable (N/a).
5. Write a closed question that can be applied to a regulatory standard

[8] The author has extensive experience with these standards and has worked as a data creator, data user, and data curator.

  i. Justification for the score will be included in the framework, for greater comparison between standards.

These steps create the framework that allows the DQ indicators of a regulatory standard to be assessed.

The application of this framework assessment method is as follows:

6. An environmental monitoring standard will be assessed using the framework questions. If the standard demonstrates that it includes that DQ indicator , it is given a score of 1 for yes and the justification is also given (reference to the part of the standard where it answers the question.
    i. If the standard does not demonstrate the DQ indicator, it is given a score of 0 for no. Justification should also be entered.
    ii. As for 4(i), some indicators may be categorised as N/a.
    iii. This is designed as an Excel spreadsheet.
7. The framework is then used to assess additional environmental monitoring standards concerned with the Three Waters industry and greenhouse gas emission calculations
    i. Standards from New Zealand, Australia, and the European Union will be assessed
    ii. Results are entered into the framework spreadsheet
8. A summary table of the scores for each DQ indicator will be compiled
9. As each group will have a different number of DQ indicators, the percentage of indicators will be used to compare groups between standards.
10. The results of the framework can then be used to compare the data quality requirements of the standards that were assessed.
    i. A case study of the existing and proposed new standards for drinking water in New Zealand will be made.

3.2 Method Development

3.2.1 Identifying Data Quality Indicators

Both empirical data quality methods and FAIR assessment methods obtain DQ indicators from data users. As the author has extensive work experience as a data user of compliance data collected for the DWSNZ, DQ indicators were compiled using the example of a typical compliance sample “life cycle” process. This approach identified which points in that process would, if they were not required, reduce the regulator’s trust in the validity of the sample result as part of a compliance report. These indicators are included as Table 3.1 below.

Data Quality Indicator
Sample collection method
Sampling technician training
Sample container requirements
Sample location requirements
Sample location GPS requirements
Sample transport requirements
Sample delivery time frame requirements
Field data instrument calibration requirements
Allowable time between continuous monitoring data points
Continuous monitoring instrument calibration frequency and technician training
Continuous monitoring instrument GPS location
Laboratory analytical test method
Laboratory accreditation requirements
Laboratory technician training
Uncertainty reporting for laboratory analysis
Data transfer from instrument or testing laboratory to supplier
Calculation method used
Calculated data: input requirements
Unit system
GPS datum
Parameter units
Method of sending data to the regulator: electronic or manual?
Can data be sent directly from the testing laboratory to the regulator?
Does the regulator's database have the capacity for M2M transfer
Regulator's file type requirements
Data standard referenced
Data ownership transfer from supplier to regulator

Table 3.1 Data Quality Indicators

3.2.2 Defining Data Quality Groups

By grouping DQ indicators in process stages, the framework allows standards to be compared with each other at a hierarchical level as well as by DQ indicator. As the DQ indicators were identified from the "life cycle" of a compliance sample, the process stages were defined as follows:

- How is the sample collected?
- How is the sample transported to the analytical laboratory?

- How is it analysed at the laboratory, or how is it analysed by an on-site meter (for example, continuous monitoring)?
- How does the data get from the result processing place (the continuous monitoring meter or the testing laboratory) to the supplier?
- How does the data get from the supplier to the regulator?

These process stages can be visualised in a flow chart, illustrated by Figure 3.1.

Once the stages of the compliance sample process were identified, the DQ indicators in Table 3.2 were allocated to the process stages. This is illustrated by Figure 3.2.

Finally, the DQ indicators can be combined into Groups which allow a high-level overview of the way a regulatory standard incorporates data quality. This is illustrated by Figure 3.3.

The Groups are a way of categorising the process stages that are common to any environmental monitoring process. This allows the framework to be applied to standards from different industry types.

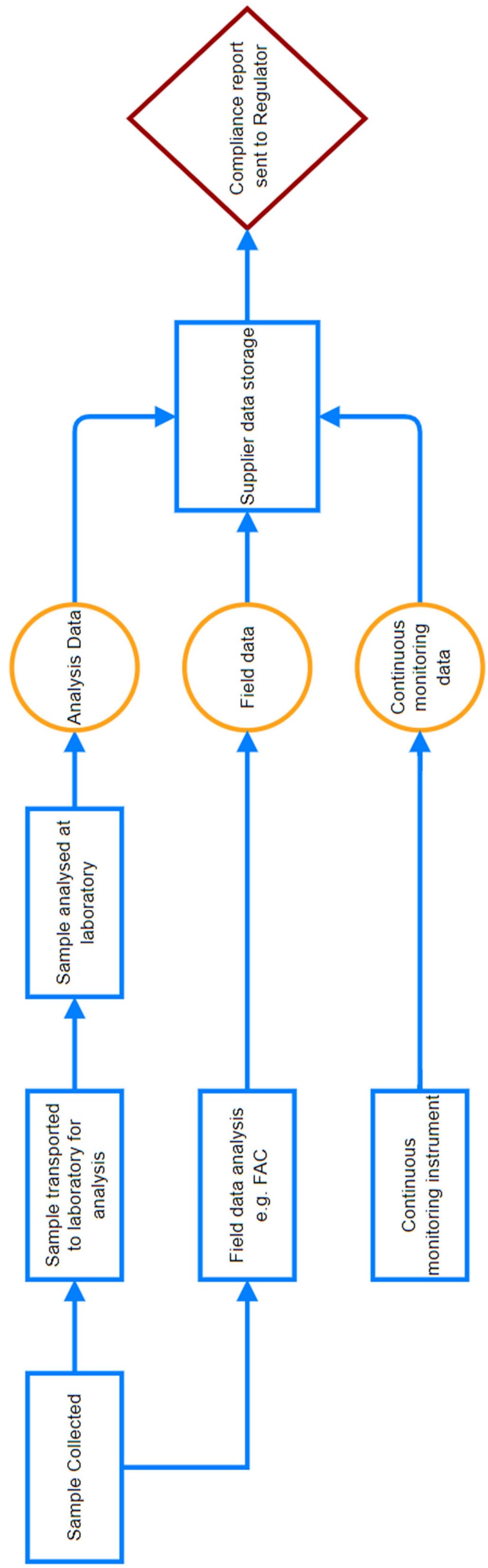
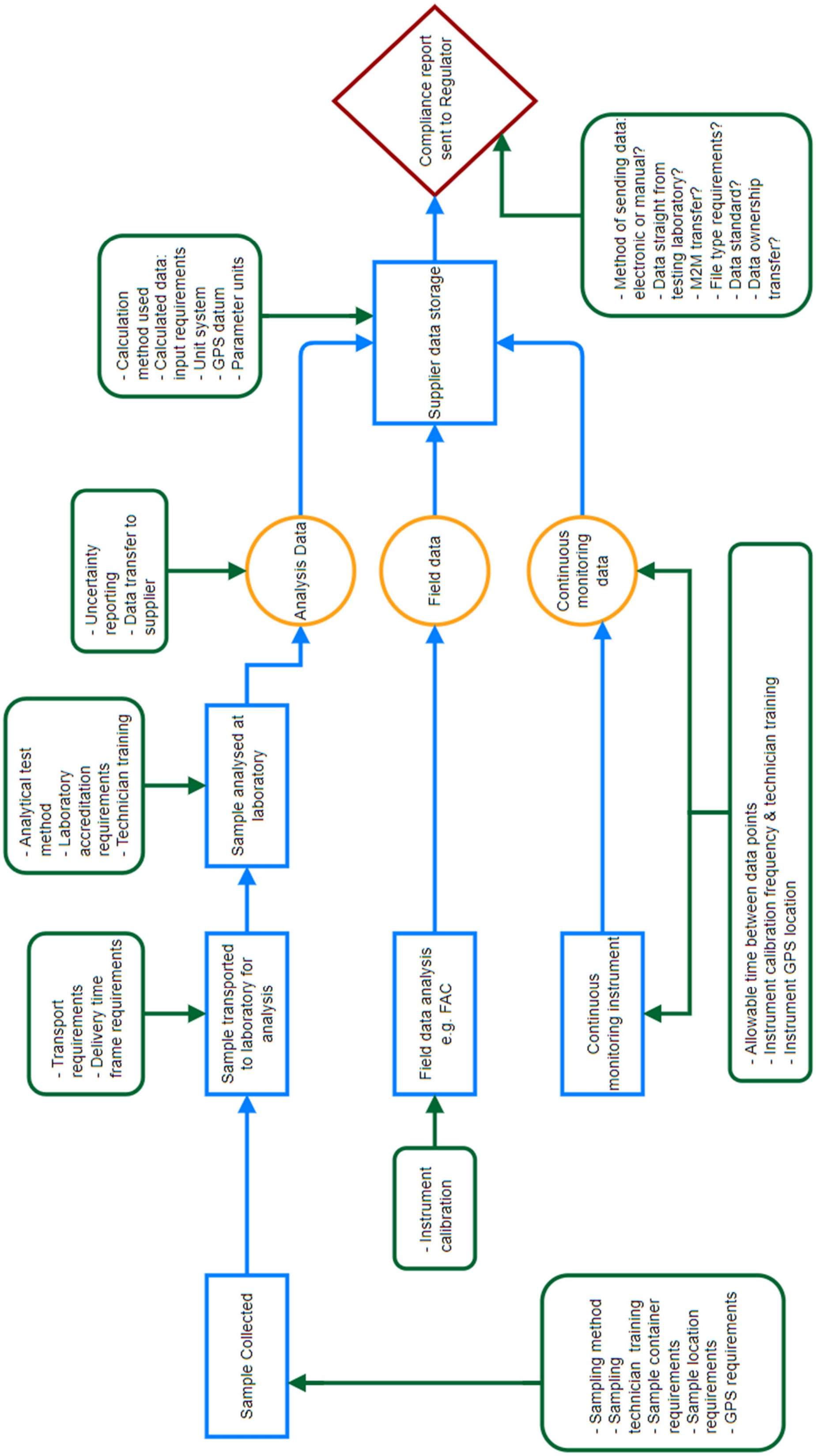


Figure 3.1 Compliance sample process stages



:

Figure 3.2 Data quality indicators allocated to compliance sample process stages

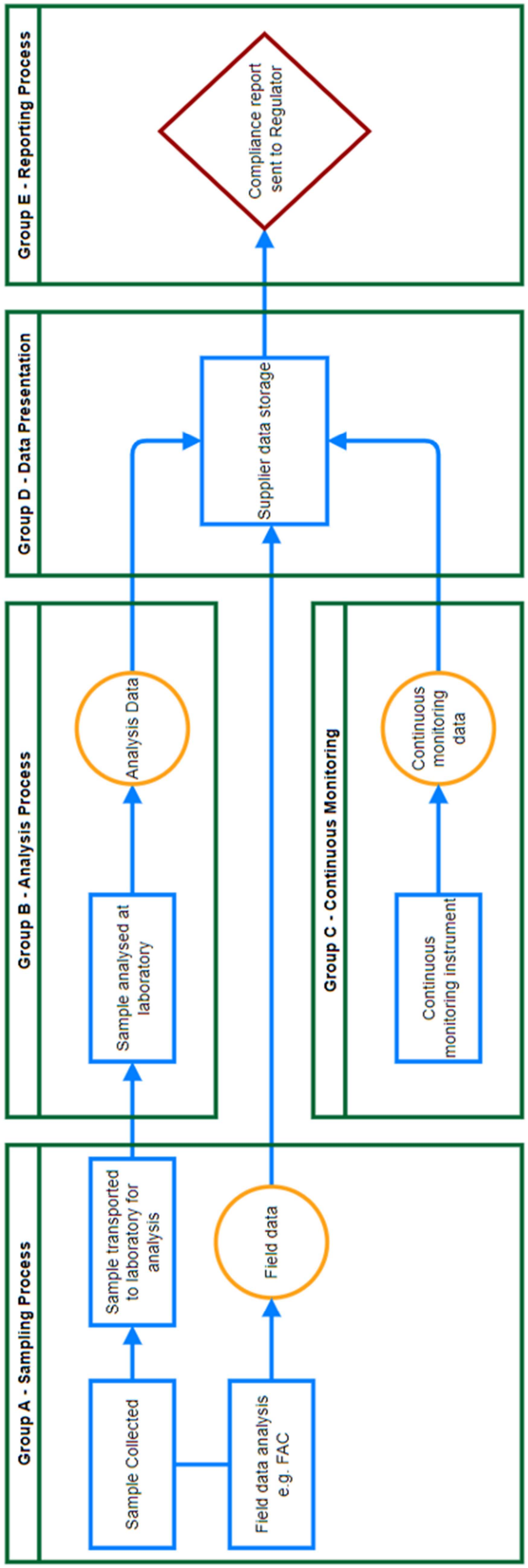


Figure 3.3 The compliance sample process stages organised into groups

3.2.3 Group A: Sampling Process

This group focuses on the data quality of the process of collecting a sample for analysis. This is different from continuous monitoring as a sample represents a fixed point in time. These samples are important because not all parameters can be measured in the field; some determinands require analysis that can only be carried out at a laboratory, such as the Colilert method of testing for *E. coli*, where incubation at a set temperature in media is required [66]. Another consideration is that while some determinands could be measured in the field, the accuracy of the test method might be too low to be accepted by the compliance standards. For example, if the accuracy of the test method is +/- 0.2mg/L but the maximum allowable value (MAV) is 0.01mg/L, any results obtained using this test method would always be out of compliance due to the level of uncertainty of the measurement.

The focus of this group is to provide more certainty about the quality of the sample result. Therefore, if a standard has a high number of DQ indicators from this group, there is more certainty that the sample result is accurate and believable. If the score for this group is low, it can indicate that the data submitted for that standard may have a higher level of variability. This is because even though two datasets are submitted for the same compliance reason, if there is no standard method used, then one cannot be sure that the results of the two datasets are comparable.

The indicators, descriptions, and questions for this group are included in Table 3.2 below.

Code: A1	**Indicator:** Sampling Method
Descriptor	The method of collecting the sample. May include reference to a standard methodology or may refer to certain requirements such as aseptic technique.
Question	Does the standard refer to or include one of the following: - A standard sampling method that should be followed - Requirement for the sampling and testing laboratory to have a standard or accredited method - A description of a specific sampling method
Effect on Data Quality	If the method is prescribed, then the sample process must utilise this method in order for the data to be compliant. Where the method is prescribed, results are unlikely to be the result of inconsistent sampling procedures, and so are more reliable. In addition, if datasets use the same sampling method, their data can be compared.

Code: A2	**Indicator:** Sample container requirements
Descriptor	The requirements of the sample container, for example if it requires a preservative or if it must be sterile. May include reference to a standard methodology.
Question	Does the standard refer to or include one of the following: - The container requirements of a standard sampling / analytical method - Requirement for the sampling and testing laboratory to have a standard or accredited method that includes container requirements - A description of specific container requirements
Effect on Data Quality	Some determinands require samples to have a preservative added to prevent degradation of the determinand between the sample being taken and delivered to the testing laboratory. As well, microbiological determinands are typically required to be sampled in sterile containers. If the standard does not address sample container requirements, then the results may not be as reliable. If this indicator is addressed in the standard, there is an increased ability to compare samples with others that had similar sample container requirements.
Code: A3	**Indicator:** Sample transport requirements
Descriptor	The requirements of the sample transport method, for example if it must be stored at or below a certain temperature, or away from light sources. May include reference to a standard methodology.
Question	Does the standard refer to or include one of the following: - The transport requirements of a standard sampling / analytical method - Requirement for the sampling and testing laboratory to have a standard or accredited method that includes transport requirements - A description of specific transport requirements
Effect on Data Quality	If the determinand being sampled requires that the sample be transported under certain conditions, these should be noted. Otherwise the sample result may not be valid, as the influence of an external factor (such as temperature) may alter the sample so that it is no longer representative of the water being tested. As with indicator A2, this increases the ability to compare samples with others where this indicator was addressed.

Code: A4	**Indicator:** Sample delivery time frame requirements
Descriptor	Requirements of the time frame regarding sample collection and delivery to the testing laboratory, e.g. the sample must be delivered to the testing laboratory within 24 hours of the sample time. May include reference to a standard methodology.
Question	Does the standard refer to or include one of the following: - The delivery time frame requirements of a standard sampling / analytical method - Requirement for the sampling and testing laboratory to have a standard or accredited method that includes acceptable sample delivery time frames - A description of specific sample delivery time frames
Effect on Data Quality	The determinand being tested for may require the analysis to commence within a certain time frame. For example, an analysis for *E. coli* may be required start within 4 hours of the sample time, and no more than 24 hours after the sample was taken. The outcome is an increased ability to compare samples. Related to this indicator is the consideration that the laboratory method may only be accredited for samples delivered within a certain timeframe.
Code: A5	**Indicator:** Sampling location
Descriptor	Requirements for where the sample must be taken in order to be representative of the process or environment being sampled.
Question	Does the standard describe the sampling location requirements of field samples taken for compliance monitoring?
Effect on Data Quality	Standards describe where a sample should be taken from. For example, the DWSNZ require that an *E. coli* sample at a water treatment plant should be taken “at a point after the prescribed disinfection contact time has elapsed but before the first consumer” ([42],Section 4.2.7.1, page 25). Therefore the *E. coli* results will be representative of water leaving the treatment plant, and not water at a point in the treatment process prior to disinfection, or water in the distribution network. This results in a sample that is representative of the process it is being used to evaluate. The outcome is increased reusability of the data. If the sample called "WTP [name] Treated Water" is required to be taken "at a point after the prescribed disinfection contact time has elapsed but before the first consumer", it gives location context to the sample; a new user of the data would therefore not be able to interpret it as water prior to disinfection.

Code: A6	**Indicator:** Sampling technician qualification
Descriptor	Requirement that the sampling technician has had sufficient training or holds a relevant qualification.
Question	Does the standard require the sampler to hold an appropriate qualification? Alternatively, does the standard require the sampler to be accredited in the sampling method (to show sufficient training)?
Effect on Data Quality	This indicator shows that the technician taking the sample is required to have a qualification that shows they are adequately trained. This provides certainty about the validity of the result.
Code: A7	**Indicator:** Field data instrument calibration
Descriptor	Calibration requirements for instruments used to collect field data, for example the interval between calibrations.
Question	Does the standard specify calibration requirements for field data instruments?
Effect on Data Quality	If the standard has requirements for field data instrument calibration, it means that field data is more reliable and likely to be accurate. If the standard has no calibration requirement then the data point may not be as reliable. This increases the validity of the data point, its potential for reuse, and the ability to compare samples.
Code: A8	**Indicator:** Sample GPS data
Descriptor	Requirements for the GPS coordinates of the sample location.
Question	Does the standard require the GPS coordinates of the sample location to be recorded?
Effect on Data Quality	This indicator considers if standard requires the GPS coordinates of the sample location to be part of the sample data, as GPS coordinates would allow that location to be verified. This indicator conveys more trust that the sample was taken at the correct location, and provides greater reusability of the dataset.

Table 3.2 Group A Data Quality Indicators

3.2.4 Group B: Analysis Process

This group assesses the way those field samples are processed. For example, if the method for testing a sample was not described, then it would allow many different methods to be used. For *E. coli* tests, it means the type of media and the temperature and duration of incubation is no longer standard. In practice, longer incubation times may result in higher *E. coli* counts; this would be misrepresentative of both the water that was tested with the long incubation period, but also the water that was tested with the short incubation period.

This group therefore focuses on laboratory analysis procedures. If a standard scores highly in this group, it means that the laboratory analysis procedure for any result submitted against that standard is either present in the standard or referred to in another. Often this is addressed by requiring the testing laboratory to hold an appropriate accreditation for the analysis method. For example, in New Zealand a laboratory may be required to hold IANZ[9] accreditation for a particular method in order for the result to be accepted as part of a compliance dataset.

Indicator B6, "laboratory reporting requirements – method of sending data to supplier" refers to the standard requiring that the test results of the compliance sample are to be sent directly from the testing laboratory to the regulator, bypassing the supplier. If this indicator is present, the data quality of the dataset is increased, as the regulator receives data that cannot have been edited in any way by the supplier. This provides the regulator with greater certainty that the data is correct.

The indicators, descriptions, and questions for this group are included in Table 3.3 below.

Code: B1	**Indicator:** Laboratory technician qualification
Descriptor	Requirement that the laboratory technician conducting the analysis has had sufficient training or holds a relevant qualification.
Question	Does the standard require the laboratory technician to hold an appropriate qualification? Alternatively, does the standard require the laboratory technician to be accredited in the analytical method (to show sufficient training)?
Effect on Data Quality	Complementary to indicator A6, this indicator conveys an understanding that the technician performing the laboratory analysis is required to have a qualification that shows they are adequately trained. This provides certainty about the validity of the analysis result.

[9] International Accreditation New Zealand, IANZ

Code: B2	**Indicator:** Laboratory instrument calibration requirements
Descriptor	Calibration requirements for laboratory instruments used to perform an analytical procedure, for example the interval between calibrations.
Question	Does the standard specify calibration requirements for laboratory instruments?
Effect on Data Quality	As with field data instrument calibration, this indicator provides certainty that the result determined by the analysis was accurate and not due to a lack of calibration. This contributes to reusability, sample comparability, and proof of validity.
Code: B3	**Indicator:** Analytical test method
Descriptor	The method used to analyse the sample. May include reference to a standard methodology or may refer to certain requirements such as aseptic technique.
Question	Does the standard refer to or include one of the following: - A standard analytical method that should be followed - Requirement for the testing laboratory to have a standard or accredited method - A description of a specific analytical method
Effect on Data Quality	If the test method specified is an internationally accepted method, the analysis result is more believable than if an untested method was used. In addition, if the method is known, the data point can be viewed alongside any other data points that used that method, so it results in greater interoperability and reusability of the data.
Code: B4	**Indicator:** Laboratory accreditation requirements
Descriptor	Requirement that the testing laboratory holds accreditation for the analyses performed on compliance samples.
Question	Does the standard require the testing laboratory to hold accreditation?
Effect on Data Quality	Laboratories receive accreditation from auditors who assess many aspects of the analytical process such as technician qualifications and how technicians adhere to standardised test methods. If a standard requires an accredited laboratory to be used, it conveys a higher certainty that other indicators are being met, such as intervals between instrument calibrations, and adherence to specific test methods. It is an indicator that supports the DQ of other indicators.

Code: B5	**Indicator:** Laboratory reporting requirements - uncertainty
Descriptor	Requirement for the testing laboratory to report the level of uncertainty along with the sample result.
Question	Does the standard require the uncertainty to be reported with the result?
Effect on Data Quality	The uncertainty of an analytical method conveys the amount by which the final result could vary, due to measurement restrictions. For example, a result of 0.25 mg/L with an uncertainty of ± 0.01 mg/L means the final result could be in the range of 0.24 -0.26 mg/L. Conversely, a result of 0.15 mg/L with an uncertainty of ± 0.1 mg/L could mean the final result is in the range of 0.05 - 0.25 mg/L. In practice, when the minimum allowable value of a determinand is 0.2 mg/L, understanding the uncertainty measure allows the regulator to determine if a result of 0.25 mg/L is compliant, or if the uncertainty calls compliance into question.
Code: B6	**Indicator:** Laboratory reporting requirements - method of sending data to supplier
Descriptor	If a required method of data transfer from the testing laboratory to the supplier is not noted in the standard, then this could be made electronically or manually.
Question	Does the standard prohibit non-electronic data transfer from testing laboratory to supplier?
Effect on Data Quality	Manual data transfer introduces the risk of transcription errors, which reduce the believability of the data point. Manual data transfer also restricts audit checks such as checking the recorded value in a compliance report sent to the regulator against what was sent by the laboratory.

Table 3.3 Group B Data Quality Indicators

3.2.5 Group C: Continuous Monitoring

While samples taken in the field are important, it is also possible to continuously monitor some determinands. Continuous monitoring here refers to when an instrument measures a parameter at fixed intervals. The intervals can either be time based (e.g., every 60 seconds or 15 minutes or 3 hours), or when a change occurs (e.g., a turbidity meter may only take a reading when the turbidity changes by more than 10% (plus or minus) from the last reading).

Depending on the application, this can result in a large amount of data. A standard may not require all of this data to be sent for compliance reporting purposes, but might instead have other requirements.

One example is the time between measurements – in the DWSNZ it is required to be not more than 60 seconds for turbidity measurements taken at the water treatment plant (WTP) ([42], section 3.2, page 16). This is a DQ indicator because if the standard goes on to require that the turbidity of treated water at a WTP is less than 1.0 NTU[10] for 100% of the day, and the maximum interval between readings is 60 seconds, a regulator reading the compliance report is able to extrapolate the number of readings where turbidity was less than 1.0 NTU (if the plant was operational for the full 24 hours, this is 1440 readings). If the maximum interval is not specified, then the readings might be taken once every hour – meaning only 24 readings may be taken. If this indicator was not defined, both of these readings might be considered compliant, but one is a better representation of the quality of the treated water, and therefore is more likely to be reused.

Another aspect of continuous monitoring is the calibration of the meters. If this is not specified by the standard, it means there is less confidence in the data.

The GPS coordinates of the compliance monitoring meter are also a DQ indicator. This is perhaps less important for small water treatment plants (the more relevant indicator would be A5 in Sampling Process, where the standard defines where the determinand should be monitored), but it is important for larger treatment plants, and any sort of environmental monitoring, especially wastewater where the monitoring locations are often required to be a set number of meters upstream or downstream of a discharge point. GPS coordinates would benefit the supplier (they allow confirmation that the sample was taken in the correct location, or if not, how far away it was – a concern for stream sampling after rainfall); the regulator (they allow confirmation that the sample was taken where the standard requires it to be, e.g. 50 metres upstream or downstream of a discharge point with fixed coordinates), and anyone wishing to reuse the dataset.

The indicators, descriptions, and questions for this group are included in Table 3.4 below.

Code: C1	**Indicator:** Time between data points
Descriptor	Continuous monitoring of parameters is carried out by meters installed at monitoring points. Meters will be programmed to capture data points either at fixed intervals, or in response to change away from a baseline. If the standard does not specify the maximum allowable interval between these readings, calculations such as average, maximum, and minimum may not be comparable between datasets.
Question	Is the maximum allowable interval between continuous monitoring data points noted?

[10] Nephelometric turbidity units, NTU

Effect on Data Quality	Due to practical considerations (such as cost, available radio bandwidth, on-site data storage capacity), a supplier's definition of "continuous" may vary. Therefore a standard should define "continuous". This is normally represented as 'the required time between data points'. By defining this, datasets of continuous monitoring data can be compared, and the data has increased reusability.
Code: C2	**Indicator:** Instrument calibration requirements - frequency
Descriptor	The maximum allowable interval between calibrations for continuous monitoring instruments.
Question	Does the standard specify the maximum allowable interval between calibrations for continuous monitoring instruments, or reference that the manufacturer's specifications be followed?
Effect on Data Quality	As with field data instrument calibration, this data quality indicator provides certainty that the results recorded by the continuous monitoring meter were accurate and not due to a lack of calibration. This contributes to reusability, comparability, and proof of validity.
Code: C3	**Indicator:** Instrument calibration requirements - technician
Descriptor	Requirement that the technician performing the instrument calibration has had sufficient training or holds a relevant qualification.
Question	Does the standard require the instrument calibration technician to hold an appropriate qualification or record of training?
Effect on Data Quality	This makes the calibration result more reliable, and as a result, increases the believability of the compliance monitoring data recorded by the meter.
Code: C4	**Indicator:** Instrument GPS location data
Descriptor	Requirement that the continuous monitoring instrument also logs GPS data.
Question	Does the standard require data from a continuous monitoring instrument to include GPS coordinates?
Effect on Data Quality	Similar to the sample collection GPS requirement, the difference here is that if an instrument is required to record GPS coordinates alongside the continuous monitoring data, it confirms that the instrument was located at the correct place, and that the data represents the required process step.

Table 3.4 Group C Data Quality Indicators

3.2.6 Group D: Data Presentation

The DQ indicators in this group are concerned with the way the data gathered in the first three groups is presented. For example, some standards may assess compliance based on a calculated determinand. An example of this is C.t[11] as a measure of the amount of time a disinfectant is in contact with treated water at a WTP before it is distributed. The DQ indicators in this example would consist of considerations such the calculation method – if this is prescribed in the standard, then all calculated data submitted as part of that standard can be viewed as being calculated in the same way. Another DQ indicator would be if a standard requires the input data to be reported alongside the calculated data – this increases the reusability of the data set (while not every data re-user may be interested in contact time, they may find utility in the disinfectant concentration, or the volume of water in the reservoir); this would also assist the audit process – the raw data would not need to be audited, instead it would be present alongside the calculated data.

Other data presentation DQ indicators include the unit system used – if the standard requires the units to be based in the SI[12] system, this should be noted, otherwise uncertainty may arise when comparing datasets. This is linked to the DQ indicator where the standard specifies the reporting units for the parameter. An example of where this might have an effect is an *E. coli* test. *E. coli* in the DWSNZ is required to be reported as MPN/100mL, which means "the most probable number of *E. coli* cells per 100mL of sample". This means that the result cannot be zero, as the test is designed to give the "most probable number"; therefore, the lowest possible result is <1. However, another microbiological test method is to count the number of colonies that are visible on an agar plate (after incubation). This test is measured in "cfu/mL" or colony forming units per 1mL of sample. Because the result is based on a visual assessment of the number of colonies, it can return a result of 0cfu/mL if no colonies are present. These two methods are not comparable, not only because they have different test methods, but because the minimum possible result of <1 is not the same as 0.

Other benefits of this DQ indicator are conversions (e.g., nitrate or nitrogen as nitrate), and orders of magnitude – if the concentration of chlorine is 2mg/L, it could be reported as 0.002g/L; if the reporting units are not specified, making comparison of the two results less straightforward by requiring the data user to convert the data.

Finally, the data presentation of GPS units is a DQ indicator. There are many different GPS datum systems in place around the world. If GPS coordinates are required at any point in the standard, the datum should either be specified by the standard, or required to be presented in the compliance report. If this is not the case, then the data has reduced reusability and interoperability.

The indicators, descriptions, and questions for this group are included in Table 3.5 below.

---

[11] C.t is the concentration of a disinfectant in mg/L multiplied by exposure or contact time in minutes. The unit for C.t is min.mg/L ([42], A1.3.3 page 98).

[12] International System of Units

Code: D1	**Indicator:** Calculation method
Descriptor	Some compliance data may be calculated from sampling results. If this is required, the standard should provide the calculation method, so that the same method is applied to all compliance datasets.
Question	If calculations are required, is the method / formula included in the standard?
Effect on Data Quality	If a standard requires a calculated result to be part of the compliance data, it should provide the calculation method. If not, there is no standardisation between any calculated data points. While most conversions are straightforward, some are complex. By including the calculation method, the reliability of the calculated result increases.
Code: D2	**Indicator:** Calculation - Input data requirements
Descriptor	Requirement for the input data to be reported with the calculated data. Allows the regulator to check the calculation was made correctly.
Question	Does the standard require the input data to be reported with the calculated data?
Effect on Data Quality	If the input data is submitted alongside the calculated data, then it can be easily verified.
Code: D3	**Indicator:** Units - system used
Descriptor	The system of units used by the standard, for example SI units.
Question	Does the standard specify the unit system used?
Effect on Data Quality	Reduces the different ways in which the data can be interpreted, which increases reusability of the dataset.
Code: D4	**Indicator:** GPS datum
Descriptor	If GPS coordinates are required alongside compliance data, the GPS datum used should be specified in the standard or recorded with the coordinate data.
Question	If GPS data is required, is the datum specified?
Effect on Data Quality	If the GPS datum is not specified, it reduces the reusability of the data.

Code: D5	**Indicator:** Units - reporting
Descriptor	Results for compliance parameters should be reported in the same units as the allowable limits in the standard, so that a unit conversion is not required.
Question	Does the standard specify the reporting units for the sample data?
Effect on Data Quality	If the parameter being measured is not reported in the correct units, then it cannot be compared with other data points for that parameter. An example of this is microbial monitoring – two commonly used units are “MPN/100mL” and “cfu/mL”. MPN/100mL means ‘most probable number of cells per 100mL of sample”. The result cannot be zero; the lowest possible reading is <1. Cfu/mL means “colony forming units per 1mL of sample”. The result can be zero if no colonies are detected.

Table 3.5 Group D Data Quality Indicators

3.2.7 Group E: Reporting Process

This group looks at the DQ indicators of communicating the compliance reporting data. The inclusion of these DQ indicators may be a relatively new development for some environmental monitoring standards. The reason why this is a measure of data quality is because it shows the points in the process (from data generation to regulator) where the data point could be changed, either by accident (mis-transcription), or deliberate action (unfavourable data points are not reported).

One aspect of this is to check whether the standard requires that any laboratory testing data is sent directly from the testing laboratory to the regulator. The related DQ indicator of this is whether this has to be done electronically or if allows a transcription step to be introduced. For example, if the testing laboratory sent a result to the regulator as a static report (e.g., a paper report, or an electronic such as PDF that requires transcription), it requires the regulator to manually enter this data into their database – introducing a possible point of error.

By requiring the data to be sent directly from the testing laboratory to the regulator, it also removes the possibility that the supplier could change the data. If the data can be changed at a point of transfer, then the quality of the data in the regulator’s database is not as high as a dataset that was sent straight from the laboratory to the regulator without any transcription requirements. This requires the regulator to have a method of receiving electronic data, which is captured in other indicators in this group, including if the regulator specifies the file types for electronic data transfer, or alternatively if it references an international data standard requirement.

Therefore, this group provides an indication of the level of trust that can be afforded to the dataset when it is received by the regulator. If all other DQ indicators in the previous groups were present, yet the standard still allows the dataset to be edited, then the data quality of the dataset is not being prioritised by the standard.

Finally, there is an indicator included in this group which relates to reusability: E7, which addresses ownership of the compliance dataset. It asks if the standard references the ownership of the data once it has been received by the regulator. If this is addressed in the standard, then the ability of the data to be reused is more transparent. If this is not present in the standard, then not only does it inhibit the reuse of the data, it raises where the responsibility lies if there is an error in the data. For example, this DQ indicator considers which party is responsible for updating the dataset if it has been published by the regulator, but during an audit the laboratory discovered that their incubators for processing *E. coli* samples were all 5°C colder than required? By referencing the transfer of ownership of the data, it conveys the meaning that "this data was considered correct at the time it was sent to the regulator". This allows the regulator to say they own the data that was sent to them, and are not responsible for any effects of subsequent reuse if it is found that the data sent to them was incorrect.

The indicators, descriptions, and questions for this group are included in Table 3.6 below.

Code: E1	**Indicator:** Method of sending data to Regulator - electronic data transfer
Descriptor	The data may be sent to the regulator by many different methods. Manual (or non-electronic) methods such as paper reports are the most susceptible to human errors such as transcription errors. Standards that require datasets to be sent electronically convey a higher level of data quality to that dataset as its reusability is higher, as well as its credibility.
Question	Does the standard prohibit non-electronic data transfer?
Effect on Data Quality	Any data transfer method other than "laboratory to regulator" or "data historian to regulator" means that the data could be edited, which reduces the reliability of that data.
Code: E2	**Indicator:** Method of sending data to Regulator - data pathway
Descriptor	The standard may require the sample analysis data to be sent directly from the testing laboratory to the regulator, so the supplier is not able to edit the data.
Question	Does the standard require lab data to be sent directly to the regulator (i.e., a data transfer connection between the testing lab and the regulator, bypassing the supplier)?

Effect on Data Quality	If the testing data is able to be sent from the laboratory directly to the regulator, it conveys a higher level of confidence to the regulator that the data is being accurately reported. If the testing data is able to be sent from the supplier to the regulator, the regulator does not have access to the original data, and cannot be certain that it has been accurately transcribed. This would require an audit of the supplier where the original laboratory report is checked.
Code: E3	**Indicator:** Method of sending data to Regulator - M2M pathway
Descriptor	Machine to machine (or M2M) data transfer removes the likelihood of transcription errors being made. If this is possible, the standard should describe the method, so that there is not another risk of transcription error.
Question	Does the standard describe a method for M2M data transfer from the supplier's data capture system to the regulator?
Effect on Data Quality	If the testing data is able to be sent from the supplier's data capture system directly to the regulator, it conveys a higher level of confidence to the regulator that the data is being accurately reported. This reduces transcription errors and increases data believability.
Code: E4	**Indicator:** Method of sending data to Regulator - M2M requirement
Descriptor	Data transfer from supplier to regulator is only allowed by M2M transfer.
Question	Does the standard require M2M data transfer from the supplier to the regulator?
Effect on Data Quality	If the testing data is required to be sent from the supplier's data capture system directly to the regulator, it removes the risk of transcription errors, and ensures the data is accurate.
Code: E5	**Indicator:** Method of sending data to Regulator - file types
Descriptor	If a certain file type is required by the standard, it should be noted. This also increases the reusability of the datasets.
Question	Where electronic data transfer is allowed by the standard, is/are the allowable file type(s) specified?
Effect on Data Quality	If the file type is specified, it ensures data sent to the regulator is in a format that can be utilised by the regulator's database. This supports interoperability.

Code: E6	**Indicator:** Data storage by the regulator
Descriptor	The data format requirements of the standard. May be described or a reference to a data standard may be made.
Question	For electronic data transfer, is a data standard referenced?
Effect on Data Quality	Data standards describe format requirements of databases. If a data standard is referenced, it increases the interoperability of datasets submitted to the regulator.
Code: E7	**Indicator:** Data access and use rights
Descriptor	If there is a change of ownership of the compliance dataset once it is sent from the supplier to the regulator, it should be noted in the standard. This allows subsequent reuse of the data, as it will be appropriately referenced.
Question	Does the standard note any transfer of ownership of the data once it is sent to the regulator?
Effect on Data Quality	If the standard specifies the data owner of the dataset once it is sent to the regulator, it ensures that it is only accessible and interpretable by authorised users in a specific context of use.

Table 3.6 Group E Data Quality Indicators

3.3 Method Application

The goal of the framework is to assess the data quality of compliance reporting; therefore, the focus of the evaluation was on the standard's compliance criteria and the associated monitoring requirements. Some standards require compliance reports to include details of any remedial actions taken, however as these are typically submitted as incident reports[13] due to their variable nature they are unlikely to be standardised in the same way that compliance monitoring is.

The framework also includes a column for Justification, where after the score has been given, notes can be added which refer to the relevant part of the standard that was assessed, and the reasoning behind the score. By including this reference, the framework will act not only as a method of comparison, but also as a point of reference for the author of the comparison.

[13] Based on personal work experience – the author has written incident reports for water treatment processes in Auckland, New Zealand and submitted them for compliance assessment.

3.3.1 Framework Scoring System

The DQ indicators were designed to be answered by yes/no questions, so that the application of the framework would be consistent across standards. If the standard provides evidence of including the indicator, it is given a score of 1 for yes. If no evidence is found in the standard, then the indicator receives a score of 0 for no.

However not every DQ indicator will be relevant to each standard – most notably, the indicators in Group C – continuous monitoring. Some standards reviewed did not have any reference to continuous monitoring data. For these standards, the indicators in Group C were all categorised as not applicable (N/a).

The standards are assessed against every indicator in each group. The summary table provides the score for each group, along with a weighted percentage and a non-weighted percentage. This is to account for the indicators that returned a score of N/a. For example, the NZ Carbon standard [67] provides monitoring methods for on-site analysis of parameters. Due to the analysis method, the samples are not transported to a testing laboratory, therefore the DQ indicators involving sample transport are considered to be N/a. Therefore, for Group A, only 5 out of 8 indicators are relevant to the standard. The non-weighted percentage is calculated as follows:

$$\frac{\text{(Number of indicators in Group A with a score of 1)}}{\text{(Total number of indicators in Group A)}} \; x\ 100$$

The weighted percentage removes the N/a indicators:

$$\frac{\text{(Number of indicators in Group A with a score of 1)}}{\text{(Total number of indicators in Group A)} - \text{(Number of indicators in Group A with a score of N/a)}} \; x\ 100$$

The non-weighted percentage is 50%, and the weighted percentage is 80%.

Both the weighted and non-weighted percentages are presented to show which standards have indicators that were N/a. This comparison is useful when viewed alongside the justification comments, as it allows an assessor to see why an indicator was N/a in a standard.

3.3.2 Primary and Secondary Standards

In applying the framework to the standards, it was not always possible to determine an answer for each DQ indicator from the standard by itself. This was because the standards sometimes referred to others as guidelines. For example, the EU Carbon Directive referred to the Intergovernmental Panel on Climate Change (IPCC) 2006 guidelines on methodologies for measuring GHG emissions ([54], Annex V, part 3, page 65).

The way this was managed by the framework application was as follows.

A primary standard was chosen to be assessed. This standard had the purpose of setting out compliance requirements for environmental monitoring. If the primary standard referred to a secondary standard in such a way that meant that compliance would be assessed based on methodologies or requirements in the secondary standard, then the secondary standard was checked. For example, the EU Carbon Directive received a score of 1 for indicator B3 (analytical test method), because the EU Carbon Directive considered compliance to be achieved if the methods used to monitor emissions were in line with the IPCC. The IPCC contains an exhaustive list of reference methodologies for both sampling and analysis, and so this DQ indicator was considered to be present in the EU Carbon Directive.

The justification for each DQ indicator was particularly critical for such examples, and provides a useful second level of comparison as well. For example, all three carbon standards assessed scored a 1 for indicator B3. When reading the justification notes for each, they show that the NZ carbon standard uses the methodology in the United States Environmental Protection Agency (USEPA) standard 7E; the Australian carbon standard includes its own methodology; and the EU Directive uses the IPCC methodology. So while these standards include methodologies, they are all different, and so additional translation work may be required to achieve interoperability between the datasets.

4 Framework Application

4.1 Regulatory Standards Assessed by the Framework

The framework was applied to the following standards:

Area	**Industry Type**	**Standard Name**	**Code**
New Zealand	Drinking water	Drinking-water Standards for New Zealand 2005 (Revised 2018)	DWSNZ
New Zealand	Drinking water	Drinking Water Quality Assurance Rules 20 December 2021 (Draft)	NZ DW TA
New Zealand	Water Takes (environmental)	Resource Management (Measurement and Reporting of Water Takes) Regulations 2010	NZ EnvW
New Zealand	Wastewater	ECAN Consent CRC051724[14]	NZ WW Chch
New Zealand	Wastewater	ECAN Consent CRC030999.1[14]	NZ WW Ash
New Zealand	GHG emissions	Climate Change (Stationary Energy and Industrial Processes) Regulations 2009	NZ Carbon
Australia (New South Wales)	Drinking water	Drinking Water Monitoring Program December 2005 (updated October 2011)	AU NSW DW
Australia (Queensland)	Environmental Waters	Monitoring and Sampling Manual 2018: Environmental Protection (Water) Policy 2009	AU QLD EnvW
Australia	GHG emissions	Carbon Credits (Carbon Farming Initiative – Domestic, Commercial and Industrial Wastewater) Methodology Determination 2015	AU Carbon
European Union	Drinking Water	Directive (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020 on the quality of water intended for human consumption	EU DW
European Union	Wastewater	Council Directive of 21 May 1991 concerning urban wastewater treatment	EU WW

[14] As there are no country-wide regulations for wastewater monitoring in New Zealand, two current consents have been reviewed instead. These consents apply to different councils but were both issued by Environment Canterbury (ECAN), who in this example fills the role of regulator.

Area	Industry Type	Standard Name	Code
		(91/271/EEC), amended[15] in 1998, 2003, 2008, and 2013	
European Union	Groundwater	Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy; and Directive 2006/118/EC of the European Parliament and of the Council of 12 December 2006 on the protection of groundwater against pollution and deterioration	EU GW
European Union	GHG emissions	Regulation (EU) 2018/1999 of the European Parliament and of the Council of 11 December 2018 on the Governance of the Energy Union and Climate Action, amending Regulations (EC) No 663/2009 and (EC) No 715/2009 of the European Parliament and of the Council, Directives 94/22/EC, 98/70/EC, 2009/31/EC, 2009/73/EC, 2010/31/EU, 2012/27/EU and 2013/30/EU of the European Parliament and of the Council, Council Directives 2009/119/EC and (EU) 2015/652 and repealing Regulation (EU) No 525/2013 of the European Parliament and of the Council	EU Carbon

Table 4.1 Regulatory standards assessed by the data quality framework

Standards concerning drinking water monitoring requirements include the DWSNZ and the draft Drinking Water Quality Assurance Rules (DWQAR) published by Taumata Arowai. These two standards both apply at a national level; as a counterpoint, the NSW Health monitoring program from Australia was also assessed, as this applies at the state level and not at a national level. As the Australian

[15] Amendments:
Commission Directive 98/15/EC of 27 February 1998
Regulation (EC) No 1882/2003 of the European Parliament and of the Council of 29 September 2003
Regulation (EC) No 1137/2008 of the European Parliament and of the Council of 22 October 2008
Council Directive 2013/64/EU of 17 December 2013

Drinking Water Guidelines do not prescribe any monitoring requirements or compliance reporting, they were not assessed.

In New Zealand there is no set of national-level regulations for wastewater monitoring requirements. Under the RMA, regional councils assign resource consents (RCs) to organisations operating wastewater treatment plants. However, as the new Three Waters regulator, Taumata Arowai are expected to write a set of national-level regulations specific to wastewater monitoring. Therefore, in order to understand the current data quality requirements of existing RCs, the framework has been applied to two existing RCs. Both RCs were written by Environment Canterbury (ECAN) but were issued to two different councils. This application of the framework explores whether it can be used to assess the DQ requirements of RCs as well as regulatory standards. The current utility of this application would be for regional councils managing new and existing consents, as it would allow the data quality requirements to be standardised across the RCs. In addition, should Taumata Arowai commence writing wastewater monitoring regulations, this application could provide an understanding of the current data quality indicators present in wastewater monitoring consents, and what the impact on the regional councils may be if the DQ indicators were changed.

As environmental monitoring of source waters is linked to drinking water monitoring, regulatory standards concerning this industry have also been assessed. Of note is the Monitoring and Sampling Manual 2018: Environmental Protection (Water) Policy 2009 for the state of Queensland in Australia. This standard sets out monitoring requirements for samples that may be used in prosecution environments, and so has particularly stringent data quality requirements. In comparison, the Resource Management (Measurement and Reporting of Water Takes) Regulations 2010 for New Zealand is more closely linked to the DWSNZ, and so the DQ indicators present are different.

Due to the increasing focus of greenhouse gas emissions monitoring for wastewater treatment processes, regulations concerned with GHG emissions calculations have also been assessed. While in New Zealand there is no GHG emission calculation methodology specific to wastewater treatment processes, Australia does have a regulatory standard that addresses this. Unlike the drinking water standards, this regulatory standard applies at a national level and not a state level, providing another point of comparison.

While in Australia some standards apply at a state level and some at a national level, European Union Directives apply to all Member States. Therefore these Directives ensure that the compliance datasets generated are able to be compared with each other, increasing their interoperability and reusability. By assessing the EU Directives for drinking water, wastewater, groundwater, and GHG emissions monitoring, the data quality indicators that contribute to dataset reusability and interoperability may be determined. Using this as a point of comparison with the Australian and New Zealand standards allows regulators such as Taumata Arowai to not only benchmark proposed new standards against international standards, but also provides the justification for how data quality in one standard is greater than another.

4.2 Application of the Data Quality Assessment Framework

This section contains the notes for the specific standards that were assessed, along with the score for each indicator and, where applicable, the justification for this score. The completed framework includes the following:

- DQ indicator codes
- DQ indicator groups
- DQ indicator names
- DQ indicator descriptions
- The question applied to each standard to determine if the DQ indicator was present
- Explanation for why the presence of the DQ indicator increases the DQ of the resulting compliance dataset
- DQ indicator scores for the standards listed in Table 4.1
- Justifications for these scores

The framework is designed as a spreadsheet and is included in Appendix 1 as A1.1.

4.2.1 New Zealand – Drinking Water – Current Standard

Legislation: Drinking-water Standards for New Zealand 2005 (Revised 2018) [42]

Code: DWSNZ

This standard was primarily assessed using Section 3, which, as outlined in the structure of the document, “discusses compliance with and transgressions of the water quality standards (determinands), and the general criteria used to determine whether a water supply complies with the DWSNZ” ([42], section 1.3, page 2).

Section 3 has more broad compliance requirements, such as the following: “sample sites must be representative of the water being tested. Procedures for sample collection, preservation, transport and storage, test methods and reporting must be agreed beforehand with the Ministry of Health recognized laboratory that will carry out the analysis” ([42], section 3.1, page 13), and “the steps necessary to demonstrate that a drinking-water supply is in bacterial, protozoal, cyanotoxin, chemical and radiological compliance with the DWSNZ are defined in their specific compliance criteria sections” ([42], section 3.1.1, page 13).

The justification notes for this standard are therefore important to show in which part of the standard the DQ indicator was described. In most instances the DQ indicators were present in Sections 3 and 4 (which do not concern specific treatment processes), and the appendices.

This approach allows more detailed comparisons to be made with other standards, and will also be utilised when assessing the December 2021 draft of the Drinking Water Quality Assurance Rules published by Taumata Arowai, allowing a more in-depth level of comparison between the existing standard and that which is being written to supersede it.

Unless otherwise specified, the document referenced in Table 4.2 is DWSNZ [42].

Code	Score	Justification
A1	1	Section 4.2.6.1, page 25: "Procedures for sample collection and storage, testing and reporting must be appropriate (sections 3.1, 3.2)." Section 4.2.6.2 details specific requirements for *E. coli* sampling.
A2	1	As for A1
A3	1	As for A1
A4	1	As for A1

Code	Score	Justification
A5	1	Section 4.2.6.1, page 25: "Sampling sites and frequencies are discussed in sections 4.2.7 and 4.2.8 (water leaving the WTP[16]), and 4.3.3 and 4.3.4 (water in the DZ[17])".
A6	1	Section 3.1.1, page 14: "The DWA[18] must assess the competence of the analyst for commonly performed treatment plant or distribution system analyses (field tests). Analysts must be certified as competent if carrying out compliance testing."
A7	1	Section A2.3.2 of Appendix 2 (page 104) describes the standardisation requirements for portable turbidimeters.
A8	0	Not present in this standard.
B1	1	Section A2.1 of Appendix 2 (page 103) allows laboratories to "use test methods for which IANZ[19] has assessed them and found them to be competent to perform".
B2	1	Section A2.3.1 of Appendix 2 (page 104) describes the standardising requirements for bench-top turbidimeters.
B3	1	Section 3.1.1 page 15: "Laboratories conducting chemical tests may use the test methods for which IANZ has assessed them and found them to be competent to perform. Laboratories conducting tests for bacteria in drinking-water compliance need to use a referee method specified in the DWSNZ, or a method that has been calibrated against a referee method." Appendix 2 details specific reference methods..
B4	1	Section 3.1.1, page 14: "Organisations conducting compliance or validation testing must be recognised for the purpose by the Ministry of Health. This requires demonstration of compliance with the relevant clauses of the *General Requirements for the Competence of Testing and Calibration Laboratories* (NZS ISO/IEC 17025) (IANZ 2005)."
B5	1	Appendix 1, Section A1.2.2, page 97: "The general requirements for the competence of testing and calibration laboratories requires laboratories to calculate their uncertainty of measurement".

[16] Water Treatment Plant (WTP)
[17] Distribution Zone (DZ)
[18] Drinking Water Assessor (DWA)
[19] International Accreditation New Zealand (IANZ)

Code	Score	Justification
B6	0	Section 3.1 page 13: "Procedures for sample… reporting must be agreed beforehand with the MoH-recognised laboratory that will carry out the analysis". The standard does not define what these are.
C1	1	These intervals are detailed in Section 3.2, page 16.
C2	1	Section 3.2, point 2, page 16: "Continuous monitors used for compliance testing must be standardised at least as frequently as recommended by the equipment suppliers."
C3	1	Appendix 2, Section A2.1, page 103: "Accredited or recognised laboratories must standardise any online instrumentation used for compliance testing of water… the water supplier must record the result, together with any adjustments that are made to the instrument, and the identity of the operator(s)."
C4	0	Not present in this standard.
D1	1	Conversion calculations are detailed in Appendix 1, Section A1.3, pages 98 – 100.
D2	0	Free available chlorine (FAC) and pH are two parts of the C.t calculation and they are required to be included in a compliance report. However they are required due to their own implications for compliance, and not to support the C.t calculation. The standard does not specifically state that input data must be part of the compliance report alongside calculated data.
D3	1	Appendix A1, Section A1.1, page 97
D4	0	Not present in this standard.
D5	1	Appendix A1, Section A1.2.1, page 97: "To establish whether a transgression has occurred, the test result (measurement) must be compared with the MAV[20]." Appendix A1, Section A1.3, pages 98 - 100, states the reporting units required.
E1	0	The standard does not require electronic reporting, therefore non-electronic reporting is accepted.
E2	0	The standard does not require laboratory data to be sent directly to the regulator.
E3	0	Not present in this standard.

[20] Maximum Allowable Value (MAV)

Code	Score	Justification
E4	0	Not present in this standard.
E5	0	Not present in this standard.
E6	0	Not present in this standard.
E7	0	Not present in this standard.

Table 4.2 Data Quality Indicator Scores and Justifications for the Drinking-water Standards for New Zealand 2005 (Revised 2018)

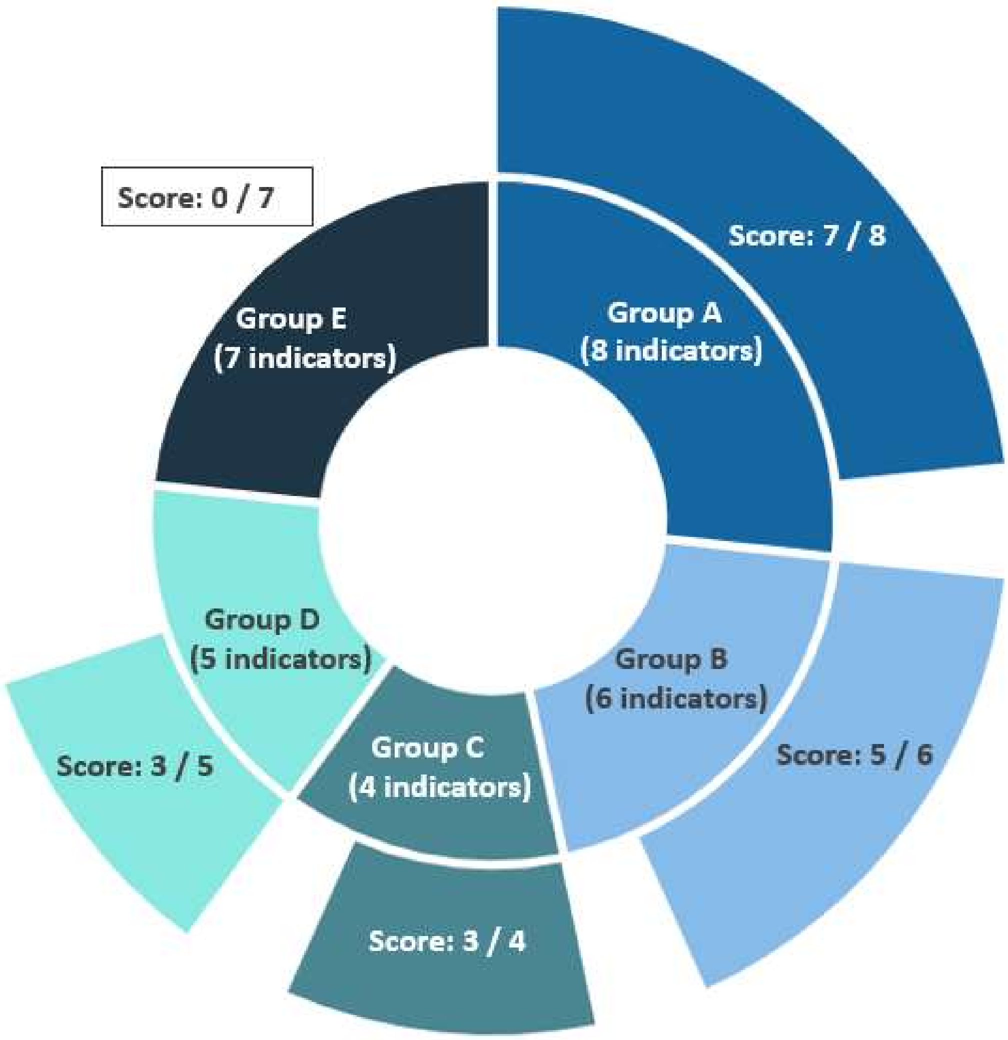


Figure 4.1 Summary graph illustrating the total number of data quality indicators possible for each group and the number of indicators that were present in the Drinking-water Standards for New Zealand 2005 (Revised 2018)

4.2.2 New Zealand – Drinking Water – Proposed New Standard

Legislation: Drinking Water Quality Assurance Rules 20 December 2021 (Draft) [46]

Code: NZ DW TA

Similar to the DWSNZ, Section 3 of the draft Drinking Water Quality Assurance Rules concerns compliance and reporting requirements. Treatment-specific rules provide a framework for suppliers to demonstrate compliance against. Unlike the DWSNZ, the determinands of interest and their maximum allowable values (MAVs) are present in a separate document, called the Drinking Water Standards for New Zealand (Draft, 20 December 2021) [47]. For the purposes of this framework application, these documents are assessed together, to assist the comparison with the existing DWSNZ.

The December 2021 draft of the NZ DW TA contains little detail on the reporting requirements of compliance data. Not all DQ indicators present in the DWSNZ are included in the new draft either; for example, the unit system used. This is likely due to the fact that the new Rules are still in the draft phase; however, it provides a useful application of the framework, wherein the differences between the two standards can be highlighted, ensuring that they are not missed when the new Rules are being compiled. Some other DQ indicators, such as the requirement to use an accredited lab for compliance sampling, are outlined in the Water Services Act 2021 [45]; where this is the case, it is noted in the Justification section.

Unless otherwise specified, the document referenced in Table 4.3 is NZ DW TA [46].

Code	Score	Justification
A1	0	No standard method is referenced or described.
A2	0	No standard method is referenced or described.
A3	1	Rule G4, page 25.
A4	1	Rule G4, page 25.
A5	1	Detailed in each rule module table. For example, Rule Section T3, Table 20, page 54 determines that turbidity is required to be monitored “on the inlet and outlet of [the] sedimentation process "
A6	0	Not present in this standard.
A7	1	Rule G5, page 25.
A8	0	Not present in this standard.
B1	1	The Water Services Act 2021 Subpart 11 Section 73 Subsection 1 requires a drinking water supplier to use an accredited laboratory “as part of any monitoring requirements in compliance rules” [45]. As with the DWSNZ, the requirements of

Code	Score	Justification
		an accredited laboratory ensure this element is represented in subsequent compliance monitoring datasets.
B2	1	As for B1.
B3	0	Not present in this standard.
B4	1	As for B1.
B5	0	Not present in this standard.
B6	0	No standard method is referenced or described.
C1	1	Rule G9, page 26.
C2	1	Rule G8, page 26.
C3	0	Not present in this standard.
C4	0	Not present in this standard.
D1	0	While calculations are referred to, no supporting appendix like that of the DWSNZ is included.
D2	1	The tables in Section T3 list the parameters that need to be monitored, and the calculations that they are used in.
D3	0	Not present in this standard.
D4	0	Not present in this standard.
D5	1	All parameters required to be monitored have units listed in the draft Drinking Water Standards [47].
E1	0	Not present in this standard.
E2	0	Not present in this standard.
E3	0	Not present in this standard.
E4	0	Not present in this standard.
E5	0	Not present in this standard.
E6	0	Not present in this standard.
E7	0	Not present in this standard.

Table 4.3 Data Quality Indicator Scores and Justifications for the Drinking Water Quality Assurance Rules 20 December 2021 (Draft)

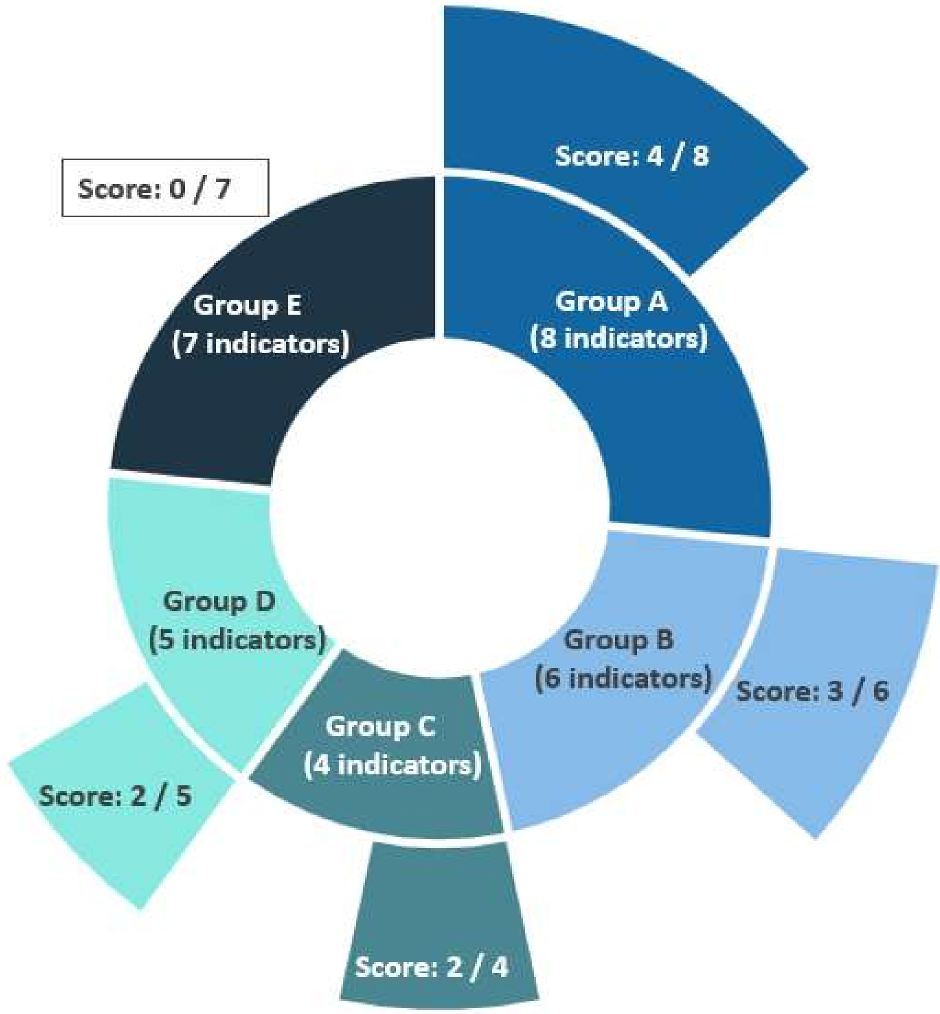


Figure 4.2 Summary graph illustrating the total number of data quality indicators possible for each group and the number of indicators that were present in the Drinking Water Quality Assurance Rules 20 December 2021 (Draft)

4.2.3 New Zealand – Water Takes

Legislation: Resource Management (Measurement and Reporting of Water Takes) Regulations 2010 [68]

Code: NZ WT

These regulations apply to water permits granted under the Resource Management Act 1991 [48]. They only apply to a water permit that allows fresh water to be taken at a rate of 5 L/s or more.

These regulations only apply to continuous monitoring, so the DQ indicators in groups A (sampling process) and B (lab analysis process) are all scored as not applicable (N/a), except for A5. The document does not specify a need for the GPS coordinates of the continuous monitoring instrument, but it does say that the instrument needs to be installed at the location of the water take. This is better represented by DQ indicator A5, so the score for C4 is zero.

While these regulations do not require instrument calibration, they do require instrument verification. This has the same impact on the DQ of the dataset – because instrument verification is required, it gives the regulator confidence that the data collected by the instrument is accurate. Therefore indicators C2 and C3 have been assessed using the verification requirements of the regulations.

Indicator D2 concerns the requirement to report calculated data with the data used in the calculation (the input data). While point 6 subpoint 5 (page 4) states “the records must be kept in a format that, in the opinion of the regional council that granted the permit, is suitable for auditing”, the regulations do not specifically require the input data to be included as part of a compliance report. The onus of maintaining auditable data is given to the supplier, but the definition of “auditable” would be determined by the regional council that granted the permit. In practice, if all the compliance datasets collected under these regulations were compiled, it is likely that there would not be one standard audit system, as each regional council could potentially have a different method. Therefore these regulations do not convey the DQ of indicator D2 to all compliance data submitted under them, and so the score for this indicator is zero.

With regards to indicator E1, these regulations require electronic data transfer (point 7A subpoint 4 page 5). However they also allow records to be sent in writing if the regional council requests it. The wording of the question for E1 would imply that this indicator would receive a score of zero, but similar to the D2 example, the regional council is the only reason that non-electronic data transfer would be accepted. In other words, the regulations require electronic data transfer, and any change is only due to the preference of the council. Therefore this indicator is given a score of 1.

Unless otherwise specified, the document referenced in Table 4.4 is NZ WT [68].

Code	Score	Justification
A1	N/a	
A2	N/a	
A3	N/a	
A4	N/a	
A5	1	See note for C4.
A6	N/a	
A7	N/a	
A8	N/a	
B1	N/a	
B2	N/a	
B3	N/a	
B4	N/a	
B5	N/a	
B6	N/a	
C1	1	Described in point 6 subpoint 2(a) page 4.
C2	1	Described in Point 7, subpoints 2 and 3, page 5.
C3	1	Point 7, subpoint 4, page 6: "verification must have been performed by a person who… is suitably qualified."
C4	0	Point 6, subpoints 6(e) (i) and (ii), page 4 give the location requirements for the monitoring instrument.
D1	1	Point 5 on page 3 provides the method for determining flow rate in L/s.
D2	0	Point 6, subpoint 5, page 4: "the records must be kept in a format that, in the opinion of the regional council that granted the permit, is suitable for auditing."
D3	0	Not present in these regulations.
D4	0	GPS data is not required, so the datum is not specified.
D5	1	This may differ for each permit, based on the requirement of the regional council that issued the permit. However point 4, subpoint 1, page 3 states that "these regulations apply only to a water permit that allows fresh water to be taken at a rate of 5 litres/second or more". In addition, the calculations in point 5 are all based around L/s.

E1	1	Point 7A, page 5 applies to permit holders who keep records under point 6, subpoint 2a, and point 7A, subpoint 4 (page 5) states that "the records must be provided electronically". However point 8 subpoint 4 page 6 also allows the records to be provided in writing if requested by the regional council. As per the discussion in the introduction to this table, this indicator is given a score of 1.
E2	N/a	Not present in these regulations.
E3	0	Not present in these regulations.
E4	0	No - point 7A, subpoint 4 (page 5) only requires "electronic reporting", which could be an electronic document such as a pdf.
E5	0	No file types are specified. However this may be due to the fact that the file type preference would be by the regional council.
E6	0	As for E5.
E7	N/a	As these regulations are not associated with a national-level regulator in the manner of the DWSNZ, this indicator is N/a.

Table 4.4 Data Quality Indicator Scores and Justifications for the Resource Management (Measurement and Reporting of Water Takes) Regulations 2010

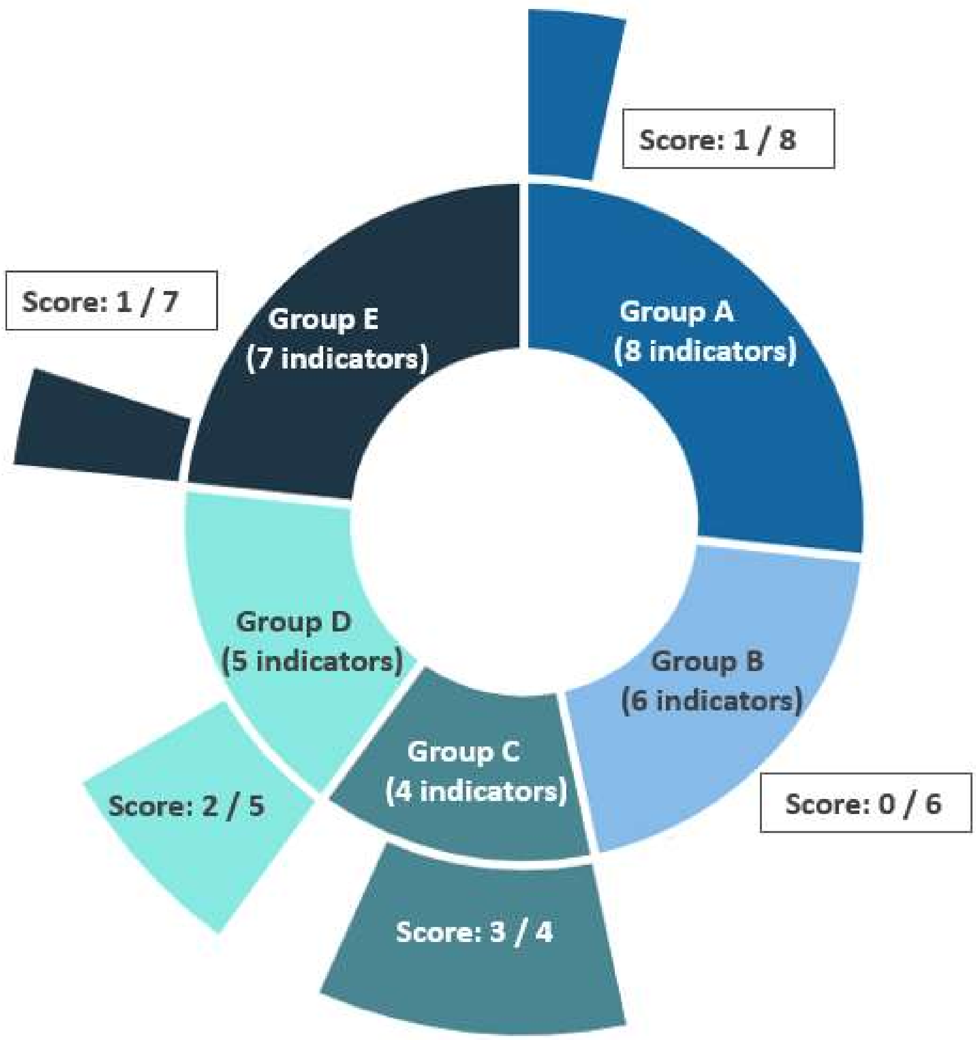


Figure 4.3 Summary graph illustrating the total number of data quality indicators possible for each group and the number of indicators that were present in the Resource Management (Measurement and Reporting of Water Takes) Regulations 2010

Secondary Legislation: Resource Management (National Environmental Standards for Sources of Human Drinking Water) Regulations 2007 [69]

Code: NES-DW

These regulations were not assessed as they do not provide compliance requirements for monitoring activities. They are written to require regional councils to consider the effects of activities on drinking-water sources when granting water permits or discharge permits or changing rules around permitted activities. Because these regulations do not have any monitoring requirements, there are no DQ indicators associated with data produced as a requirement to comply with these regulations, so the framework has not been applied.

However, if they were to be expanded to include monitoring requirements, then the framework could serve as a guideline to align the requirements with the DWSNZ and the NZ WT, so that there would be increased interoperability between these databases.

4.2.4 New Zealand – Wastewater – Christchurch WWTP

Legislation: Environment Canterbury Consent CRC051724 for Christchurch City Council [70]

Code: NZ WW Chch

Condition 11 of this resource consent states that “the volume of treated wastewater…. shall be measured and recorded continuously”, therefore the continuous monitoring group of DQ indicators were assessed.

Indicators B1 and B2 are not explicitly mentioned in the consent, however these are given a score of one due to the consent’s requirement to use an accredited laboratory (condition 28c). This is because in order to achieve the accreditation, the laboratory must ensure that technicians are appropriately qualified (indicator B2), and that instruments used for analysis are maintained according to the manufacturer’s requirements (i.e., calibrated accordingly, indicator B3). Another requirement of accredited laboratories is to include uncertainty calculations in their report. However indicator B5 asks if the consent itself requires the uncertainty to be reported. While the supplier may have the data, if it is not included with the compliance dataset, then the overall data quality of the compliance data is not prioritised. Therefore, this indicator is scored as zero.

Unless otherwise specified, the document referenced in Table 4.5 is NZ WW Chch [70].

Code	Score	Justification
A1	1	Condition 28(a) states that "all sampling required under this consent shall be undertaken by competent persons using the most appropriate scientifically recognised method(s)." While this does not refer to a specific method, it is given a positive score because the effect of the condition is to increase the data quality of the monitoring data collected.
A2	0	This is not included in this consent. While it could be argued that this is covered by the statement in condition 28(a), other standards that use similar wording to that condition also note the container requirements. This standard does not.
A3	0	As for A2.
A4	0	As for A2.
A5	1	The consent describes a number of different monitoring locations. In particular, condition 25(a) requires the following: “Samples to be collected from locations 100, 200, 500, 1,000 and 2,500 metres along shore from the centre point of the diffuser (both north and south) and 100, 200 and 500 metres inshore and offshore from the diffuser (total 16 sites, 80 samples).”

Code	Score	Justification
A6	1	As for A1, the wording requires a competent person.
A7	0	Not present in this consent, even though some of the analytical processes required in condition 13c would be field tests (e.g., BOD_5[21] and temperature).
A8	0	While the location requirements are specific, there is no mention of GPS coordinates. The sample locations in condition 25(a) would benefit from having GPS coordinates, as they would confirm that the correct distances between locations were used.
B1	1	As for A1.
B2	1	As for A1.
B3	1	Condition 28b.
B4	1	Conditions 28b and 28c.
B5	0	See notes in the introduction to this table.
B6	0	Not present in this consent.
C1	0	Condition 11 requires continuous monitoring, but does not state the allowable interval between data points.
C2	0	Condition 9 requires that "any equipment used to monitor flow rate…. shall be maintained in good working order at all times". However no reference is made to calibration or validation requirements.
C3	0	As for C2.
C4	0	Not present in this consent.
D1	N/a	No calculations are detailed in this consent.
D2	N/a	As for D2.
D3	0	Not present in this consent.
D4	0	Not present in this consent.
D5	1	Condition 13 states the units that the analytes must be reported in.
E1	0	Condition 31h states that "reporting of all data shall be in hardcopy and electronic formats".
E2	0	Not present in this consent.
E3	0	Not present in this consent.
E4	0	Not present in this consent.

[21] Biological Oxygen Demand, BOD_5

Code	Score	Justification
E5	0	Not present in this consent.
E6	0	Not present in this consent.
E7	0	Not present in this consent.

Table 4.5 Data Quality Indicator Scores and Justifications for the Environment Canterbury Consent CRC051724 for Christchurch City Council

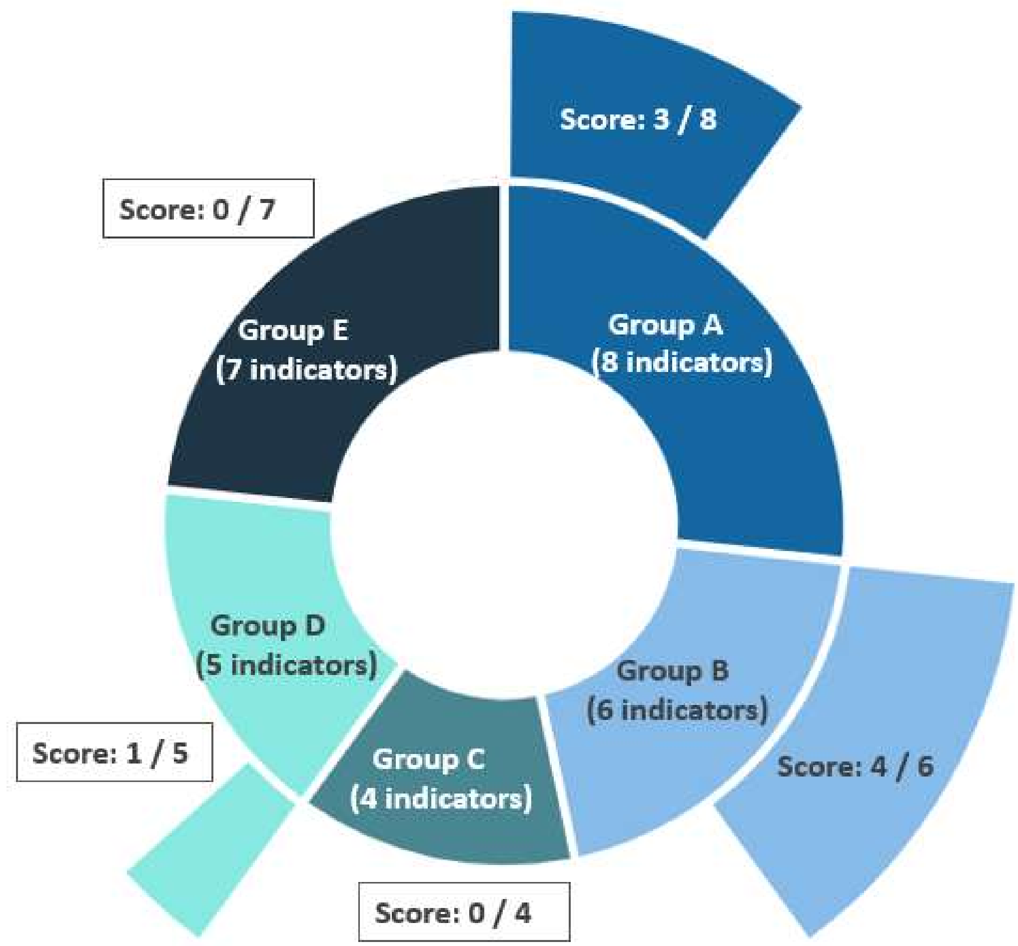


Figure 4.4 Summary graph illustrating the total number of data quality indicators possible for each group and the number of indicators that were present in the Environment Canterbury Consent CRC051724 for Christchurch City Council

4.2.5 New Zealand – Wastewater – Ashburton WWTP

Legislation: Environment Canterbury Consent CRC030999.1 for Ashburton District Council [71]

Code: NZ WW Ash

For this resource consent, indicator A1 is scored as a 1 due to the statement "using the most appropriate scientifically recognized and current methods". It does not refer to a specific method due to the time frame associated with the consent. This consent was granted in 2004 and expires in 2039. If a particular method was specified in 2004, but changes to technology resulted in a new industry standard in the subsequent 35 years for which the consent is valid, then this would negatively impact the quality of the data captured for compliance with the consent. By including this statement, the consent prioritises the quality of the monitoring method, and so contributes to the overall data quality of the data set.

No continuous monitoring is required under this consent, so the indicators in Group C are all scored as N/a.

Indicator D4 is scored as a yes because the consent specifies the datum that it is working in. While the consent does not require the GPS coordinates of the sample locations (indicator A8), it does specify the required sample locations (indicator A5).

Unless otherwise specified, the document referenced in Table 4.6 is NZ WW Ash [71].

Code	Score	Justification
A1	1	Condition 24a states that "all sampling required under this consent shall be undertaken by a competent person using the most appropriate scientifically recognised and current methods". Note, conditions 12 and 13 give method guidance for bore samples as well.
A2	0	Not present in this consent.
A3	0	Not present in this consent.
A4	0	Not present in this consent.
A5	1	Locations are specified when monitoring is required. An example is seen in condition 6(i).
A6	1	As for A1.
A7	0	Not present in this consent.
A8	0	Not present in this consent.
B1	1	This is covered under the requirements of conditions 24b and 24c to use an accredited laboratory.

Code	Score	Justification
B2	1	As for B1.
B3	1	Condition 24b.
B4	1	Conditions 24a and 24b.
B5	0	As for B5 in table 4.5 (NZ WW Christchurch).
B6	0	Not present in this consent.
C1	N/a	
C2	N/a	
C3	N/a	
C4	N/a	
D1	N/a	No calculations are required by this consent.
D2	N/a	
D3	0	Not present in this consent.
D4	1	Condition 2 states that the map reference is NZMS 260 (New Zealand Map Series 260).
D5	0	Condition 3 says the volume is measured in m3/day, but all of the parameters to be analysed in condition 6 do not have any units specified.
E1	0	Condition 28 only says that the consent holder needs to provide a report. No required format is given.
E2	0	Not present in this consent.
E3	0	Not present in this consent.
E4	0	Not present in this consent.
E5	0	Not present in this consent.
E6	0	Not present in this consent.
E7	0	Not present in this consent.

Table 4.6 Data Quality Indicator Scores and Justifications for the Environment Canterbury Consent CRC030999.1 for Ashburton District Council

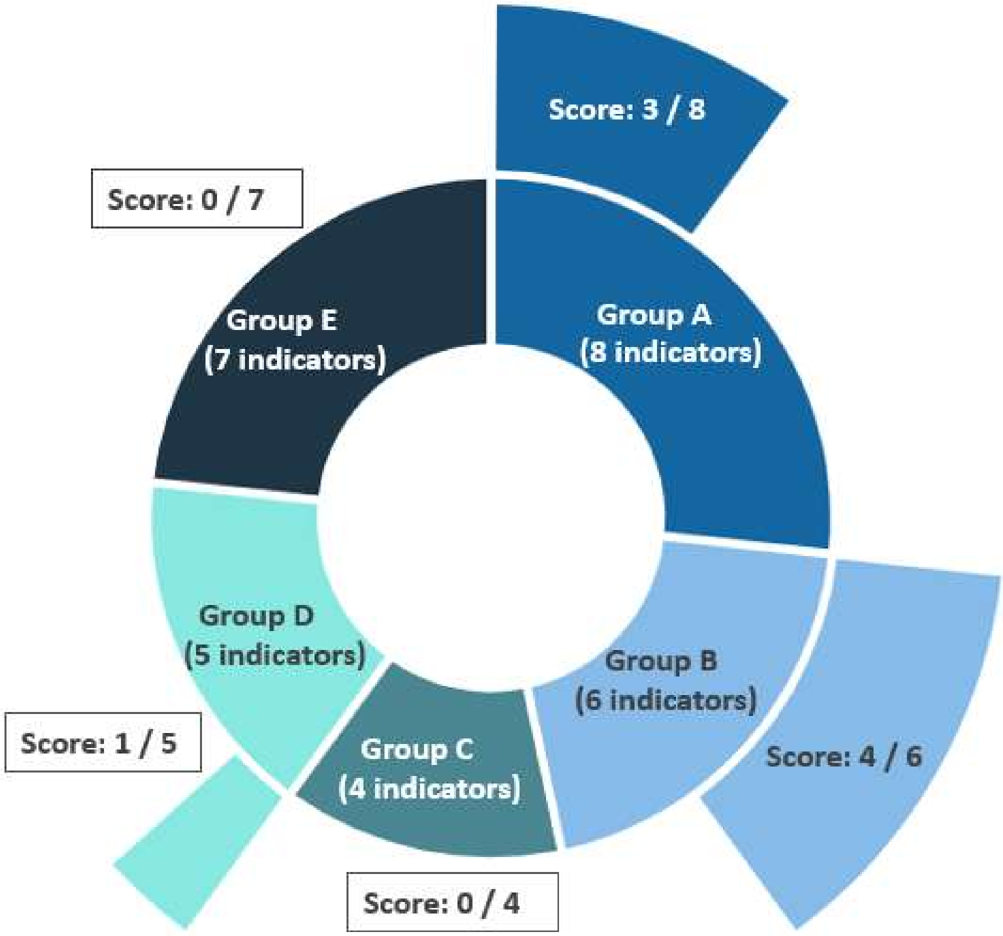


Figure 4.5 Summary graph illustrating the total number of data quality indicators possible for each group and the number of indicators that were present in the Environment Canterbury Consent CRC030999.1 for Ashburton District Council

4.2.6 New Zealand – Carbon

Legislation: Climate Change (Stationary Energy and Industrial Processes) Regulations 2009 [67]

Code: NZ Carbon

In New Zealand, the industry body Water NZ published a set of guidelines to assist councils in assessing greenhouse gas (GHG) emissions of wastewater treatment plants [52]. While GHG emission regulations for a number of industry types exist at a national level in New Zealand, they do not specifically apply to wastewater treatment processes. Therefore the Climate Change (Stationary Energy and Industrial Processes) Regulations 2009 have been reviewed to determine what the data quality of a GHG emission regulation for New Zealand is like, and what regulations specific to wastewater GHG emission calculations may look like if they are developed in the future. These regulations were made pursuant to section 163 of the Climate Change Response Act 2002 (CCA) [72].

These regulations contain requirements for a number of different industries; aside from the general conditions, the industry-specific conditions that will be assessed are 21 to 26C as these apply to municipal waste, which of the industries covered in the regulations is the most closely related process to wastewater treatment. The DQ indicators of these regulations can then be compared with those of the wastewater consents in sections 4.2.4 and 4.2.5, as well as indicators from standards in Australia and the European Union.

Monitoring methods for greenhouse gas emissions differ to those for water samples in that water samples can be transported back to a laboratory for analysis, while GHG emission monitoring is completed on site. This means that indicators A2, A3, A4, and B1 are not applicable. However some Group B indicators are still applicable, because the on-site monitoring is required to be conducted by an accredited laboratory. Also, indicator A1 is not the same as B2; A1 concerns the method by which the sample is taken, and B2 the method by which it is analysed – even though this is done on site.

These regulations do not detail the sampling and analysis methodology requirements; instead, Condition 24 (1) (a) (ii) on page 34 details the reference test methods. Of these, USEPA Method 3A [73] has been considered as part of the assessment of these regulations. Section 8.0 of Method 3A ([73], page 3) discusses sample collection requirements. This section refers the reader to section 8.1 of Method 7E [74]. As such, indicators that were assessed based on the requirements in Method 7E have this noted in their justifications.

Indicator D2 is a positive, because the CCA requires the input data to be kept as a record ([72], Clause 62(d), page 153). This means that the calculation data is more easily auditable, and therefore the overall DQ is of a higher level.

Unless otherwise specified, the document referenced in Table 4.7 is NZ Carbon [67].

Code	Score	Justification
A1	1	Method 7E: Section 8, page 7-13 [74].
A2	N/a	The analysis is done on site, not at a testing laboratory.
A3	N/a	As for A2
A4	N/a	As for A2
A5	1	Method 7E: Section 8.1, page 8 [74].
A6	1	Specified in Condition 24 (1) (a) (i) on page 34.
A7	1	Method 7E: Section 9, pages 14-16, 'Summary Table of QA/QC' [74].
A8	0	Not present in these regulations.
B1	N/a	As for A2
B2	1	Method 7E: Section 9, pages 14-16, 'Summary Table of QA/QC', and Section 10, page 17 [74].
B3	1	As for A1
B4	1	Specified in Condition 24 (1) (a) (ii) on page 34.
B5	1	Method 7E, Section 12, pages 17-20 [74] gives calculation instructions for how to calculate calibration error, which (because this is all field sampling) is the uncertainty reading. However the Regulations do not require that this be reported. But as the data is captured as part of the sampling process, and with the positive score for D2, these indicators contribute to an increased level of data quality, so the score is 1.
B6	0	Not present in these regulations.
C1	1	Condition 24 (1) (a), page 34.
C2	1	Method 7E, Section 9, 'Summary Table of QA/QC', page 15 [74]: the table entries regarding System Performance specify that system calibration needs to be done before and after each run, and is categorised as mandatory.
C3	1	Method 7E, Section 10 (2), page 17 [74] specifies that the measurement system calibration report must include the name of the person or manufacturer who carried out the calibration.
C4	0	Not present in these regulations.
D1	1	Calculation formulae are described in Condition 24 (1) (c) on page 35.
D2	1	Condition 24 (b) requires that gas pressure and temperature are recorded. These data points are used in the following calculations, and so can be considered 'input data'. While the regulations do not specify that the input data need to be reported, Clause 62(d) (page 153) of the CCA [72] requires that

Code	Score	Justification
		records of the data and calculations be kept. This increases the data quality of the calculated data, so this score is a 1.
D3	0	Not present in these regulations.
D4	0	Not present in these regulations.
D5	1	Conditions 24 (b)-(f), pages 34 – 36, specify the reporting units for the calculations.
E1	0	Not present in these regulations.
E2	0	Not present in these regulations.
E3	0	Not present in these regulations.
E4	0	Not present in these regulations.
E5	0	Not present in these regulations.
E6	0	Not present in these regulations.
E7	0	Not present in these regulations.

Table 4.7 Data Quality Indicator Scores and Justifications for the Climate Change (Stationary Energy and Industrial Processes) Regulations 2009

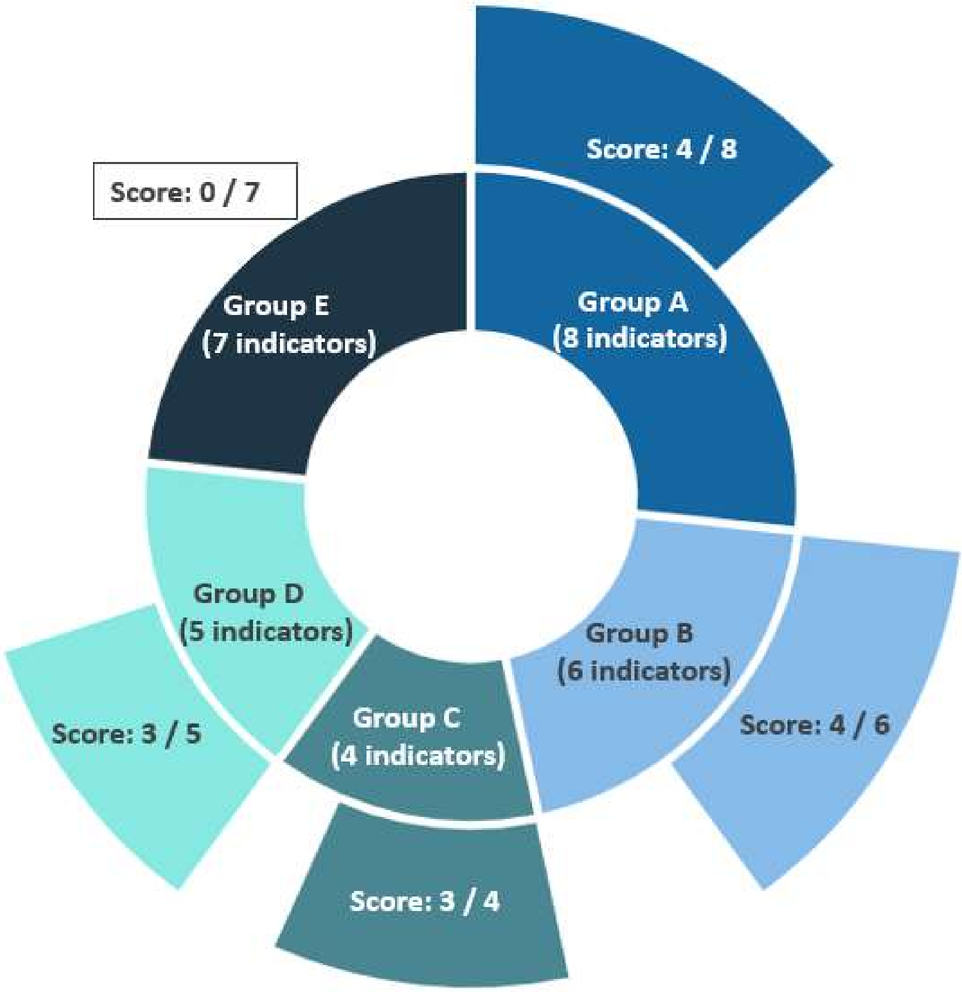


Figure 4.6 Summary graph illustrating the total number of data quality indicators possible for each group and the number of indicators that were present in the Climate Change (Stationary Energy and Industrial Processes) Regulations 2009

4.2.7 Australia – NSW – Drinking Water – NSW Health

Legislation: Drinking Water Monitoring Program December 2005 (updated October 2011) [40]

Code: AU NSW DW

Similar to the NZ Carbon regulations assessed in section 4.2.6, the AU NSW DW program also references supporting documents. The AU NSW DW program supports monitoring programs for New South Wales water utilities by providing free testing at NSW Health laboratories ([40], page 1). These laboratories, including the Division of Analytical Laboratories (DAL), are accredited by NATA (the National Association of Testing Authorities, [40], page 13). The Guide for Submitting Water Samples to DAL for Analysis (DAL Guide) [75] is a reference document to ensure samples are correctly collected and transported prior to analysis at the DAL. The DAL Guide also contains a full list of analyses that are available ([40], page 6). As such, the DAL Guide is considered alongside the AU NSW DW for this assessment. In addition, the monitoring program refers to General Accreditation Criteria published by NATA; where applicable, these Criteria have been assessed for clarification of the data quality requirements. The two Criteria considered are the 'Equipment assurance, in-house calibration and equipment verification' [76] and the 'Proficiency Testing Policy' [77].

As the AU NSW DW only considers samples taken in the field and sent to a laboratory for analysis, and does not include continuous monitoring sampling, the indicators in Group C are all scored as not applicable.

Similarly, indicators D1 and D2 are also scored as N/a as no calculations are required by the monitoring program.

Unless otherwise specified, the document referenced in Table 4.8 is AU NSW DW [40].

Code	Score	Justification
A1	1	DAL Guide Section 1.2 pages 5-7 [75].
A2	1	DAL Guide Section 1.1 page 4 [75].
A3	1	DAL Guide Section 1.3 pages 7-8 [75].
A4	1	DAL Guide Section 1.3 pages 7-8 [75].
A5	1	Section 5 pages 10-12
A6	1	DAL Guide Section 1.2 page 5: "Samples should be collected by a person trained in correct sampling techniques" [75].
A7	0	Not present in this program or the DAL Guide.
A8	0	Not present in this program or the DAL Guide.

Code	Score	Justification
B1	1	While this is not specified by the AU NSW DW, an accredited laboratory is required (Section 6.2, page 13). As qualified technicians are a requirement of laboratory accreditation (NATA Criteria, [77]), this is a positive score.
B2	1	While this is not specified by the AU NSW DW, an accredited laboratory is required (Section 6.2, page13). As instrument calibrations are a requirement of laboratory accreditation (NATA Criteria, [76]) this is a positive score.
B3	1	Only accredited laboratories can perform the tests (Section 6.2, page 13). In addition, the DAL Guide Section 6.1 ([75], page 32) notes that the *E. coli* and total coliform analysis is by Colilert methodology.
B4	1	Section 6.1 (page 13) lists the accredited testing laboratories. Section 6.2.1 (page 13) states that if a different laboratory is used, it must hold NATA accreditation for the analyses.
B5	0	Not present in this program or the DAL Guide.
B6	0	Section 6.2.5, page 14: “Laboratories should enter all results into the NSW Drinking Water Database. If the laboratory does not have internet access, then the water utility should enter the results.” As provision is given for data transfer from the utility to the supplier, this is a score of zero.
C1	N/a	
C2	N/a	
C3	N/a	
C4	N/a	
D1	N/a	
D2	N/a	
D3	0	The unit system is not defined in the program or the Guide.
D4	0	Not present in this program or the DAL Guide.
D5	1	DAL Guide Section 5.2 includes the units for the analytes [75].
E1	0	As for B6.
E2	1	Even though indicators B6 and E1 are zero, this indicator receives a positive score, because the AU NSW DW supports the transfer of testing data directly from the testing laboratory to the regulator.
E3	1	As for E2, the design of the barcode system in the AU NSW DW (Section 7.1, page 15) facilitates M2M data transfer, if the databases are appropriately configured.

Code	Score	Justification
E4	0	M2M transfer is not required. Other methods are allowed (as per indicators B6 and E1).
E5	1	Data are entered into the NSW Drinking Water database (Section 6.2.5 page 14), therefore the regulator does not receive files.
E6	0	Not present in this program or the DAL Guide.
E7	0	Not present in this program or the DAL Guide.

Table 4.8 Data Quality Indicator Scores and Justifications for the Drinking Water Monitoring Program December 2005 (updated October 2011)

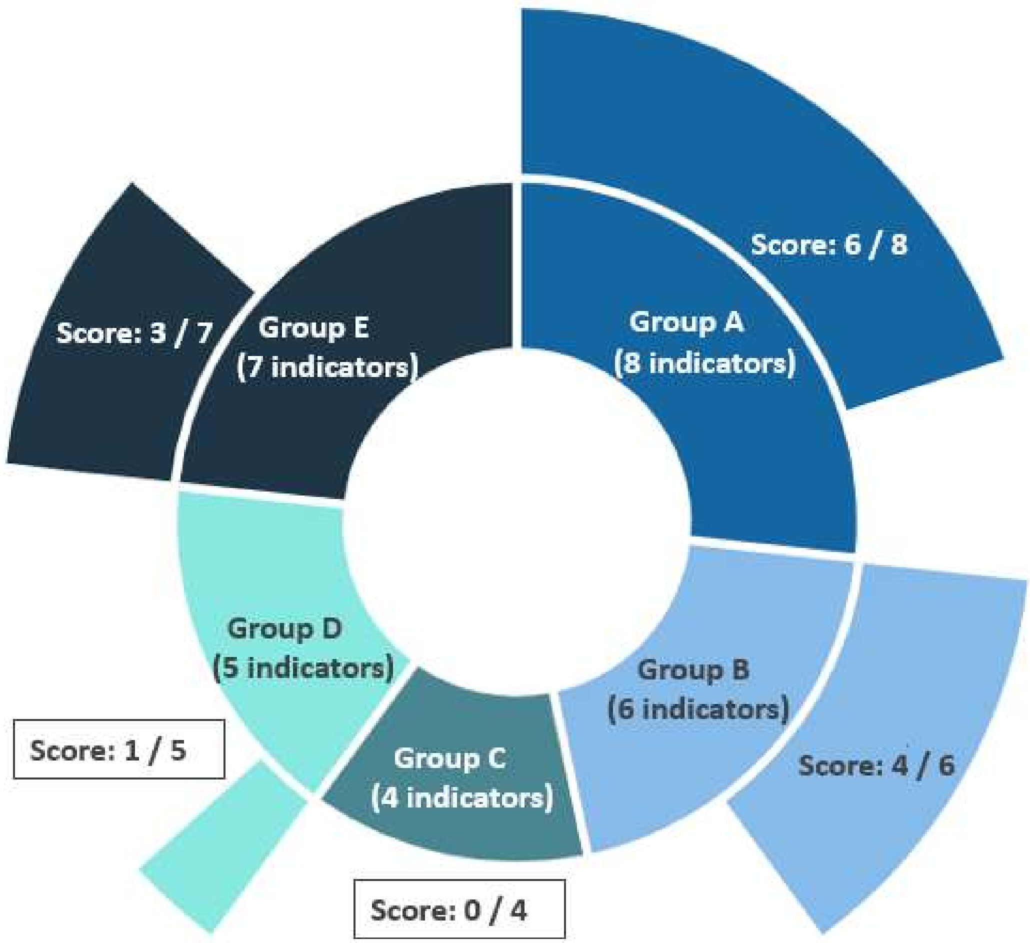


Figure 4.7 Summary graph illustrating the total number of data quality indicators possible for each group and the number of indicators that were present in the Drinking Water Monitoring Program December 2005 (updated October 2011)

4.2.8 Australia – Queensland – Environmental Waters

Legislation: Monitoring and Sampling Manual 2018: Environmental Protection (Water) Policy 2009 [41]

Code: AU QLD EnvW

While this Policy is not a set of regulations requiring compliance, its purpose is described as follows: "When monitoring is required under legislation to be done under a protocol, including the Environmental Protection (Water) Policy 2009 and the Environmental Protection Regulation 2008, the Monitoring and Sampling Manual 2018 is the primary document to decide the protocols." ([41], page 2).

Similar to the AU NWS DW, this policy recommends that NATA accredited laboratories be used ([41], section 9.3, page 41). As with the assessment in section 4.2.7, as qualified technicians and instrument calibrations are requirements of laboratory accreditation (NATA Criteria, [77] and [76]), indicators B1 and B2 are both given a positive score.

The AU QLD EnvW is also concerned with field sampling only, and does not cover continuous monitoring requirements. Therefore Group C indicators are all scored as not applicable.

A notable inclusion of this policy is Section 1.4 on page 259, which states that "all data, whether generated by government or by organisations external to government, must be managed by a data custodian." This puts the burden of ensuring data quality on the Data Custodian. Section 1.5 on page 260 further details good data management practices.

Unless otherwise specified, the document referenced in Table 4.9 is AU QLD EnvW [41].

Code	Score	Justification
A1	1	This document extensively details sampling methodology for a range of analyses.
A2	1	Section 2.6.6.1 (page 11) addresses sample container requirements.
A3	1	Section 2.6.6.2 (page 12) addresses sample storage and holding time requirements. Section 2.6.7 (pages 12-13) discusses transportation requirements.
A4	1	Section 2.6.6.2 page 12 states that "the samples need to be delivered before the maximum holding times, with enough time for the laboratory to analyse the samples".
A5	1	Section 1.8 (page 6) discusses sample location considerations; Section 6 (pages 24-26) discusses control and reference sites.

Code	Score	Justification
A6	1	Section 2.5 page 9 states that staff should “have the necessary skills to undertake the methods that will be used in the sampling program." In addition, the method sections each have a paragraph that addresses the skillset required.
A7	1	Calibration requirements are described in Section 2.7.1.2, page 49.
A8	1	Section 3.7.2.5 (page 17) states that the GPS location must be recorded.
B1	1	Section 9.3 (page 41) recommends NATA accredited laboratories be used. As with B1 in AU NSW DW, this conveys a positive score.
B2	1	Section 9.3 (page 41) recommends NATA accredited laboratories be used. As with B2 in AU NSW DW, this conveys a positive score.
B3	1	Section 9.4 page 41 states that analytical methods should be accredited or equivalent to NATA.
B4	1	Section 9.3, page 41 states that analytical laboratories should: "demonstrate Good Laboratory Practice and perform according to the standard AS ISO/IEC 17025-2005 (R2016): General requirements for the competence of testing and calibration laboratories. This is usually demonstrated by labs holding accreditation with the National Association of Testing Authorities (NATA)."
B5	1	Section 9.5 page 42: states that “the Limits of Reporting (LOR) for an analytical method should be lower than the benchmark… to which results will be compared". In order to confirm the LOR was acceptable, it will have to be part of a report.
B6	0	Not present in this policy.
C1	N/a	
C2	N/a	
C3	N/a	
C4	N/a	
D1	1	Conversion factors are listed in section 2.2 on page 262. It is noted that “it is important to compare like with like and convert the results if needed.”
D2	0	While D1 is positive, the policy does not include a requirement for the input data for the conversions to be included in the report.
D3	0	Not present in this policy.
D4	1	Section 8.7.1.1, Table 4, page 38 says the recommended datum is GDA 94. Section 7.3 (page 33) provides information on how to work with other datum

Code	Score	Justification
		systems. Section 2.6 (page 261) requires that the data submitted must include the GPS datum used.
D5	N/a	Because this is a guide for sampling that may be undertaken in response to different situations (e.g., compliance or prosecution), there is no list of MAVs to refer to.
E1	1	Section 1.6 page 261 states that "the data must be supplied to the Queensland Government in the specified electronic file format and/or template where requested".
E2	0	Not present in this policy.
E3	0	Not present in this policy.
E4	0	Not present in this policy.
E5	1	Section 2.6 page 261 states that "acceptable file formats may include excel, text or comma separated value files". PDFs are described as not being an acceptable file format.
E6	1	Section 1.6 page 260 states that "any water monitoring data provided to the Queensland Government should be collected in accordance with this document and other relevant standards, guidelines, and policies."
E7	1	Section 1.6 page 261 states that "the Queensland Government also reserves the right to use monitoring data that has been provided to it by any organisation for any purpose it sees fit including supply of data to a third party".

Table 4.9 Data Quality Indicator Scores and Justifications for the Monitoring and Sampling Manual 2018: Environmental Protection (Water) Policy 2009

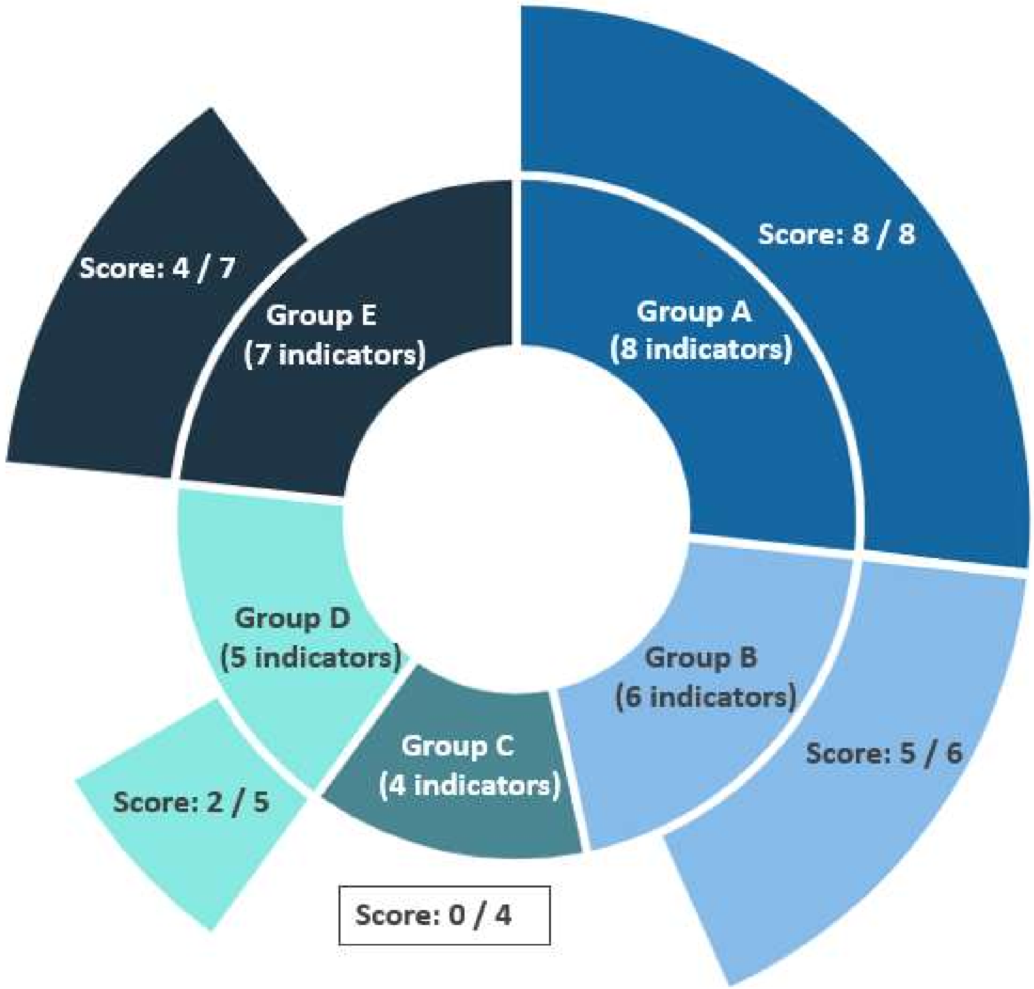


Figure 4.8 Summary graph illustrating the total number of data quality indicators possible for each group and the number of indicators that were present in the Monitoring and Sampling Manual 2018: Environmental Protection (Water) Policy 2009

4.2.9 Australia – Carbon – Wastewater

Legislation: Carbon Credits (Carbon Farming Initiative – Domestic, Commercial and Industrial Wastewater) Methodology Determination 2015 [53]

Code: AU Carbon

While in New Zealand, regulations specific to wastewater treatment plant greenhouse gas (GHG) emission calculations have not yet been determined, in Australia a methodology covers this. By assessing this methodology, the data quality requirements of GHG emission monitoring data used in a carbon credit trading scheme can be evaluated.

Similar to the NZ Carbon regulations, the AU Carbon methodology refers to a supplementary document that contains further details on the method requirements. This document is the National Greenhouse and Energy Reporting (Measurement) Determination 2008 (NGER) [78].

Unless otherwise specified, the document referenced in Table 4.10 is AU Carbon [53].

Code	Score	Justification
A1	1	The measurement procedure for each parameter is included in the table in Division 2, Section 45 (pages 26 – 35).
A2	1	As for A1; if the parameter requires laboratory analysis, the method is referenced.
A3	1	As for A2.
A4	1	As for A2.
A5	1	As for A1, the sample locations are included in this table.
A6	1	As for A2.
A7	1	As for A2.
A8	0	Not mentioned in this methodology.
B1	1	As for A2.
B2	1	As for A2.
B3	1	As for A2.
B4	1	As for A2.
B5	1	NGER 2008 Chapter 8 is on the assessment of uncertainty [78].
B6	0	Not mentioned in this methodology.

Code	Score	Justification
C1	1	As for A1, if continuous monitoring is possible for that parameter, it is noted in this table.
C2	1	Note 2 on page 39 states that "any equipment or device used to monitor a parameter must be calibrated by an accredited third party technician at intervals, and using methods, that are in accordance with the manufacturer's specifications".
C3	1	As for C2.
C4	0	Not mentioned in this methodology.
D1	1	As for A1, each parameter in the table has this noted.
D2	1	The table in Division 2, Section 45 (pages 26 – 35) separates the measurement procedures (input data) from the determination of the parameter (output data).
D3	1	NGER 2008, section 1.8 (Definitions), page 4, states that the "Australian legal unit of measurement has the meaning given by the National Measurement Act 1960" [78].
D4	0	Not mentioned in this methodology.
D5	1	As for A1, each parameter in the table has this noted.
E1	0	Not mentioned in this methodology.
E2	0	Not mentioned in this methodology.
E3	0	Not mentioned in this methodology.
E4	0	Not mentioned in this methodology.
E5	0	Not mentioned in this methodology.
E6	0	Not mentioned in this methodology.
E7	0	Not mentioned in this methodology.

Table 4.10 Data Quality Indicator Scores and Justifications for the Carbon Credits (Carbon Farming Initiative – Domestic, Commercial and Industrial Wastewater) Methodology Determination 2015

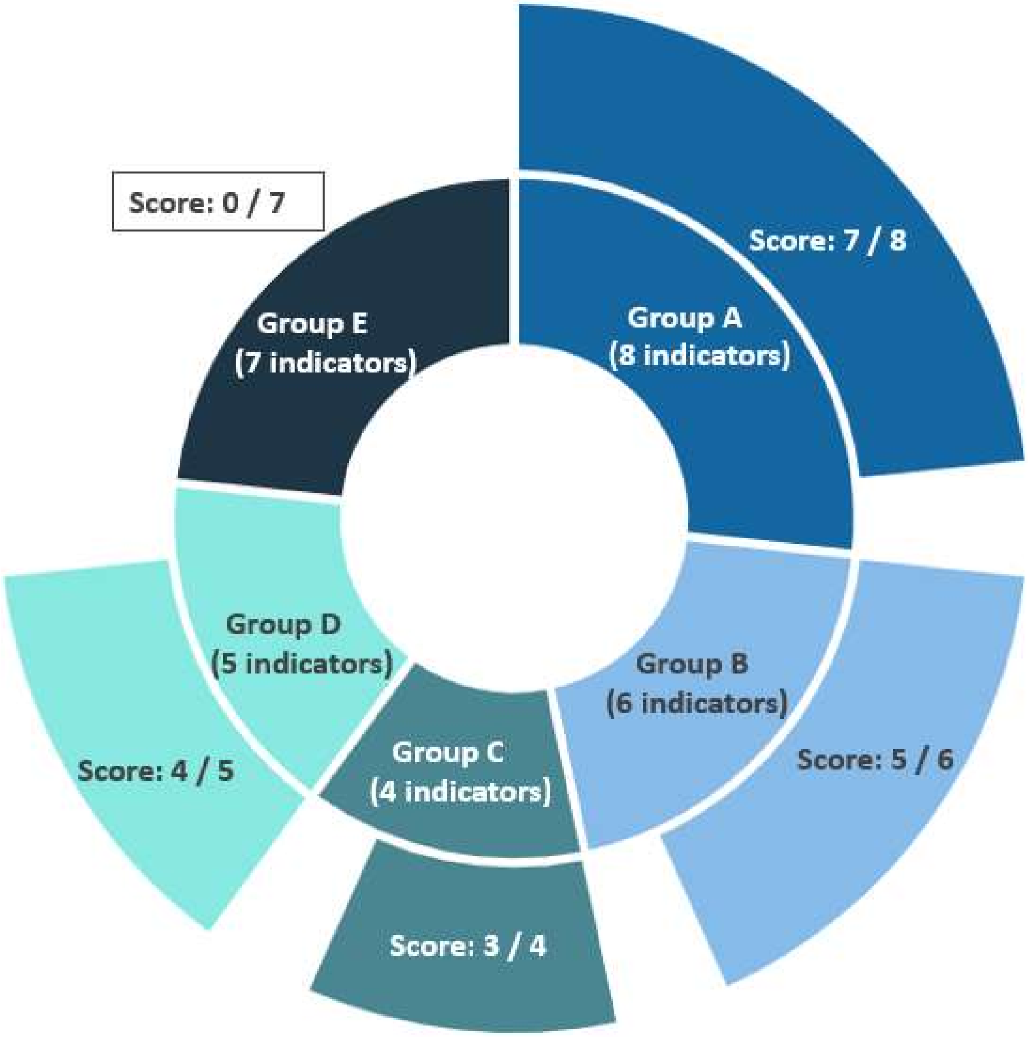


Figure 4.9 Summary graph illustrating the total number of data quality indicators possible for each group and the number of indicators that were present in the Carbon Credits (Carbon Farming Initiative – Domestic, Commercial and Industrial Wastewater) Methodology Determination 2015

4.2.10 EU – Drinking Water

Legislation: Directive (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020 on the quality of water intended for human consumption [49]

Code: EU DW

Unlike the regulations and standards reviewed for New Zealand and Australia, this Directive provides instruction to the Member States of the EU as to what needs to be included in country-level drinking water standards. Article 4(b) (page 15) of this Directive requires drinking water to meet “the minimum requirements set out in Parts A, B and D of Annex I”. This means that the EU DW determines the compliance requirements for drinking water monitoring, and the Member States are then required to create legislation to ensure these are met. Therefore, by evaluating the EU DW, the expected compliance requirements of the Member States can be assessed. As such, the application of this framework to the EU DW could serve as a guide for any Member States wishing to update their drinking water standards.

As with the NZ Carbon standard, this Directive also refers to supplementary documents. Of note are Directive 2007/2/EC which establishes an infrastructure for spatial information in the EU [79], and EU 2003/4/EC, which concerns public access to environmental information [80]. Where pertinent, these Directives have been referenced in the DQ indicator justifications. Group E in particular is affected by these supplementary directives, due to EU DW Point 43 (page 11), which states that: “to ensure that the appropriate infrastructure for public access, reporting and data-sharing between public authorities exists, Member States should base the data specifications on Directive 2007/2/EC and its implementing acts”, one of which is 2003/4/EC.

Unless otherwise specified, the document referenced in Table 4.11 is EU DW [49].

Code	Score	Justification
A1	1	Annex II Part D points 2-4 (page 48) state that sampling shall be in accordance with ISO standards.
A2	1	As for A1.
A3	1	As for A1.
A4	1	As for A1.
A5	1	Annex II Part D (page 48) states that "Sampling points shall be determined so as to ensure compliance with Article 6(1)". Article 6(1) outlines the point of compliance, i.e., where samples must be taken from.
A6	1	As for A1.

Code	Score	Justification
A7	1	As for A1.
A8	1	Article 18 (1) (page 30) states that "where possible, spatial data services as defined in point (4) of Article 3 of Directive 2007/2/EC shall be used to present the data sets referred to."
B1	1	As for B3.
B2	1	As for B3.
B3	1	Annex III (page 49) states that "Member states shall ensure that the methods of analysis used for the purposes of monitoring and demonstrating compliance with this Directive… are validated and documented in accordance with EN ISO/IEC 17025". In addition, Annex III Part A (page) 50 details specific methods.
B4	1	Annex III (page 49) states that "Member states shall ensure that labs or parties contracted by labs apply quality management system practices in accordance with EN ISO/IEC 17025".
B5	1	Annex III part B Table 1 (page 51) details the uncertainty of measurement for each parameter. As with indicator B5 for AU QLD EnvW, this is scored as a 1.
B6	0	Not present in this Directive.
C1	0	Because turbidity is required to be monitored continuously (Annex II part A points 2(b) and 3, page 43), this section is not all N/a. However, the standard does not include any further requirements for continuous monitoring - see also Annex III (page 49) where the analysis methods are referenced excepting turbidity.
C2	0	Not present in this Directive.
C3	0	Not present in this Directive.
C4	0	Not present in this Directive.
D1	1	Annex I (pages 34 – 42) includes calculation methods for relevant parameters.
D2	1	For certain parameters in Annex I (pages 34 – 42), the input data is required to be reported alongside the calculated data.
D3	0	Not present in this Directive.
D4	1	Article 18 (1) (page 30) requires datasets to be presented using spatial data services defined in 2007/2/EC. Article 8 (b) (page 6) of Directive 2007/2/EC requires “a common framework for the unique identification of spatial objects, to which identifiers under national systems can be mapped in order to ensure interoperability between them”.

Code	Score	Justification
D5	1	Annex I (pages 34 – 42) details parameters with their reporting units.
E1	1	As EU DW Point 43 (page 11) suggests, this is based on Directive 2007/2/EC and its implementing act 2003/4/EC. Of 2003/4/EC, Article 7(1) (page 5) requires Member States to ensure that “environmental information progressively becomes available in electronic databases” [80].
E2	0	Not present in this Directive.
E3	0	Not present in this Directive.
E4	0	Not present in this Directive.
E5	1	As for E1, the overarching requirement of the three Directives is for the data to be interoperable. For this to be achieved, the file formats must be specified. This indicator is scored as 1, because the intention of the Directives is to encompass this DQ indicator.
E6	1	The standard referenced is EU 2007/2/EC [79].
E7	1	EU 2003/4/EC Article 8 (1) (page 6) requires that “Member States shall, so far as is within their power, ensure that any information that is compiled by them or on their behalf is up to date, accurate and comparable” [80].

Table 4.11 Data Quality Indicator Scores and Justifications for Directive (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020 on the quality of water intended for human consumption

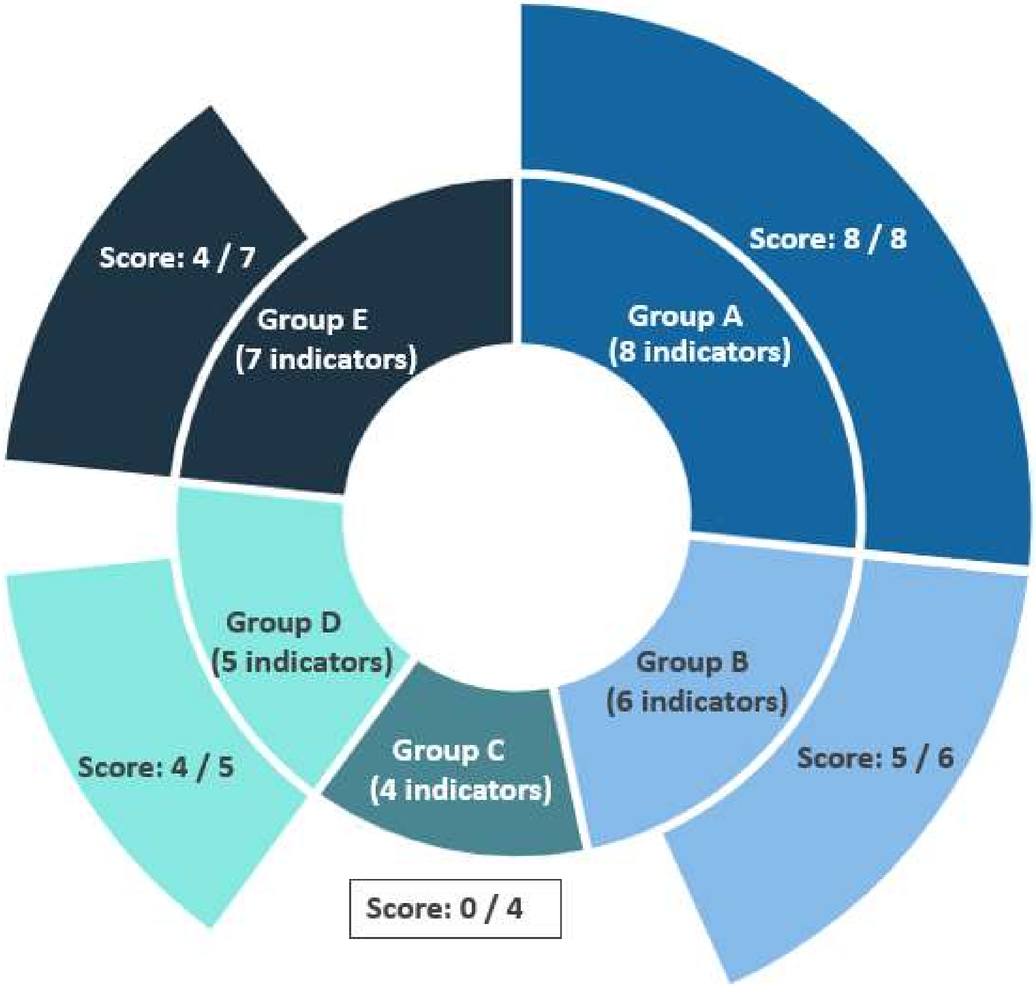


Figure 4.10 Summary graph illustrating the total number of data quality indicators possible for each group and the number of indicators that were present in Directive (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020 on the quality of water intended for human consumption

4.2.11 EU – Wastewater

Legislation: Council Directive of 21 May 1991 concerning urban wastewater treatment (91/271/EEC) [51]

Code: EU WW

Article 3, Point 2, Amendment M3 ([51], page 4) states that the "collecting systems described in paragraph 1 shall satisfy the requirements of section A of Annex I." Article 4, Point 3, Amendment M3 ([51], page 5) states that "discharges from urban waste water treatment plants… shall satisfy the relevant requirements of section B of Annex I." Therefore, as with the EU DW Directive, this Directive provides guidance to Member States on the monitoring requirements of wastewater treatment plants. Article 15, Point 1 ([51], page 10) states that wastewater treatment plant discharges shall be monitored "to verify compliance with the requirements of Annex I B", determining that this Directive sets out the compliance requirements for wastewater monitoring, and therefore can be compared with the Resource Consents assessed in sections 4.2.4 and 4.2.5.

As with the AU NSW DW, this Directive only considers samples taken in the field and sent to a laboratory for analysis and does not include continuous monitoring sampling; as such, the indicators in Group C are all scored as not applicable.

Unless otherwise specified, the document referenced in Table 4.12 is EU WW [51].

Code	Score	Justification
A1	1	Annex I, Section D, Point 2 (page 13) states that samples should be flow-proportional or time-based 24 hour samples.
A2	1	Annex I, Section D, Point 2 (page 13) states that "good laboratory practices aiming at minimising the degradation of samples between collection and analysis shall be applied".
A3	1	As for A2.
A4	1	As for A2.
A5	1	Annex I, Section B, Point 1 (page 12) states that "representative samples of the incoming waste water and of treated effluent can be obtained before discharge to receiving waters". This explains that the sample monitoring locations are at the inlets and outlets of the treatment plants.
A6	0	Not present in this Directive.
A7	0	Not present in this Directive.
A8	0	Not present in this Directive.

Code	Score	Justification
B1	0	Not present in this Directive.
B2	0	Not present in this Directive.
B3	1	Annex I Table 1 (page 14) gives the reference methods of measurement for each parameter.
B4	0	Not present in this Directive.
B5	0	Not present in this Directive.
B6	0	Not present in this Directive.
C1	N/a	
C2	N/a	
C3	N/a	
C4	N/a	
D1	1	Annex I Table 2 (page 15): Note 2 gives the calculation for total nitrogen.
D2	0	Not present in this Directive.
D3	0	Not present in this Directive.
D4	0	Not present in this Directive.
D5	1	Annex I Tables 1 and 2 (pages 14-15) specify the units for the parameters being measured.
E1	0	Not present in this Directive.
E2	0	Not present in this Directive.
E3	0	Not present in this Directive.
E4	0	Not present in this Directive.
E5	0	Not present in this Directive.
E6	0	Not present in this Directive.
E7	0	Not present in this Directive.

Table 4.12 Data Quality Indicator Scores and Justifications for Council Directive of 21 May 1991 concerning urban wastewater treatment (91/271/EEC), amended in 1998, 2003, 2008, and 2013

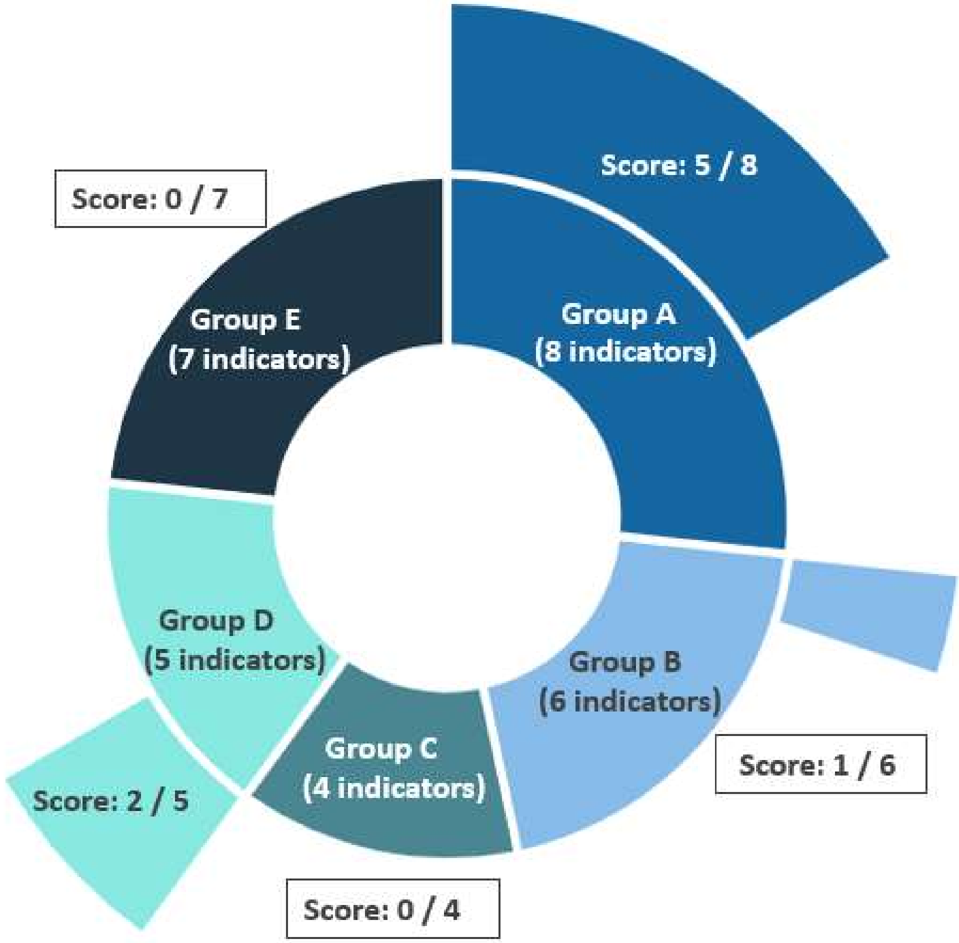


Figure 4.11 Summary graph illustrating the total number of data quality indicators possible for each group and the number of indicators that were present in Council Directive of 21 May 1991 concerning urban wastewater treatment (91/271/EEC), amended in 1998, 2003, 2008, and 2013

4.2.12 EU – Groundwater

Legislation: Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy [81]

Code: EU GW

Initially, Directive 2006/118/EC was considered for assessment, as this Directive concerns the protection of groundwater against pollution and deterioration [53]. However, Note 17 ([53], page 5) states that "Articles 8(3) and 20 of Directive 2000/60/EC provide for the adoption of standardised methods for analysis and monitoring of water status and, where necessary, of guidelines on implementation including monitoring." In order to draw comparisons with the compliance monitoring requirements of the other standards that have been assessed, Directive 2000/60/EC [52] has been reviewed as the standard for groundwater monitoring requirements in the EU. Where required, Directive 2006/118/EC is utilised as a supplementary document.

As with previous standards, neither Directive describes continuous monitoring requirements, so Group C indicators are all scored as not applicable.

Unless otherwise specified, the document referenced in Table 4.13 is EU GW [81].

Code	Score	Justification
A1	0	While Annex V Section 2.4 (pages 61-63) includes guidance on groundwater surveillance and operational monitoring, information on sampling methods or references to standard methods are not present. No supporting methodology is present in 2006/118/EC [53].
A2	0	As for A1.
A3	0	As for A1.
A4	0	As for A1.
A5	1	Article 4 Point 3 of 2006/118/EC ([53], page 9) states that "Choice of the groundwater monitoring sites as to satisfy the requirements of Section 2.4 of Annex V to Directive 2000/60/EC". 2000/60 Annex V Section 2.4 (pages 61-63) describes the method of selecting appropriate monitoring sites.
A6	0	As for A1.
A7	0	As for A1.
A8	0	As for A1.
B1	0	Not present in either Directive.

Code	Score	Justification
B2	0	Not present in either Directive.
B3	0	None specified in 2006/118/EC. In 2000/60/EC, the parameters are listed in Annex V Section 2.4.2 (page 62), but no standards or methods are mentioned.
B4	0	Not present in either Directive.
B5	0	Not present in either Directive.
B6	0	Not present in either Directive.
C1	N/a	
C2	N/a	
C3	N/a	
C4	N/a	
D1	1	Annex V Section 2.4.5 (page 63) describes that the calculated mean value is what compliance is based on.
D2	0	As indicator D1 specifies that only the mean value is used for compliance, there is no additional requirement to include input data.
D3	0	Not present in either Directive.
D4	0	Not present in either Directive.
D5	0	2006/118/EC Annex I point 1 ([53], page 14), specifies two quality standards with their relevant units. However, 2000/60/EC Annex V Section 2.4.2 (page 62) lists the parameters that must be monitored, and none has units. As well, more parameters are present in 2000/60/EC than 2006/118/EC. Therefore this indicator is scored as a zero, as the overall affect does not increase the data quality of the dataset.
E1	0	Annex V Section 2.4.5 (page 63) says that the required reporting method is a colour-coded map.
E2	0	Not present in either Directive.
E3	0	Not present in either Directive.
E4	0	Not present in either Directive.
E5	0	Not present in either Directive.
E6	0	Not present in either Directive.
E7	0	Not present in either Directive.

Table 4.13 Data Quality Indicator Scores and Justifications for Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy

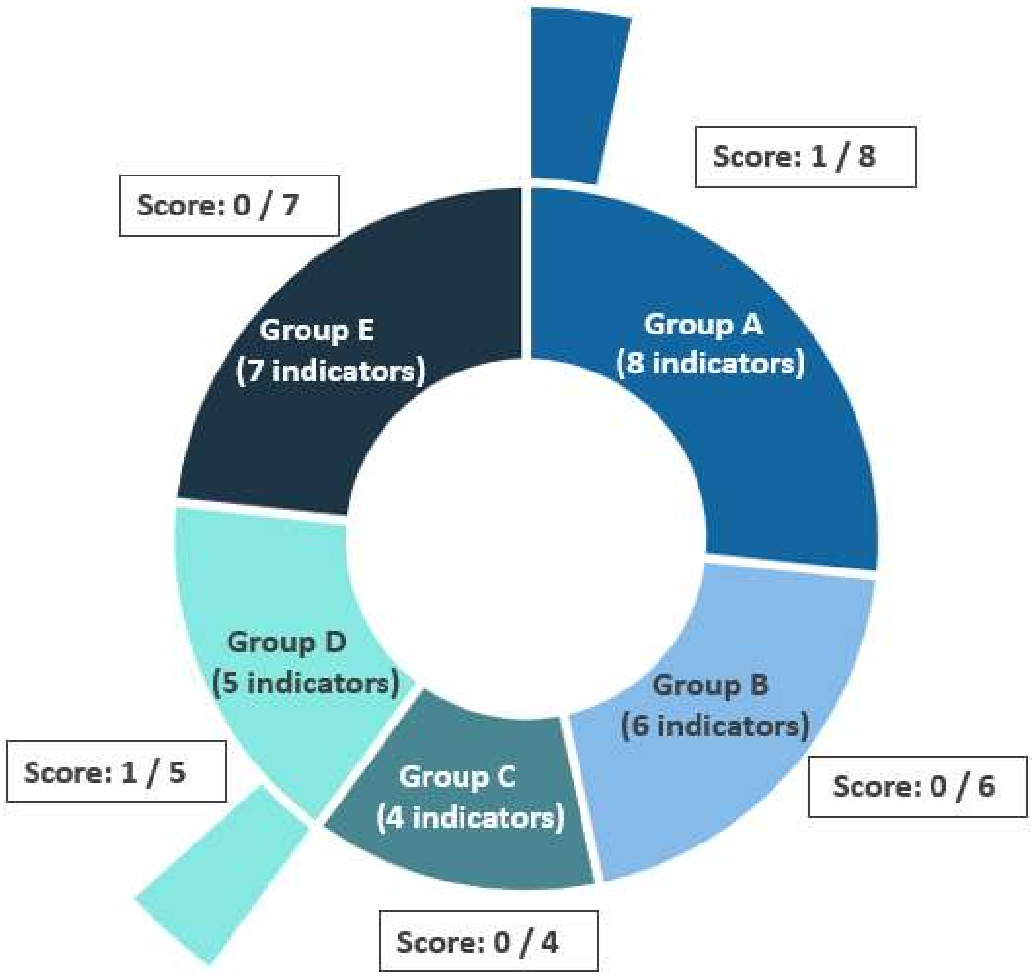


Figure 4.12 Summary graph illustrating the total number of data quality indicators possible for each group and the number of indicators that were present in Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy

4.2.13 EU – Carbon

Legislation: Regulation (EU) 2018/1999 of the European Parliament and of the Council of 11 December 2018 on the Governance of the Energy Union and Climate Action, amending Regulations (EC) No 663/2009 and (EC) No 715/2009 of the European Parliament and of the Council, Directives 94/22/EC, 98/70/EC, 2009/31/EC, 2009/73/EC, 2010/31/EU, 2012/27/EU and 2013/30/EU of the European Parliament and of the Council, Council Directives 2009/119/EC and (EU) 2015/652 and repealing Regulation (EU) No 525/2013 of the European Parliament and of the Council [54]

Code: EU Carbon

Article 18 (1b) (page 26) of these Regulations requires Member States to report information on "their national projections of anthropogenic greenhouse gas emissions by sources and removals by sinks, organised by gas or group of gases… listed in Part 2 of Annex V" ([54], page 26). Therefore, similar to the EU DW Directive, this Directive sets out the compliance requirements for the Member States to enact.

As with the NZ Carbon regulations, this Directive references a supplementary methodology document for details around the monitoring requirements. Annex V, Part 3 ([54], page 65) states that methodology is "in accordance with the 2006 IPCC Guidelines for National Greenhouse Gas Inventories". The 2006 IPCC Guidelines are extensive; however, three chapters in particular are relevant, namely Chapter 2 – Approaches to Data Collection [55], Chapter 6 – Wastewater Treatment and Discharge [56], and Chapter 8 – Reporting Guidance and Tables [57]. Where these chapters have guided the scoring for a DQ indicator, notes have been made in the justification section.

Indicators B6 and D2 are both scored as a 1 due to the verification of EU Carbon. This is detailed in Article 38, Point 2(a), where the Directive requires the comprehensive review of national inventory data submitted by the Member States to include "checks to verify the transparency, accuracy, consistency, comparability and completeness of information submitted" ([54], page 38). As this Directive requires that the datasets be audited, it conveys a greater level of data quality to the datasets.

Unless otherwise specified, the document referenced in Table 4.14 is EU Carbon [54].

Code	Score	Justification
A1	1	Annex V, part 3 (page 65) states that the methodology is "in accordance with the 2006 IPCC Guidelines for National GHG Inventories".
A2	N/a	The analysis is done on site, not at a testing laboratory.
A3	N/a	As for A2.

Code	Score	Justification
A4	N/a	As for A2.
A5	1	As for A1.
A6	1	As for A1.
A7	1	As for A1.
A8	0	Not present in these Regulations.
B1	1	As for A1. In addition, IPCC Chapter 2 (page 15) guides the reader to ISO 17025:2005 as a useful QA/QC regime, noting that "it encourages the use of standard methods by qualified personnel using suitably tested equipment" [55].
B2	1	As for B2.
B3	1	IPCC Chapter 2, Table 2.3 (page 16) details standard measurement methods [55].
B4	1	As for B2.
B5	1	As for B3, this table includes uncertainty.
B6	1	As per the discussion above, while this is not specified in the Directive, the requirement of Article 38, Point 2(a) (page 38) conveys the effect of this data quality indicator on the dataset.
C1	0	Not present in these Regulations.
C2	1	IPCC Chapter 2 (page 15) states that it is good practice to use meters that are calibrated, maintained, and regularly inspected [55].
C3	1	As B1 and C2 are both scored as 1, this shows that indicator C3 is also recognised in this methodology.
C4	0	Not present in these Regulations.
D1	1	IPCC Chapter 6 details equations for calculating GHG emissions from wastewater [56].
D2	1	As for B6.
D3	1	IPCC Chapter 8 Section 8.2.6 (page 7) specifies that SI units should be used [57].
D4	0	Not present in these Regulations.
D5	1	IPCC Chapter 8 Section 8.2.6 (page 7) specifies "emissions and removals should be expressed in mass units" [57].
E1	1	Article 28 Point 2 (page 31) states that "Member States shall use the e-platform for the purposes of submitting to the Commission the reports referred to in this Chapter."

Code	Score	Justification
E2	0	Not present in these Regulations.
E3	0	Not present in these Regulations.
E4	0	E1 only requires the report to be made via the e-platform. It does not require the report to be sent via M2M transfer.
E5	1	As E1 references an online platform, it is reasonable to expect that the allowable file types are detailed by this platform.
E6	0	Not present in these Regulations.
E7	0	Not present in these Regulations.

Table 4.14 Data Quality Indicator Scores and Justifications for Regulation (EU) 2018/1999 of the European Parliament and of the Council of 11 December 2018 on the Governance of the Energy Union and Climate Action

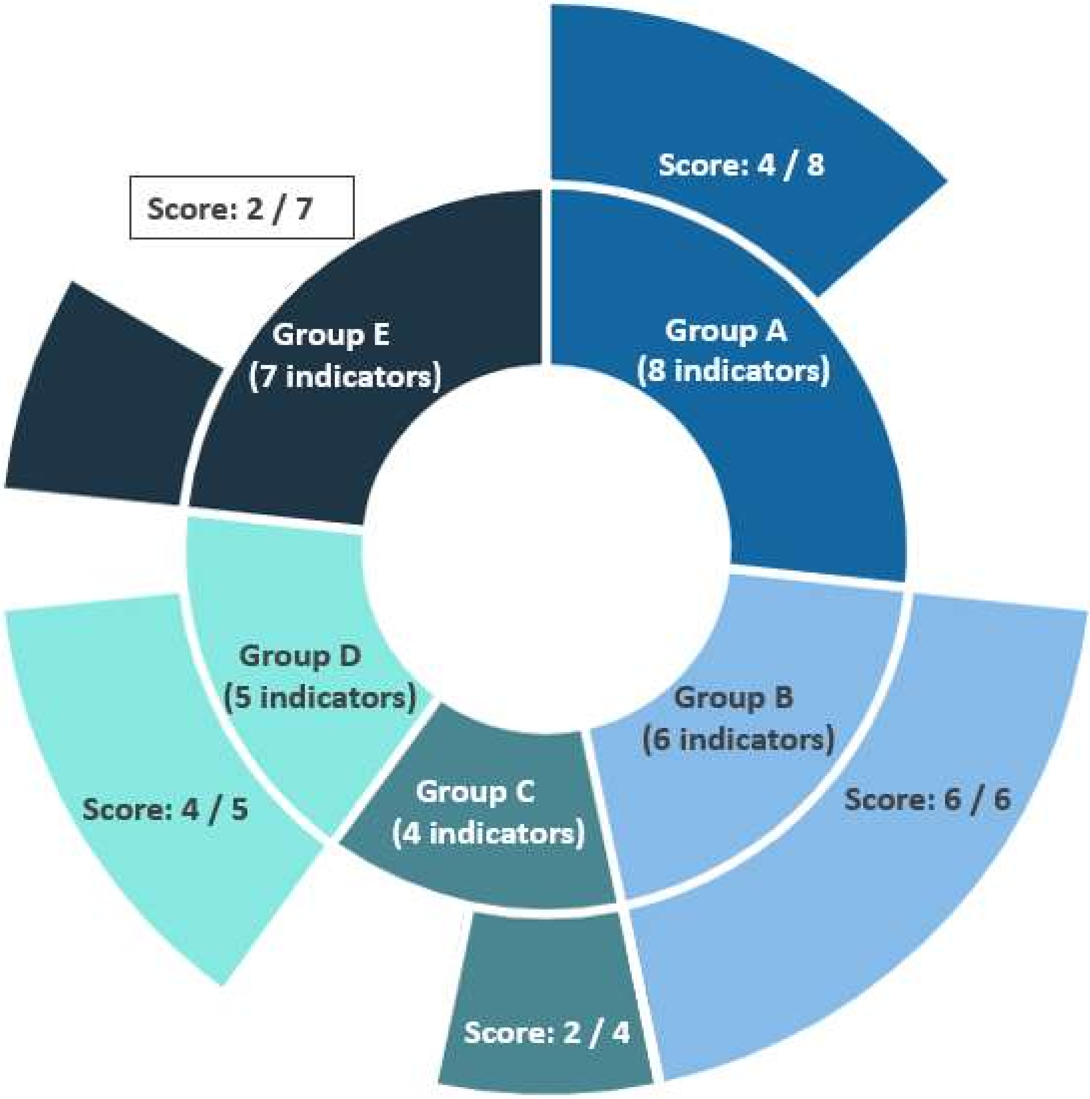


Figure 4.13 Summary graph illustrating the total number of data quality indicators possible for each group and the number of indicators that were present in Regulation (EU) 2018/1999 of the European Parliament and of the Council of 11 December 2018 on the Governance of the Energy Union and Climate Action

5 Results

5.1 Overview

The framework was developed to see if it facilitated the comparison of standards based on their data quality indicators. The application of the completed framework allows the assessor to determine the expected data quality of datasets submitted for each standard, providing an understanding of which compliance datasets can be interoperable with others from different industries, and the extent of this interoperability. In addition, by determining which DQ indicators were present or absent in each standard, the standards can be compared by their data quality requirements.

Having applied the framework to a series of regulatory standards, the output was first summarised as three summary tables. These tables are included in Appendix 1 as tabs A1.2, A1.3, and A1.4. The first and second tables both show the score each standard received for each DQ indicator, but they are grouped differently, in order to provide two means of comparing the standards. For these two tables, the indicators with a positive score are highlighted. This allows the assessor to quickly identify the DQ indicators present in each standard.

Summary Table 1 (A1.2) groups the standards by industry type, and Summary Table 2 (A1.3) groups the standards by area (country). Both tables include the totals for the DQ indicator and the standard.

Summary Table 3 (A1.4) shows the scores, weighted percentages, and non-weighted percentages for each standard, and is organised by group. This table allows the assessor to determine which stage of the compliance sample process has a greater level of DQ, and which does not.

All three tables facilitated the creation of graphs to allow a more detailed comparison of specific aspects of the DQ indicators of the standards. In addition, by focusing on select indicators or standards in the summary tables, additional comparisons of the standards can be made. The following graphs and tables are used to illustrate what data quality trends can be identified in the standards that were assessed.

5.2 Graphs and Tables

5.2.1 Data Quality performance across all standards

Graph 1 (Figure 5.1) shows the number of DQ indicators present in each standard. The number of possible DQ indicators is shown as the maximum possible score each standard could achieve, including indicators that were assessed as not applicable (N/a) for that standard. This graph illustrates the overall unweighted data quality score for the standard.

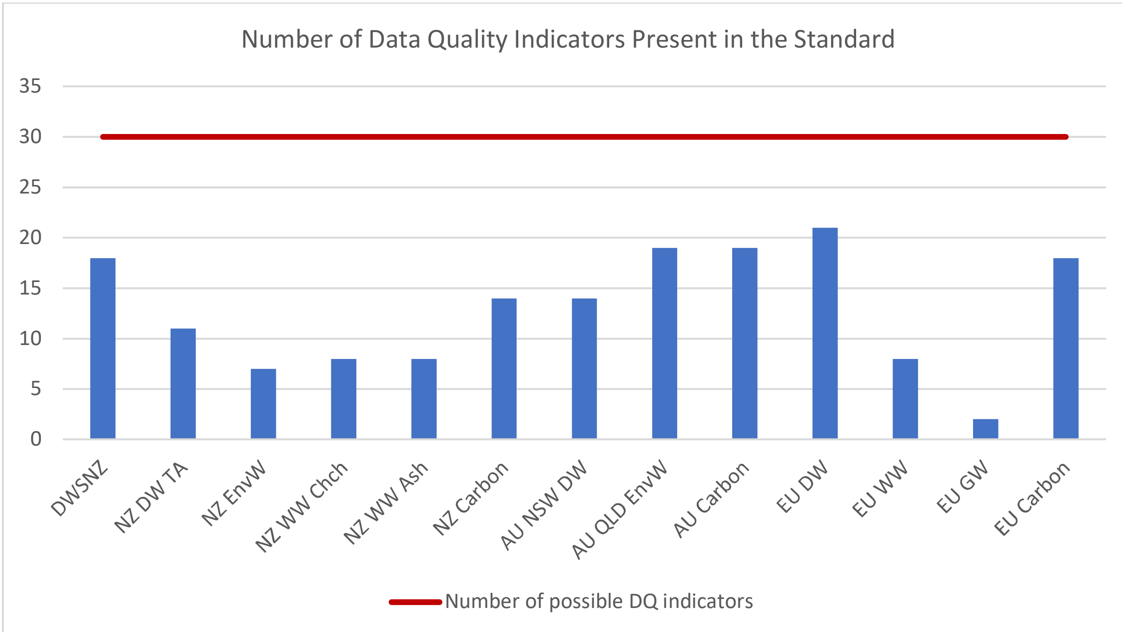


Figure 5.1 Graph 1: The number of data quality indicators present in each standard along with the total number of data quality indicators in the assessment framework

Graph 1 shows that only four standards contained more than 50% of the available DQ indicators; however, DQ indicators were present in every standard assessed.

While Graph 1 shows the unweighted DQ score for each standard, Graph 2 (Figure 5.2) shows the weighted percentage score for each standard. The weighted percentage score removes the indicators that were not applicable to the standard from the total number of possible indicators for that standard.

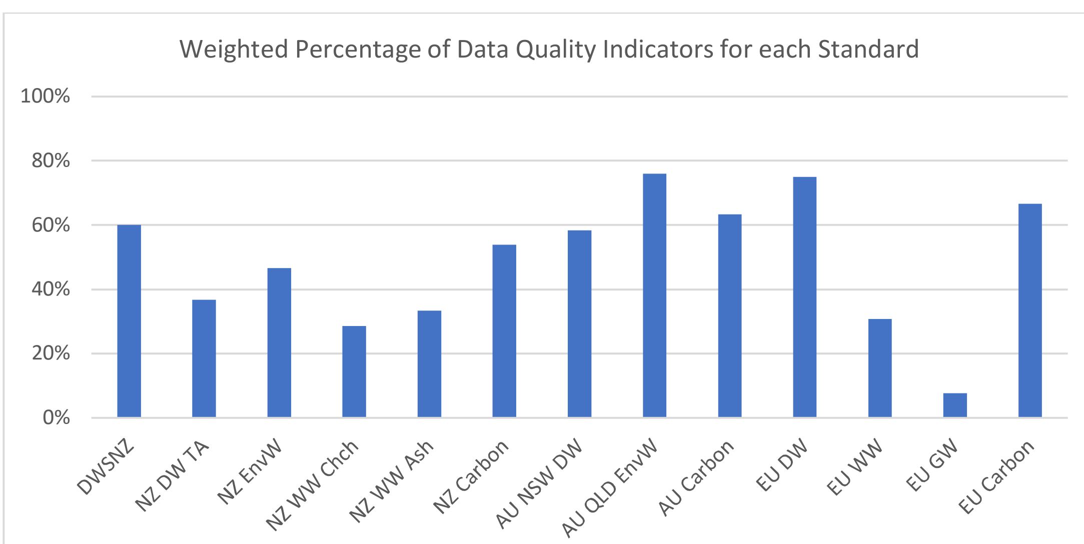


Figure 5.2 Graph 2: The weighted percentage score for the data quality indicators present in each standard

Graph 2 shows that by removing the DQ indicators that were scored as not applicable (N/a), seven standards contained more than 50% of the available DQ indicators. The comparison of these two graphs shows that the effect of DQ indicators being assessed as N/a affects the overall assessment of the level of data quality of the standard.

5.2.2 Data Quality performance by group for all standards

Graphs 3-7 (Figures 5.3 – 5.7) show the weighted and non-weighted scores by group for each standard. These graphs allow the assessor to determine the role of data quality in each stage of a compliance sample process.

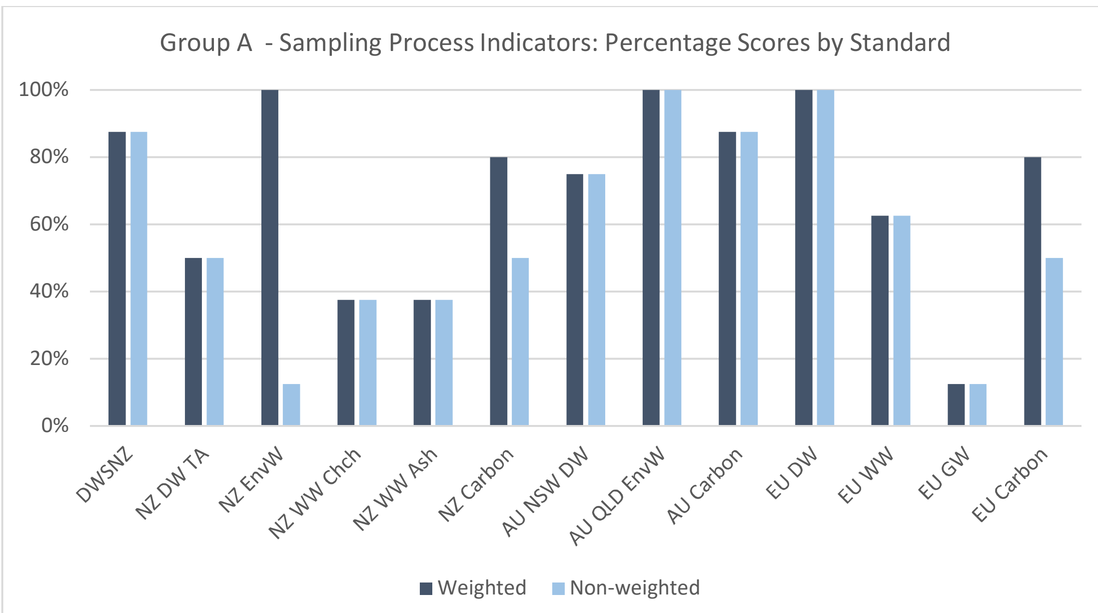


Figure 5.3 Graph 3: A comparison of the weighted and non-weighted percentage scores for the Group A – Sampling Process data quality indicators present in each standard

Graph 3 shows that while the non-weighted score for Group A DQ indicators (Sampling Process indicators) for the NZ EnvW standard [68] was low, this was due to the fact that most of the indicators were considered to be not applicable to this standard. As such the weighted score for this standard for Group A is 100%. This is confirmed by summary table 1, which shows that seven out of eight DQ indicators were assessed as being N/a, and the one indicator that was applicable was assessed as being present in the standard.

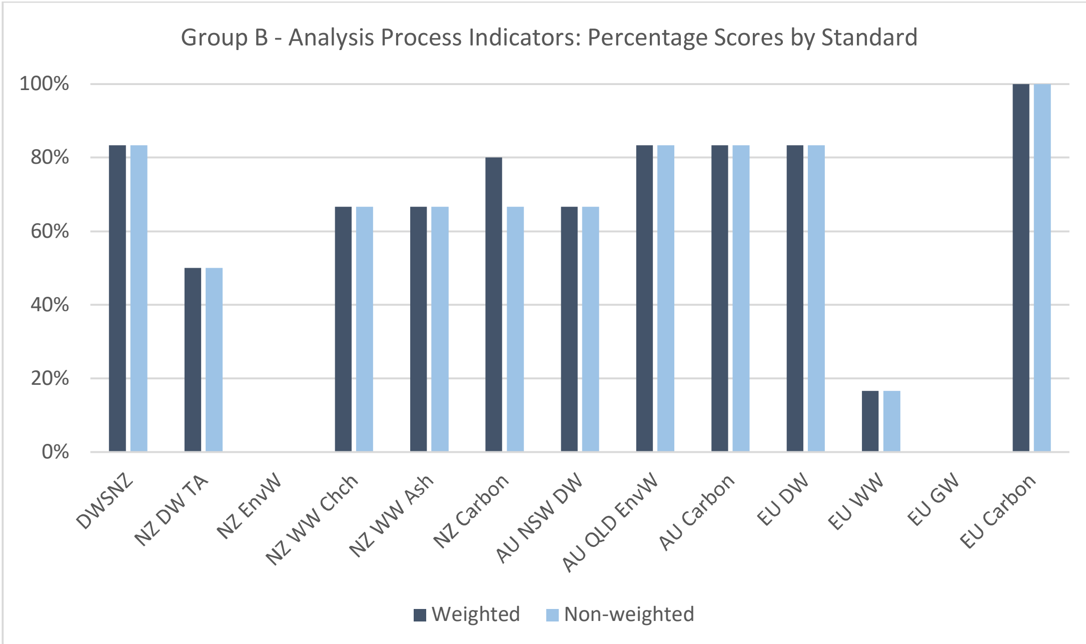


Figure 5.4 Graph 4: A comparison of the weighted and non-weighted percentage scores for the Group B - Analysis Process data quality indicators present in each standard

Graph 4 shows that none of the Group B indicators (Analysis Process indicators) were applicable to the NZ EnvW standard [68] or the EU GW Directive [52]. When compared with the summary table, this is shown to be because the Group B indicators were all scored as N/a for NZ EnvW, whereas for the EU GW Directive, they were all scored as zero.

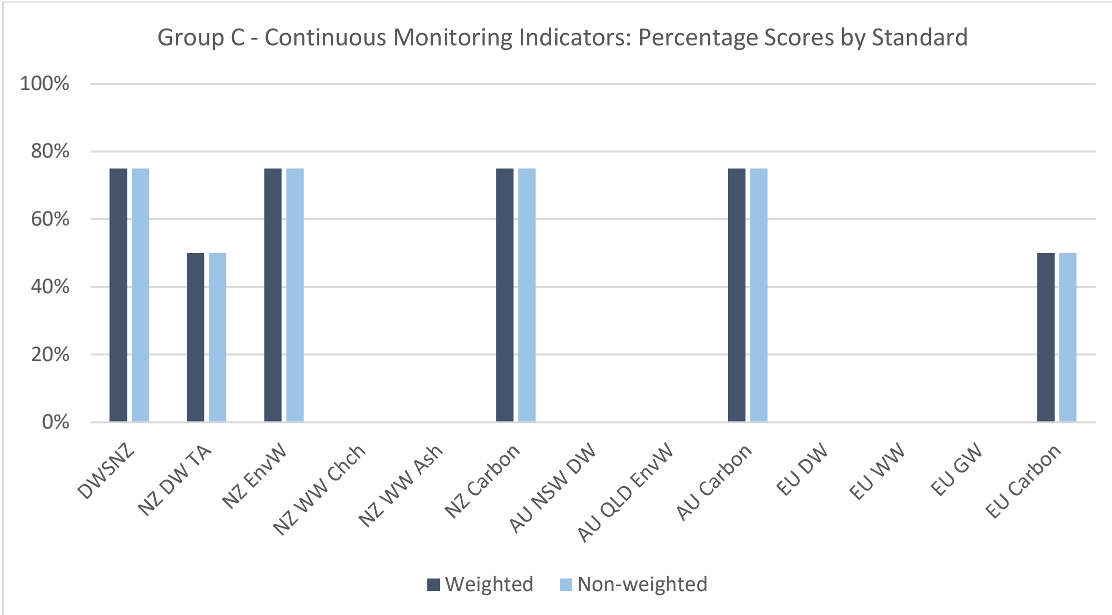


Figure 5.5 Graph 5: A comparison of the weighted and non-weighted percentage scores for the Group C – Continuous Monitoring data quality indicators present in each standard

Graph 5 shows that for standards where continuous monitoring requirements were present (Group C indicators), none of those DQ indicators was determined to be not applicable. This demonstrates that the DQ indicators were all relevant when assessing continuous monitoring requirements.

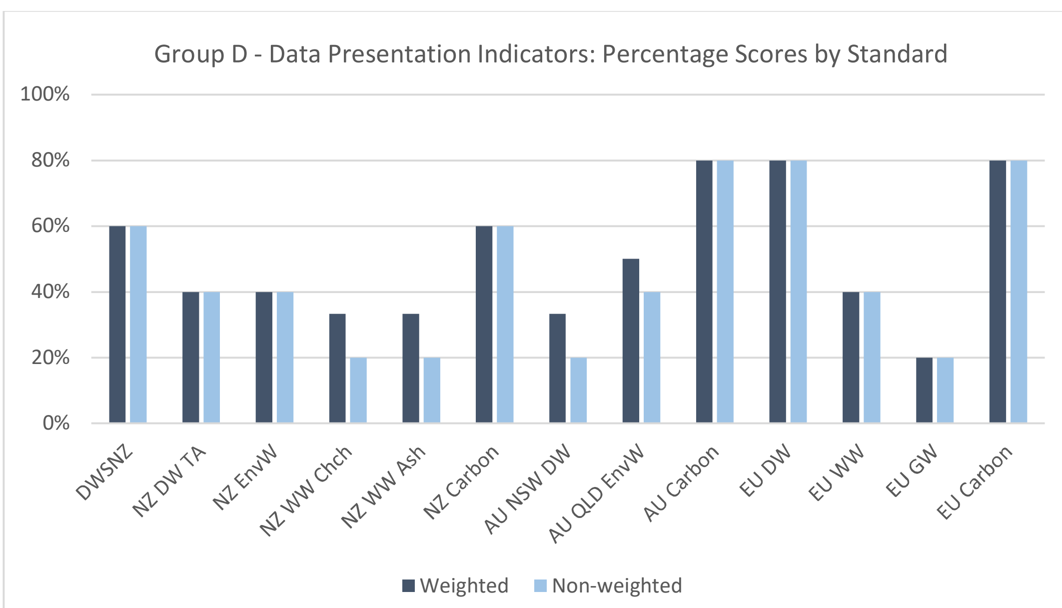


Figure 5.6 Graph 6: A comparison of the weighted and non-weighted percentage scores for the Group D – Data Presentation data quality indicators present in each standard

Graph 6 shows that the DQ indicators in Group D (Data Presentation) were not often assessed as N/a, as only four standards show a difference between the weighted and non-weighted scores.

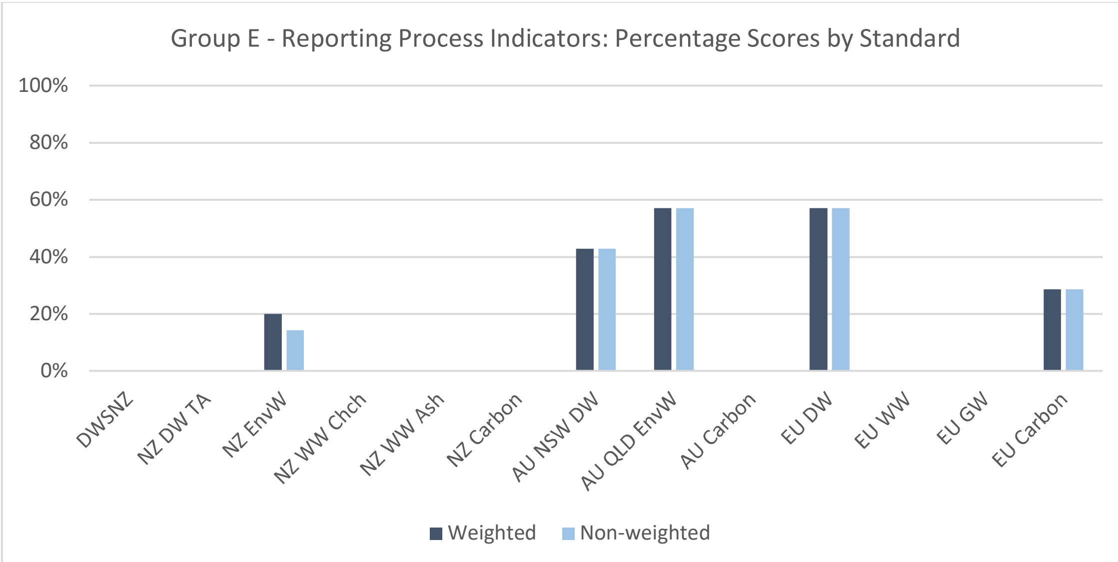


Figure 5.7 Graph 7: A comparison of the weighted and non-weighted percentage scores for the Group E – Reporting Process data quality indicators present in each standard

Graph 7 shows that only five standards contained DQ indicators from Group E (Reporting Process), and that this group is likely to be the largest barrier to the overall data quality scores for each standard.

5.2.3 Assessment of Data Quality Indicators

While Graphs 1-7 allow a high-level overview of the extent to which DQ indicators are present in the assessed standards, the framework also allows the assessor to understand which indicators are common between standards.

Firstly, Graph 8 (Figure 5.8) shows the number of standards that contained each DQ indicator. The number of standards reviewed is shown as the maximum possible score each indicator could achieve.

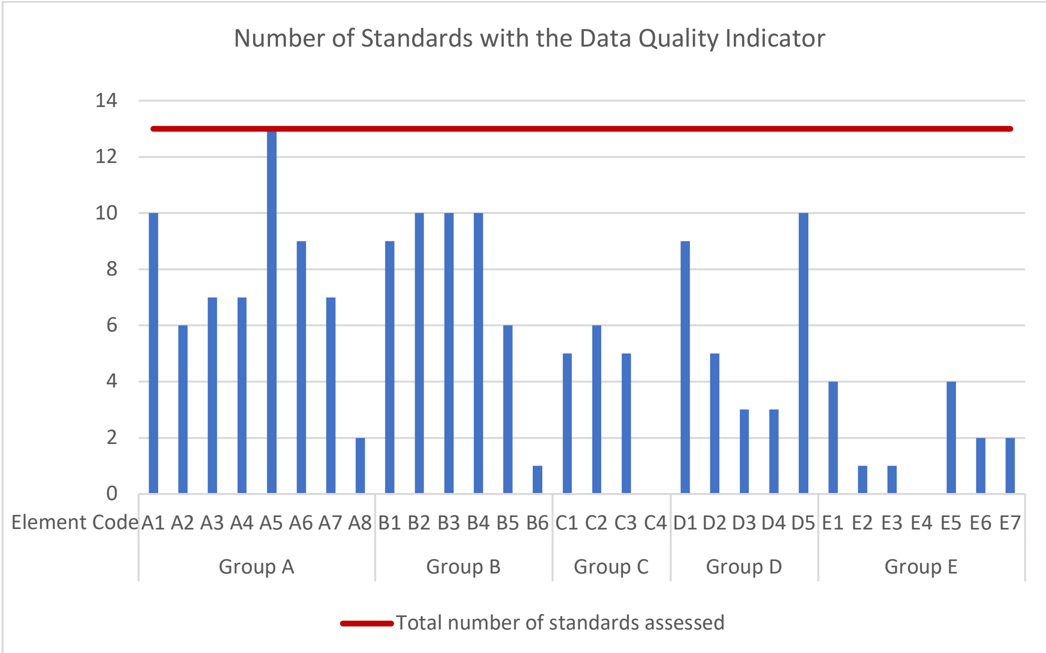


Figure 5.8 Graph 8: The number of standards that contained each data quality indicator, along with the total number of standards assessed by the framework

Graph 8 shows that indicator A5 (sampling location) was present in every standard, and confirms that indicators in Group E (Reporting Process) are not often present.

In order to assess the presence of each DQ indicator across the standards, the summary tables can be utilised. Table 5.1 below shows which Group A indicators (Sampling Process) are present in which standards.

This comparison method allows the assessor to make a series of comparisons, including:

- None of the New Zealand standards include indicator A8 (sample GPS data)
- DQ indicators A2 (sample container requirements), A3 (sample transport requirements), and A4 (sample delivery time frame requirements) were determined to be not applicable to the NZ Carbon and the EU Carbon standards, but were applicable and present in the AU Carbon standard
- Both of the New Zealand wastewater resource consents (NZ WW Chch and NZ WW Ash) had the same Group A (Sampling Process) score, and the same indicators were present
- The DWSNZ has more Group A indicators than its proposed replacement, NZ DW TA

	Group A Indicators							
Standard	A1	A2	A3	A4	A5	A6	A7	A8
DWSNZ	1	1	1	1	1	1	1	0
NZ DW TA	0	0	1	1	1	0	1	0
NZ EnvW	N/a	N/a	N/a	N/a	1	N/a	N/a	N/a
NZ WW Chch	1	0	0	0	1	1	0	0
NZ WW Ash	1	0	0	0	1	1	0	0
NZ Carbon	1	N/a	N/a	N/a	1	1	1	0
AU NSW DW	1	1	1	1	1	1	0	0
AU QLD EnvW	1	1	1	1	1	1	1	1
AU Carbon	1	1	1	1	1	1	1	0
EU DW	1	1	1	1	1	1	1	1
EU WW	1	1	1	1	1	0	0	0
EU GW	0	0	0	0	1	0	0	0
EU Carbon	1	N/a	N/a	N/a	1	1	1	0

Table 5.1 Scores for the Group A – Sampling Process data quality indicators for each standard assessed by the framework, highlighting which indicators were present in each standard

5.2.4 Interoperability of European Union Directives

Directives of the European Union were of interest because the Directives provide the Member States with compliance requirements. The Member States then write country-specific regulations to ensure that these compliance requirements can be met. Therefore, the Directives were expected to prioritise DQ indicators that would facilitate interoperability and reuse of datasets. Graph 9 (Figure 5.9) shows the weighted group scores for each Directive that was assessed by the framework.

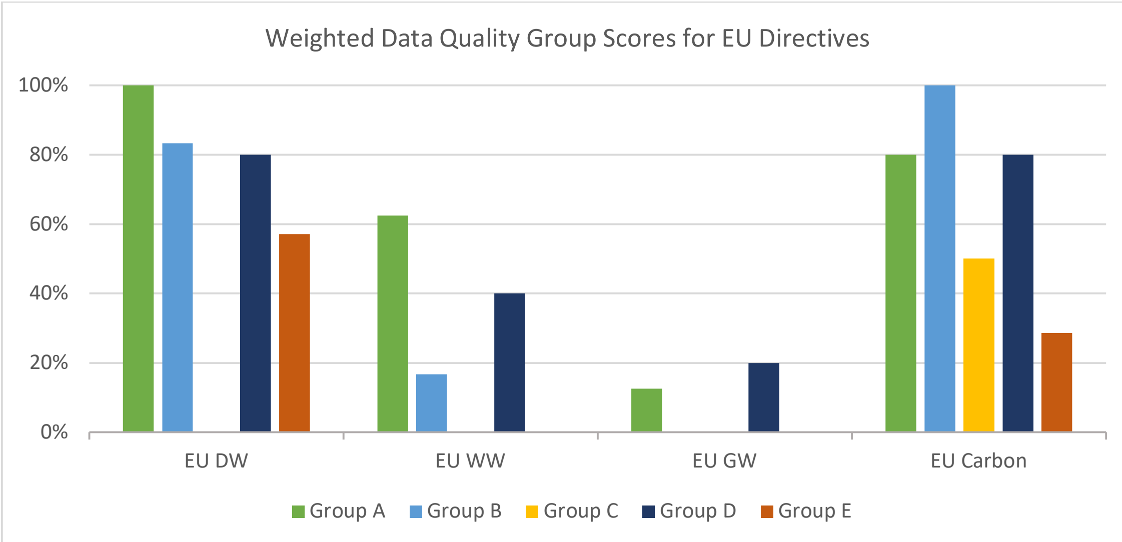


Figure 5.9 Graph 9: A comparison of the weighted percentage scores of the data quality indicator groups for the European Directives assessed by the framework

Graph 9 shows that the EU DW and the EU Carbon Directives had the most representation of DQ indicators. It also shows that all applicable Group A (Sampling Process) indicators were present in the EU DW Directive, and all applicable Group B (Analysis Process) indicators were present in the EU Carbon Directive. It also shows that the score for Group D (Data Presentation) indicators was the same for both Directives. In order to determine which DQ indicators were common across all four Directives, the summary table can be utilised (Table 5.2).

Group	Code	EU DW	EU WW	EU GW	EU Carbon
A	A1	1	1	0	1
	A2	1	1	0	N/a
	A3	1	1	0	N/a
	A4	1	1	0	N/a
	A5	1	1	1	1
	A6	1	0	0	1
	A7	1	0	0	1
	A8	1	0	0	0
B	B1	1	0	0	1
	B2	1	0	0	1
	B3	1	1	0	1
	B4	1	0	0	1
	B5	1	0	0	1
	B6	0	0	0	1
C	C1	0	N/a	N/a	0
	C2	0	N/a	N/a	1
	C3	0	N/a	N/a	1
	C4	0	N/a	N/a	0
D	D1	1	1	1	1
	D2	1	0	0	1
	D3	0	0	0	1
	D4	1	0	0	0
	D5	1	1	0	1
E	E1	1	0	0	1
	E2	0	0	0	0
	E3	0	0	0	0
	E4	0	0	0	0
	E5	1	0	0	1
	E6	1	0	0	0
	E7	1	0	0	0

Table 5.2 The scores for each data quality indicator, organised by group, for the European Directives assessed by the framework, highlighting which indicators were present in each standard

This table shows that while the Group D scores for the EU DW and EU Carbon Directives were the same, the distribution of indicators is different. The EU DW Directive did not have indicator D3 (units – system used), and the EU Carbon Directive did not have indicator D4 (GPS datum). In addition, while both standards included continuous monitoring requirements, neither specified the allowable interval between readings.

The DQ indicators present in all four standards were A5 (sampling location) and D1 (calculation method). Indicator D1 in particular supports the role of the Directives as providing guidance to Member States; by specifying calculation requirements, the calculated data can be compared across datasets from different countries. The EU DW and EU Carbon Directives further support this by including indicator D2, which requires the input data to be included alongside the calculated data.

The direct comparison of DQ indicators allows the assessor to understand the extent of the interoperability and reusability of datasets compiled for these Directives. In particular, the EU DW Directive, by including all of the Group A indicators, shows that drinking water sample data can be compared, because the sampling methods are specified. This is further supported by the presence of DQ indicators B1-B5, which relate to the laboratory analysis of those samples; by including requirements for analysis methods and laboratory accreditation, further confidence is conveyed in the validity of the result. While indicator B6 is not present (meaning the standard allows the transfer of data from the laboratory to the supplier, rather than requiring that the data is sent directly from the laboratory to the regulator), indicator E1 is, meaning that the Directive requires electronic transfer of data to the regulator. While the possibility of transcription errors are not completely removed, they are reduced by the requirement for reports to be sent electronically.

5.2.5 A Comparison of the Data Quality Scores across Drinking Water Standards

While the interoperability of EU Directives can be visualised, the framework also allows standards from different countries to be compared. Graph 10 (Figure 5.10) shows the weighted data quality group scores for drinking water standards from New Zealand and Australia compared with the EU Directive.

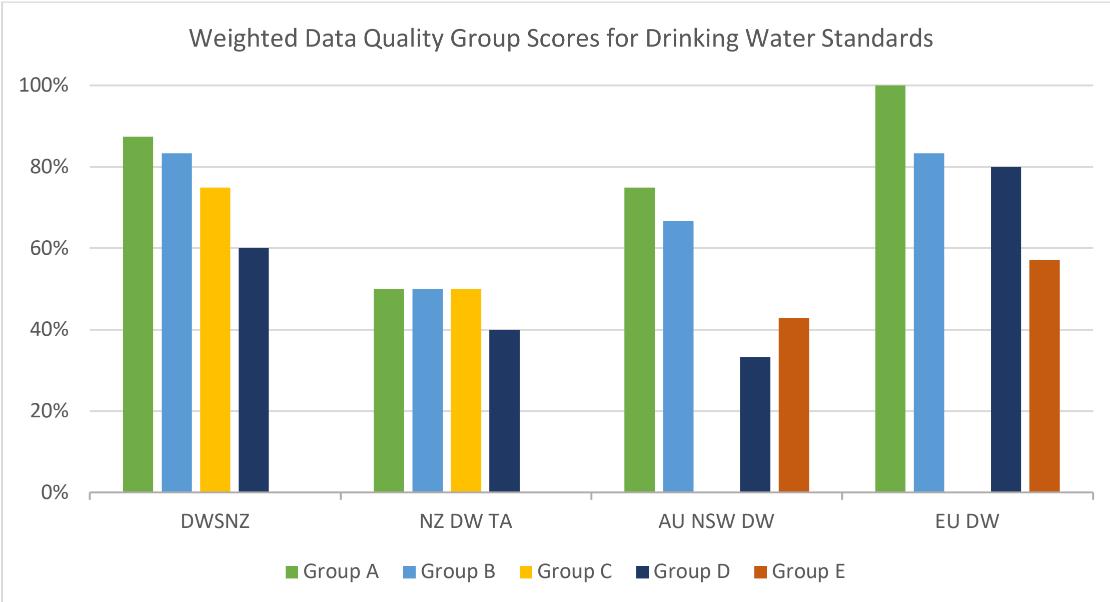


Figure 5.10 Graph 10: A comparison of the weighted percentage scores of the data quality indicator groups for the drinking water standards assessed by the framework

By combining this graph with a summary table (Table 5.3), the assessor can understand the level of data quality present in each standard.

Group	Code	DWSNZ	NZ DW TA	AU NSW DW	EU DW
A	A1	1	0	1	1
	A2	1	0	1	1
	A3	1	1	1	1
	A4	1	1	1	1
	A5	1	1	1	1
	A6	1	0	1	1
	A7	1	1	0	1
	A8	0	0	0	1
B	B1	1	1	1	1
	B2	1	1	1	1
	B3	1	0	1	1
	B4	1	1	1	1
	B5	1	0	0	1
	B6	0	0	0	0
C	C1	1	1	N/a	0
	C2	1	1	N/a	0
	C3	1	0	N/a	0
	C4	0	0	N/a	0
D	D1	1	0	N/a	1
	D2	0	1	N/a	1
	D3	1	0	0	0
	D4	0	0	0	1
	D5	1	1	1	1
E	E1	0	0	0	1
	E2	0	0	1	0
	E3	0	0	1	0
	E4	0	0	0	0
	E5	0	0	1	1
	E6	0	0	0	1
	E7	0	0	0	1

Table 5.3 The scores for each data quality indicator, organised by group, for the drinking water standards assessed by the framework, highlighting which indicators were present in each standard

For example, if datasets from each standard were to be compared, the assessor would have greater confidence in comparing sample results from the DWSNZ, the AU NSW DW, and the EU DW, as these three standards all included indicator A1, which ensures that a standard sample method was used. This indicator is not present in the current published draft of the NZ DW TA; while this may yet change with future iterations of the draft, the framework highlights that this reduces the interoperability of compliance datasets produced under that standard.

In addition, neither the current DWSNZ or the draft NZ DW TA standards include indicator E1 (which prohibits non-electronic transfer of datasets from the supplier to the regulator). However, this is present in the EU DW Directive. Therefore, if this indicator was required to be included in a future iteration of the draft NZ DW TA, the authors could use the justification for indicator E1 for EU DW to determine how that Directive ensures that electronic transfer is a requirement. Following from this, between AU NSW DW and EU DW, all Group E (Reporting Process) indicators (excepting E4, method of sending data to Regulator - M2M (machine to machine) requirement) are present; therefore, by understanding how these standards incorporated these DQ indicators, the framework illustrates how the data quality of a standard currently being written could be increased.

5.2.6 Interoperability of New Zealand Standards

Also of interest to this assessment was the ability of the framework to compare the potential interoperability of regulatory standards applicable to the three waters industry in New Zealand. Having compared the current and proposed new drinking water standards, extending this assessment to the data quality requirements of wastewater monitoring consents and water take monitoring regulations allows an assessment to be made of the current data quality requirements, and where improvements may be possible when writing new standards.

As with the comparison of the EU Directives, the weighted scores for each group are first visualised in a graph, and a summary table provides further information on which DQ indicators are present in each standard.

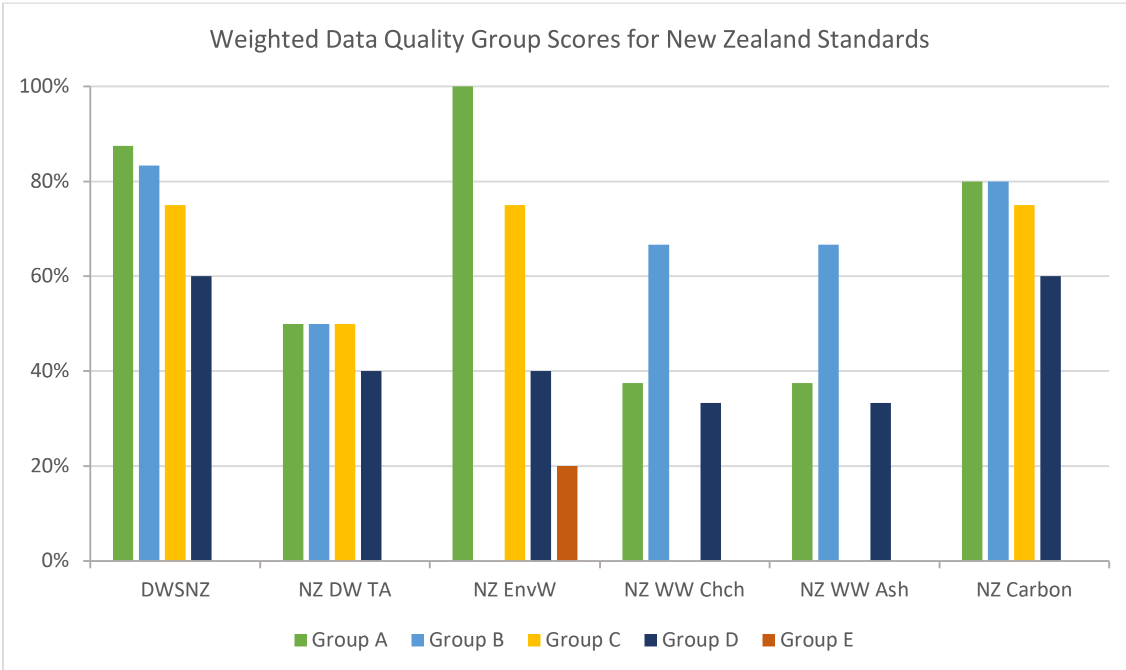


Figure 5.11 Graph 11: A comparison of the weighted percentage scores of the data quality indicator groups for the New Zealand standards assessed by the framework

This graph shows that some Group A (Sampling Process) indicators are present in all standards, while Group E (Reporting Process) indicators are only present in the NZ EnvW standard. It also shows that DWSNZ, NZ EnvW, and NZ Carbon all have the same Group C (Continuous Monitoring indicators) score.

Group	Code	DWSNZ	NZ DW TA	NZ EnvW	NZ WW Chch	NZ WW Ash	NZ Carbon
A	A1	1	0	N/a	1	1	1
	A2	1	0	N/a	0	0	N/a
	A3	1	1	N/a	0	0	N/a
	A4	1	1	N/a	0	0	N/a
	A5	1	1	1	1	1	1
	A6	1	0	N/a	1	1	1
	A7	1	1	N/a	0	0	1
	A8	0	0	N/a	0	0	0
B	B1	1	1	N/a	1	1	N/a
	B2	1	1	N/a	1	1	1
	B3	1	0	N/a	1	1	1
	B4	1	1	N/a	1	1	1
	B5	1	0	N/a	0	0	1
	B6	0	0	N/a	0	0	0
C	C1	1	1	1	0	N/a	1
	C2	1	1	1	0	N/a	1
	C3	1	0	1	0	N/a	1
	C4	0	0	0	0	N/a	0
D	D1	1	0	1	N/a	N/a	1
	D2	0	1	0	N/a	N/a	1
	D3	1	0	0	0	0	0
	D4	0	0	0	0	1	0
	D5	1	1	1	1	0	1
E	E1	0	0	1	0	0	0
	E2	0	0	N/a	0	0	0
	E3	0	0	0	0	0	0
	E4	0	0	0	0	0	0
	E5	0	0	0	0	0	0
	E6	0	0	0	0	0	0
	E7	0	0	N/a	0	0	0

Table 5.4 The scores for each data quality indicator, organised by group, for the New Zealand standards assessed by the framework, highlighting which indicators were present in each standard

Table 5.4 shows that the same Group C indicators are present for DWSNZ, NZ EnvW, and NZ Carbon. As NZ DW TA does not include indicator C3 (instrument calibration requirements – technician), these three standards provide points of reference for what is required to ensure that indicator C3 is included in a subsequent iteration of NZ DW TA.

Excepting the not applicable scores, indicators B1 (laboratory technician qualification), B2 (laboratory instrument calibration requirements), and B3 (analytical test method) are present in all standards. This shows that these standards include requirements around laboratory analysis methods and accreditation, conveying certainty that the results are accurate. By understanding the laboratory analysis methods, the results can also be compared with each other, increasing the reusability of the datasets. This is further supported by indicator D5 (units – reporting), which ensures that the results are reported with the units identified by the relevant standard. While this indicator is not present in NZ WW Ash, it is present in the other wastewater monitoring consent assessed (NZ WW Chch), therefore if a national-level wastewater regulation was drafted, there would be an example of current wastewater monitoring requirements available as a reference.

5.2.7 A Comparison of the Data Quality Scores of the New Zealand Drinking Water Standards

While section 5.2.6 highlights the similarities and differences in New Zealand standards, including, the current DWSNZ and the new NZ DW TA, the framework can also be used as a performance measure for comparing these two drinking water standards.

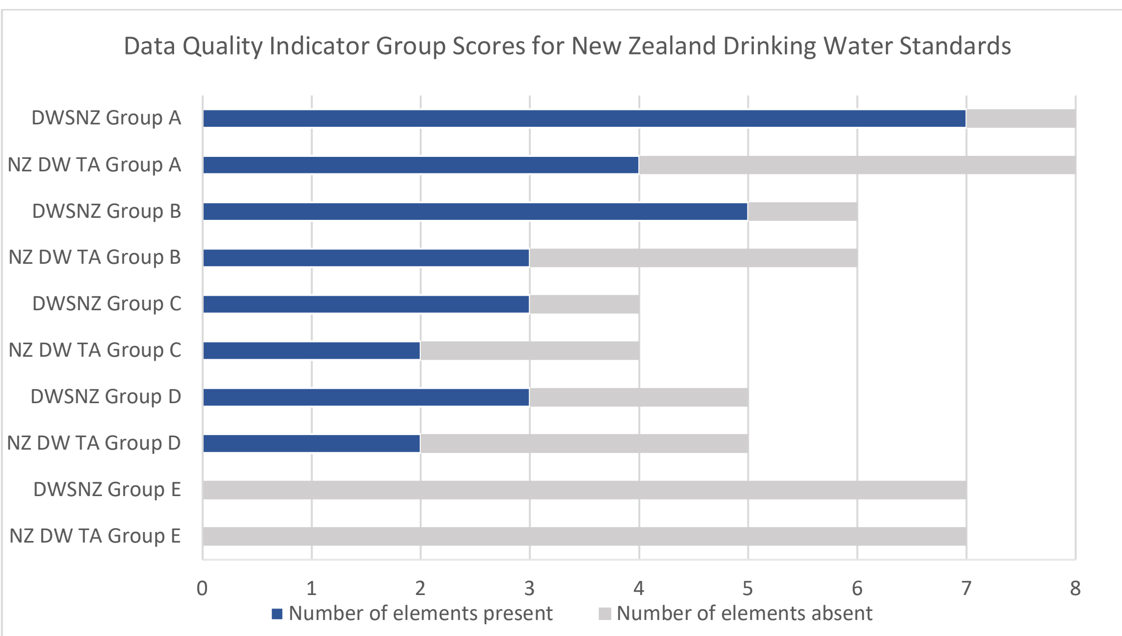


Figure 5.12 Graph 12: A comparison of the number of data quality indicators in each group that were present and absent for the two New Zealand drinking water standards assessed by the framework

Graph 12 shows the number of DQ indicators that are present, and those that are absent, for each standard. Note that for DWSNZ and NZ DW TA, no indicators were assessed as not applicable.

This method can be used as a benchmark for assessing the level of data quality present in the current draft NZ DW TA when compared with the existing DWSNZ. While the DQ assessment for the DWSNZ will remain the same, subsequent iterations of the draft NZ DW TA could be assessed to determine how the inclusion of data quality has changed. While the summary table supports the specifics of which DQ indicators are present and absent, this graph provides a high-level overview of changes over time.

6 Discussion

6.1 Assessment of the Framework

The application of the framework to environmental monitoring standards from different industries and countries demonstrated that its design was effective. While the scoring system was simple, by categorising the data quality indicators as groups, the standards could be evaluated in a range of ways. The questions were applicable to standards from across the three waters industry, and could also be applied to standards relating to greenhouse gas emission calculations. Not all data quality groups were applicable to all standards, for example, the AU QLD EnvW [41] was only applicable to field samples, and so the Group C indicators concerning continuous monitoring were not applicable. However, this was also a useful function of the framework, as the summary showed which standards covered continuous monitoring requirements, and which did not. This application of the framework would be useful for organisations that have more than one set of regulations to comply with, as it provides detail on the specific compliance requirements of the determinands being monitored.

The framework facilitated the comparison of standards at a hierarchical level as well as at the DQ indicator level. The hierarchies – represented in the framework as Groups – are a useful method of comparison through visualisation of the weighted average (calculated as the number of DQ indicators present in the standard divided by the total number of DQ indicators in the Group, where the total excludes indicators that were not applicable to that standard) for each group. This allowed the following questions to be explored:

- Which standards included DQ indicators that contributed to data quality via reporting methods (Group E)?
    - Graph 7 showed that only five standards included Group E indicators; these were NZ EnvW, AU NSW DW, AU QLD EnvW, EU DW, and EU Carbon
- Which standards included continuous monitoring requirements (Group C), but did not have a high DQ requirement for them?
    - Summary Tables 1 and 2 (Appendix 1, A1.2 and A1.3) showed that of the eight standards that had continuous monitoring requirements, six had at least two DQ indicators present; two standards (EU DW and NZ WW Chch) had no DQ indicators present.
- Which standards prioritised the DQ of sampling methods (Group A)?
    - Summary Table 3 (Appendix 1, A1.4) showed that three standards had a weighted percentage of 100% for the Group A indicators; these were NZ EnvW, AU QLD EnvW, and EU DW.

The framework also allowed the comparison of DQ indicators. For example, Summary Tables 1 and 2 highlighted which DQ indicators were present in each standard, allowing the assessor to easily compare the overall data quality requirements of the standards. In addition, Graph 8 (Figure 5.8) showed which DQ indicators were common, and which were unusual.

More in-depth analysis could also be made; for example, while Group E (Reporting Process) indicators were present in two of the three Australian standards, neither of these standards included a requirement for data transfer from the supplier to the regulator to be via a machine to machine (M2M) method (indicator E4).

6.2 The Role of the Framework in assessing Drinking Water Standards in New Zealand

The changing regulatory environment of drinking water in New Zealand is exemplified by the transition from the existing drinking water standards (DWSNZ) to the draft Drinking Water Quality Assurance Rules (NZ DW TA). This framework was developed with the goal of assessing these two sets of standards, to determine if the role of data quality in both standards could be identified and described. The application of the framework showed that this goal was achieved, as it enabled the two sets of standards to be compared by which data quality indicators were present or absent in the standards.

The framework proved to be a useful method of obtaining a high-level overview of the way each standard focuses on DQ indicators. While the NZ DW TA standard is a draft, meaning changes are likely prior to formal publication, the framework was able to identify which DQ indicators were present, and which were not. Interestingly, DQ indicators present in the DWSNZ were not always present in the NZ DW TA; most notably indicator A1, which references a standard sampling method.

As well as providing a method of comparing the DQ indicators present in each standard, the framework also allows the current data quality requirements of the DWSNZ to be summarised. This provided an indication of what the NZ DW TA rules should include (i) to be the same as the DWSNZ, and (ii) to improve on the DWSNZ. By benchmarking this assessment against drinking water standards from Australia and the European Union, the framework allowed those DQ indicators not present in the NZ DW TA standard to be identified in the AU NSW DW and EU DW standards; and by referencing the justification notes in the framework, examples of how to incorporate these DQ indicators could be determined.

While Taumata Arowai are currently writing a new drinking water standard, as the regulator for three waters in New Zealand, it is possible that regulations pertaining to wastewater monitoring and stormwater monitoring may also be developed. While assessing standards covering stormwater monitoring was outside the scope of this thesis, current wastewater monitoring and environmental monitoring standards were assessed to analyse the current data quality requirements of these standards, and how they compare with the draft NZ DW TA. The goal of this analysis was to determine if the framework could act as a comparison tool in the development of subsequent standards, to ensure the DQ indicators present in each standard contributed to the interoperability of the subsequent compliance datasets.

One of the issues identified in the literature review was that when compliance data is collected by a regulator, often the ability to share it within intergovernmental agencies is limited due to the specificity of the dataset, despite the compliance data being of use beyond the initial compliance assessment. As the regulator of the three waters industry in New Zealand, Taumata Arowai has a unique opportunity to create a series of standards where the datasets for each can be assessed together. This is an exemplar of the FAIR principles of interoperability and reusability. The comparison of DQ indicators present in the EU Directives show that even with interoperability as desirable output of the standard, it may not be achieved if the DQ indicators that support it are not aligned between the standards. However a reference framework that allows performance benchmarking (as with section 5.2.7) would be of use, as it would allow the data quality present in draft new standards to be critically assessed and improved upon with each iteration prior to final publication.

By working to align the DQ indicators for New Zealand's potential future Three Waters standards, the resulting datasets can be viewed not just for compliance purposes, but alongside other environmental monitoring datasets. This is an exemplar of the FAIR principle of reusability, where the compliance dataset is able to be reused for other goals – either within Taumata Arowai, or externally, for example as part of a research project.

Finally, using the framework assessment method to identify and improve DQ indicators also supports the FAIR principles of findability and accessibility. This is because a standard DQ system with defined reporting requirements enables datasets to be systematically archived, increasing their findability. As well, if a common method is used across different standards, this allows the datasets to be accessed by users (with appropriate authorisation) from both intergovernmental departments such as Taumata Arowai, as well as those interested in the research potential of the datasets.

6.3 The Framework as a Method of Assessing Data Quality of Regulatory Standards

The framework proved to be useful for visualising DQ indicators in varied environmental monitoring standards. It also allowed investigation into those indicators, as the assessor could reference the justification for each DQ indicator to understand in what manner it was present in the standard, or why it was not present. This application has utility for data users wishing to compare datasets produced under different regulatory systems; by understanding how each regulation prioritises data quality, the level of data quality of the dataset can be understood. If the data quality of a dataset is understood, it is more likely to be reused.

The framework also has utility for regulators such as Taumata Arowai who are in the process of developing a new standard. The framework shows which DQ indicators are present in comparable standards in other countries, and also shows the current DQ indicators for existing standards in that regulator's country. This provides a method by which to benchmark the level of data quality present in new draft standards against the existing standards and those from other countries. It also provides a method of aligning DQ requirements across standards, leading to increased interoperability and facilitating the development of open governmental data (OGD).

6.4 Potential Improvements to the Framework as a Data Quality Assessment Method

The framework focuses on environmental monitoring, and was written with a focus on drinking water, wastewater, source water, and carbon standards. While it may be applicable to other environmental monitoring standards, it has not been tested on these, and so it may not include DQ indicators relevant to those standards; as well, the DQ indicators present in this framework may be irrelevant to other types of environmental monitoring standards. An example of this was exemplified by the scores for the continuous monitoring group (Group C); as not all the standards assessed contained requirements for continuous monitoring.

In addition, by including specific metadata elements as well as DQ indicators, the framework could also be made more specific. As the DQ indicators focused on assessing the method in which the standard prioritised data quality, the output was restricted to an assessment of the standard as a whole. Conversely, if metadata elements were assessed, the framework would be able to provide specific comparisons. For example, if the metadata element of 'allowable UV validation certifications' was used in the assessment, the output of the framework would show which drinking water standards included this element, and also what the standards were. The result of this level of assessment would be that compliance datasets for standards that allowed the same UV certifications would be directly comparable; where the certification was unknown, the resulting dataset would have a lower level of confidence, and may not be included in the comparison.
This type of framework would benefit situations where the specific metadata elements are required to be identified. One example would be the EU DW Directive, where if the metadata elements are specified, the Member States would have specific points of reference when compiling national-level regulations. It would also benefit instances where new standards are being written, for example the current update to the drinking water standards in New Zealand.

However the specificity of this type of framework design is restrictive. Metadata elements relevant to drinking water standards are unlikely to be relevant to wastewater standards, and less so to greenhouse gas emission calculation standards. A framework that included metadata elements would therefore be limited to industry-specific applications, and would not be as effective as providing a hierarchical level overview of standards from different industry types.

Therefore the suggested improvement for the framework used in this thesis is to use it to identify which data quality groups in which standards require the most improvements; then, using a similar methodology, identify industry-specific metadata elements to construct a secondary framework. By combining both methods, the high-level overview of multiple standards can be achieved, and used as a performance indicator; the secondary framework then provides the detail required to make industry-specific data quality improvements. This would be useful for instances when the interoperability between data sets needs to be exactly compared, for example when a common database is being developed to allow two or more compliance systems to be combined.

7 Conclusions

7.1 Assessment of the Research Objectives and Questions

This thesis aimed to identify and define the data quality indicators of regulatory standards that apply to drinking water, wastewater, environmental water, and greenhouse gas emission calculations, and to compile these indicators in a framework that allows the data quality requirements of a set of regulatory standards to be critically compared with another. The primary research question was as follows:

> Can data quality evaluation techniques and assessment methodologies be used to develop a framework to allow the critical assessment of the data quality requirements and indicators of environmental monitoring standards, leading to the determination of the quality of datasets submitted in pursuit of compliance with these standards?

Once the data quality indicators were identified and grouped in a process flow manner, the framework was applied to the regulatory standards identified. The resulting data generated showed that while the method was straightforward, the insights it provided about the role of data quality in the standards was detailed. The framework was established as being a useful tool in the assessment of current standards, and is of use to a wide range of stakeholders in the three waters industry, including regulators, suppliers, policy writers, and researchers.

The application of the framework demonstrated that the research objectives of this thesis were achieved. The framework showed that by using data quality methodologies in a novel manner, the data quality requirements of regulatory standards could be critically assessed.

7.1.1 Assessment of the Secondary Research Questions

> What are the data quality indicators of regulatory compliance standards?

By reviewing the current drinking water standards for New Zealand (DWSNZ), a series of data quality indicators were identified. These indicators were designed to be applied to the standard as a whole, rather than focusing on specific metadata elements of drinking water treatment. As a result, these DQ indicators were able to be identified in regulatory standards concerned with wastewater monitoring, environmental water monitoring, and greenhouse gas emissions monitoring.

> Do data quality indicators of regulatory compliance standards allow the assessment of the data quality requirements of these standards?

Section 5.2.3 showed that by determining if these DQ indicators were present or absent in the assessed standards, insights into the overall data quality requirements of the standards could be made. For example, the DQ indicators in Group A (Sampling Process) showed that the current DWSNZ had more indicators present than the draft Drinking Water Quality Assurance Rules (NZ DW TA); overall, the current standard has more stringent data quality requirements around the compliance sampling process than the draft NZ DW TA.

In addition, section 5.2.3 showed that the indicators in Group E (Reporting Process) were not often present in the assessed standards. This shows that the data quality requirements of the assessed standards were more likely to be around the sampling process (Group A) and the analysis process (Group B), and there were fewer data quality requirements concerned with the method by which the data was sent to the regulator.

Can data quality indicators be used as a point of comparison for regulatory standards?

Section 5.2.6 showed that the data quality indicators present in the DWSNZ and the proposed new draft rules were not the same; the current standards contain more DQ indicators than the NZ DW TA. This showed that the DQ indicators provided insight into the overall data quality requirements of a standard.

Section 5.2.7 showed that the presence or absence of DQ indicators could be used as a method of assessing the progress of the draft NZ DW TA against the current DWSNZ. As the draft rules are expected to change before becoming the new drinking water standard for New Zealand, this section showed that the framework method of assessing the DQ indicators provided a way to visualise what DQ indicators are potentially required to be included in subsequent drafts. Viewed alongside the results from section 5.2.5, which compared the DQ indicators of drinking water standards from other countries, the new draft rules could also incorporate international best practices.

7.1.2 Results Analysis Conclusions

Data quality assessment methods are typically designed to assess datasets. This is achieved through the development of context-specific data quality indicators. However this method requires each dataset to be assessed individually. As compliance datasets have a fixed context – the dataset is created to demonstrate compliance with a regulatory standard – this thesis investigated whether an assessment of the data quality indicators of the standard could be made instead, where the resulting assessment of the standard could determine the expected data quality of all datasets submitted for compliance.

By determining the data quality indicators of regulatory standards, grouping these by process stage, and including them in a framework, a series of regulatory standards were able to be assessed. The result of this assessment showed which data quality indicators were present or absent in each standard.

The results showed that this novel approach was able to provide the assessor with a method of comparing the data quality requirements of the standards that were assessed. In addition, the presence or absence of the indicators provided a method of understanding the data quality of compliance datasets, allowing datasets with different compliance requirements to be compared with each other. The framework also showed that it is a useful method of benchmarking standards and can be used in the iterative development of new standards, to ensure the data quality requirements are improved from previous versions.

7.2 Future Work

An identified improvement noted in section 6.4 is to incorporate specific metadata elements into the framework. As a new set of drinking water standards for New Zealand is currently being written, this would be the first piece of subsequent work for this framework – to identify metadata specific to the current drinking water standards (the DWSNZ), and to use these to assess the new drinking water standards when they are published. This would provide a detailed understanding of the extent to which compliance datasets from each standard are comparable.

Following on from this application, the framework in its current state could also be tested on environmental monitoring standards not directly linked to the three waters industry. While this was considered by including greenhouse gas emission calculation standards, additional assessments relating to other industries would be beneficial. This would allow the selected data quality indicators to be evaluated to determine if their application is sufficiently broad; it could also lead to the development of additional data quality indicators, allowing further assessment of the standards reviewed in this project.

8 Appendix 1

Appendix 1 is included as a supplementary spreadsheet to this document. The contents of Appendix 1 are listed below. Each item is a tab in the spreadsheet.

- A1.1: Data Quality Framework
- Codes and Standards for A1.1
- A1.2: Summary Table 1 – Standards Grouped by Industry Type
- A1.3: Summary Table 2 - Standards Grouped by Area
- A1.4: Summary Table 3 - Data Quality Indicator Calculations

9 References

[1] S. Burton, ‘Data governance: The path to a data-driven culture’, *Applied Marketing Analytics*, vol. 6, no. 4, pp. 298–308, Jan. 2021.

[2] A. Ibrahim, I. Mohamed, and N. Safie, ‘Factors Influencing Master Data Quality: A Systematic Review’, *IJACSA*, vol. 12, no. 2, 2021, doi: 10.14569/IJACSA.2021.0120224.

[3] G. L. Kayser, U. Amjad, F. Dalcanale, J. Bartram, and M. E. Bentley, ‘Drinking water quality governance: A comparative case study of Brazil, Ecuador, and Malawi’, *Environmental Science & Policy*, vol. 48, pp. 186–195, Apr. 2015, doi: 10.1016/j.envsci.2014.12.019.

[4] G. Dunn, K. Bakker, and L. Harris, ‘Drinking Water Quality Guidelines across Canadian Provinces and Territories: Jurisdictional Variation in the Context of Decentralized Water Governance’, *IJERPH*, vol. 11, no. 5, pp. 4634–4651, Apr. 2014, doi: 10.3390/ijerph110504634.

[5] R. Y. Wang and D. M. Strong, ‘Beyond Accuracy: What Data Quality Means to Data Consumers’, *Journal of Management Information Systems*, vol. 12, no. 4, pp. 5–33, Mar. 1996, doi: 10.1080/07421222.1996.11518099.

[6] S. Canali, ‘Towards a Contextual Approach to Data Quality’, *Data*, vol. 5, no. 4, p. 90, Sep. 2020, doi: 10.3390/data5040090.

[7] L. L. Pipino, Y. W. Lee, and R. Y. Wang, ‘Data Quality Assessment’, *Communications of the ACM*, vol. 45, no. 4ve, pp. 211–218, 2002.

[8] E. Tute, I. Scheffner, and M. Marschollek, ‘A method for interoperable knowledge-based data quality assessment’, *BMC Med Inform Decis Mak*, vol. 21, no. 1, p. 93, Dec. 2021, doi: 10.1186/s12911-021-01458-1.

[9] S. Leonelli, ‘Global Data Quality Assessment and the Situated Nature of “Best” Research Practices in Biology’, *Data Science Journal*, vol. 16, p. 32, Jun. 2017, doi: 10.5334/dsj-2017-032.

[10] A. F. Karr, A. P. Sanil, J. Sacks, and A. Elmagarmid, ‘Workshop Report: Affiliates Workshop on Data Quality’. National Institute of Statistical Sciences, 2001. [Online]. Available: https://www.niss.org/sites/default/files/technicalreports/tr117.pdf

[11] I. Taleb, M. A. Serhani, C. Bouhaddioui, and R. Dssouli, ‘Big data quality framework: a holistic approach to continuous quality management’, *J Big Data*, vol. 8, no. 1, p. 76, Dec. 2021, doi: 10.1186/s40537-021-00468-0.

[12] S. Comero *et al.*, ‘A conceptual data quality framework for IPCHEM – The European Commission Information Platform for chemical monitoring’, *TrAC Trends in Analytical Chemistry*, vol. 127, p. 115879, Jun. 2020, doi: 10.1016/j.trac.2020.115879.

[13] N. Micic, D. Neagu, F. Campean, and E. H. Zadeh, ‘Towards a Data Quality Framework for Heterogeneous Data’, in *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, Exeter, Jun. 2017, pp. 155–162. doi: 10.1109/iThings-GreenCom-CPSCom-SmartData.2017.28.

[14] C. O. Schmidt *et al.*, ‘Facilitating harmonized data quality assessments. A data quality framework for observational health research data collections with software implementations in R’, *BMC Med Res Methodol*, vol. 21, no. 1, p. 63, Dec. 2021, doi: 10.1186/s12874-021-01252-7.

[15] L. Cai and Y. Zhu, ‘The Challenges of Data Quality and Data Quality Assessment in the Big Data Era’, *CODATA*, vol. 14, no. 0, p. 2, May 2015, doi: 10.5334/dsj-2015-002.

[16] M. D. Wilkinson *et al.*, ‘The FAIR Guiding Principles for scientific data management and stewardship’, *Sci Data*, vol. 3, no. 1, p. 160018, Dec. 2016, doi: 10.1038/sdata.2016.18.

[17] D. Kinkade and A. Shepherd, ‘Geoscience data publication: Practices and perspectives on enabling the FAIR guiding principles’, *Geosci. Data J.*, p. gdj3.120, May 2021, doi: 10.1002/gdj3.120.

[18] A. Jacobsen *et al.*, ‘A Generic Workflow for the Data FAIRification Process’, *Data Intellegence*, vol. 2, no. 1–2, pp. 56–65, Jan. 2020, doi: 10.1162/dint_a_00028.

[19] S. Jones *et al.*, ‘Data Management Planning: How Requirements and Solutions are Beginning to Converge’, *Data Intellegence*, vol. 2, no. 1–2, pp. 208–219, Jan. 2020, doi: 10.1162/dint_a_00043.

[20] A. Devaraju *et al.*, ‘From Conceptualization to Implementation: FAIR Assessment of Research Data Objects’, *Data Science Journal*, vol. 20, p. 4, Feb. 2021, doi: 10.5334/dsj-2021-004.

[21] E. Partescano *et al.*, ‘Data quality and FAIR principles applied to marine litter data in Europe’, *Marine Pollution Bulletin*, vol. 173, p. 112965, Dec. 2021, doi: 10.1016/j.marpolbul.2021.112965.

[22] N. S. Rajan, R. Gouripeddi, P. Mo, R. K. Madsen, and J. C. Facelli, ‘Towards a content agnostic computable knowledge repository for data quality assessment’, *Computer Methods and Programs in Biomedicine*, vol. 177, pp. 193–201, Aug. 2019, doi: 10.1016/j.cmpb.2019.05.017.

[23] M. Boeckhout, G. A. Zielhuis, and A. L. Bredenoord, 'The FAIR guiding principles for data stewardship: fair enough?', *Eur J Hum Genet*, vol. 26, no. 7, pp. 931–936, Jul. 2018, doi: 10.1038/s41431-018-0160-0.
[24] T. D. Houston and R. Hiederer, 'Applying quality assurance procedures to environmental monitoring data: a case study', *J. Environ. Monit.*, vol. 11, no. 4, p. 774, 2009, doi: 10.1039/b818274b.
[25] R. Y. Wang, 'A product perspective on total data quality management', *Communications of the ACM*, vol. 41, no. 2, pp. 58–63, 1998.
[26] S. T. Liaw *et al.*, 'Towards an ontology for data quality in integrated chronic disease management: A realist review of the literature', *International Journal of Medical Informatics*, vol. 82, no. 1, pp. 10–24, Jan. 2013, doi: 10.1016/j.ijmedinf.2012.10.001.
[27] R. Y. Wang, V. C. Storey, and C. P. Firth, 'A framework for analysis of data quality research', *IEEE Trans. Knowl. Data Eng.*, vol. 7, no. 4, pp. 623–640, Aug. 1995, doi: 10.1109/69.404034.
[28] R. Y. Wang, M. P. Reddy, and H. B. Kon, 'Toward quality data: An attribute-based approach', *Decision Support Systems*, vol. 13, no. 3–4, pp. 349–372, Mar. 1995, doi: 10.1016/0167-9236(93)E0050-N.
[29] C. Guien *et al.*, 'The French National Registry of patients with Facioscapulohumeral muscular dystrophy', *Orphanet J Rare Dis*, vol. 13, no. 1, p. 218, Dec. 2018, doi: 10.1186/s13023-018-0960-x.
[30] A. Trifan and J. L. Oliveira, 'Towards a More Reproducible Biomedical Research Environment: Endorsement and Adoption of the FAIR Principles', in *Biomedical Engineering Systems and Technologies*, vol. 1211, A. Roque, A. Tomczyk, E. De Maria, F. Putze, R. Moucek, A. Fred, and H. Gamboa, Eds. Cham: Springer International Publishing, 2020, pp. 453–470. doi: 10.1007/978-3-030-46970-2_22.
[31] W. Hasselbring, L. Carr, S. Hettrick, H. Packer, and T. Tiropanis, 'From FAIR research data toward FAIR and open research software', *it - Information Technology*, vol. 62, no. 1, pp. 39–47, Feb. 2020, doi: 10.1515/itit-2019-0040.
[32] R. D. Kush *et al.*, 'FAIR data sharing: The roles of common data elements and harmonization', *Journal of Biomedical Informatics*, vol. 107, p. 103421, Jul. 2020, doi: 10.1016/j.jbi.2020.103421.
[33] S.-T. Liaw *et al.*, 'Quality assessment of real-world data repositories across the data life cycle: A literature review', *Journal of the American Medical Informatics Association*, vol. 28, no. 7, pp. 1591–1599, Jul. 2021, doi: 10.1093/jamia/ocaa340.
[34] M. Musen, 'Making Data FAIR Requires More than Just Principles: We Need Knowledge Technologies', in *2019 15th International Conference on eScience (eScience)*, San Diego, CA, USA, Sep. 2019, pp. 530–532. doi: 10.1109/eScience.2019.00071.
[35] C. Bahim *et al.*, 'The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments', *Data Science Journal*, vol. 19, p. 41, Oct. 2020, doi: 10.5334/dsj-2020-041.
[36] R. Crystal-Ornelas *et al.*, 'A Guide to Using GitHub for Developing and Versioning Data Standards and Reporting Formats', *Earth Space Sci*, vol. 8, no. 8, Aug. 2021, doi: 10.1029/2021EA001797.
[37] C. Eder and A. Jedinger, 'FAIR national election studies: How well are we doing?', *Eur Polit Sci*, vol. 18, no. 4, pp. 651–668, Dec. 2019, doi: 10.1057/s41304-018-0194-3.
[38] J. Cyr-Gagnon and M. J. Rodriguez, 'Optimizing data management for municipal source water protection', *Land Use Policy*, vol. 100, p. 103788, Jan. 2021, doi: 10.1016/j.landusepol.2018.12.035.
[39] NHRMC, NRMMC, 'Australian Drinking Water Guidelines Paper 6 National Water Quality Management Strategy'. National Health and Medical Research Council, National Resource Management Ministerial Council, Commonwealth of Australia, Canberra, 2011.
[40] NSW Deparment of Health, 'Drinking Water Monitoring Program'. NSW Deparment of Health, 2011.
[41] DES, 'Monitoring and Sampling Manual: Environmental Protection (Water) Policy'. Brisbane: Department of Environment and Science Government, 2018.
[42] New Zealand and Ministry of Health, *Drinking-water standards for New Zealand 2005.* 2018.
[43] Ministry of Health, 'Annual Report on Drinking-water Quality 2019-2020'. Wellington: Ministry of Health, 2021.
[44] Ministry of Health, 'Supplementary Criteria for Accreditation: Ministry of Health Register of Water Testing Laboratories'. International Accreditation New Zealand, 2020.
[45] *Water Services Act 2021*. 2021. [Online]. Available: https://www.legislation.govt.nz/act/public/2021/0036/latest/LMS374564.html

[46] Taumata Arowai, ‘Drinking Water Quality Assurance Rules (Draft)’. Dec. 20, 2021. [Online]. Available: https://www.taumataarowai.govt.nz/for-water-suppliers/new-compliance-rules-and-standards/
[47] Taumata Arowai, ‘Drinking Water Standards for New Zealand (Draft)’. Dec. 20, 2021. [Online]. Available: https://www.taumataarowai.govt.nz/for-water-suppliers/new-compliance-rules-and-standards/
[48] *Resource Management Act 1991*. 1991. [Online]. Available: https://www.legislation.govt.nz/act/public/1991/0069/latest/DLM230265.html
[49] ‘Directive (EU) 2020/2184 of the European Parliament and of the Council of 16 December 2020 on the quality of water intended for human consumption’. OJ L 435/1, Dec. 23, 2020.
[50] ‘Directive 2006/118/EC of the European Parliament and of the Council of 12 December 2006 on the protection of groundwater against pollution and deterioration’. OJ L 372, Dec. 27, 2006.
[51] ‘Council Directive of 21 May 1991 concerning urban wastewater treatment (91/271/EEC) amended by Commission Directive 98/15/EC of 27 February 1998 (OJ L 67, 7.3.1998, pp.29), Regulation (EC) No 1882/2003 of the European Parliament and of the Council of 29 September 2003 (OJ L 284, 31.10.2003, pp.1), Regulation (EC) No 1137/2008 of the European Parliament and of the Council of 22 October 2008 (OJ L 311, 21.11.2008, pp.1), Council Directive 2013/64/EU of 17 December 2013 (OJ L 535, 28.12.2013, pp.8)’. OJ L 135, May 30, 1991.
[52] J. Andrews, ‘Carbon accounting guidelines for wastewater treatment: CH4 and N2O’. Water New Zealand, 2021. [Online]. Available: https://www.waternz.org.nz/Article?Action=View&Article_id=2078
[53] ‘Carbon Credits (Carbon Farming Initiative—Domestic, Commercial and Industrial Wastewater) Methodology Determination 2015’. Jan. 21, 2022. [Online]. Available: https://www.legislation.gov.au/Details/F2022C00227
[54] ‘Regulation (EU) 2018/1999 of the European Parliament and of the Council of 11 December 2018 on the Governance of the Energy Union and Climate Action, amending Regulations (EC) No 663/2009 and (EC) No 715/2009 of the European Parliament and of the Council, Directives 94/22/EC, 98/70/EC, 2009/31/EC, 2009/73/EC, 2010/31/EU, 2012/27/EU and 2013/30/EU of the European Parliament and of the Council, Council Directives 2009/119/EC and (EU) 2015/652 and repealing Regulation (EU) No 525/2013 of the European Parliament and of the Council’. OJ L 328/1, Dec. 21, 2018.
[55] G. Peng *et al.*, ‘Practical Application of a Data Stewardship Maturity Matrix for the NOAA *OneStop* Project’, *CODATA*, vol. 18, no. 1, p. 41, Aug. 2019, doi: 10.5334/dsj-2019-041.
[56] D. Knuplesch, M. Reichert, and A. Kumar, ‘A framework for visually monitoring business process compliance’, *Information Systems*, vol. 64, pp. 381–409, Mar. 2017, doi: 10.1016/j.is.2016.10.006.
[57] E. Ramezani, D. Fahland, J. M. van der Werf, and P. Mattheis, ‘Separating Compliance Management and Business Process Management’, in *Business Process Management Workshops*, vol. 100, F. Daniel, K. Barkaoui, and S. Dustdar, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 459–464. doi: 10.1007/978-3-642-28115-0_43.
[58] L. Gonzalez and A. Delgado, ‘Compliance Requirements Model for collaborative business process and evaluation with process mining’, in *2021 XLVII Latin American Computing Conference (CLEI)*, Cartago, Costa Rica, Oct. 2021, pp. 1–10. doi: 10.1109/CLEI53233.2021.9640197.
[59] A. Lucas, ‘Corporate Data Quality Management: Towards a Meta-Framework’, in *2011 International Conference on Management and Service Science*, Wuhan, China, Aug. 2011, pp. 1–6. doi: 10.1109/ICMSS.2011.5999365.
[60] P. Riebel, ‘Improving data quality within an environmental management system’, *Pulp & Paper Canada; Westmount*, vol. 100, no. 6, pp. 64–67, Jun. 1999.
[61] L. Miasayedava, K. McBride, and J. A. Tuhtan, ‘Automated environmental compliance monitoring of rivers with IoT and open government data’, *Journal of Environmental Management*, vol. 303, p. 114283, Feb. 2022, doi: 10.1016/j.jenvman.2021.114283.
[62] P. Myrseth, J. Stang, and V. Dalberg, ‘A data quality framework applied to e-government metadata: A prerequsite to establish governance of interoperable e-services’, in *2011 International Conference on E-Business and E-Government (ICEE)*, Shanghai, China, May 2011, pp. 1–4. doi: 10.1109/ICEBEG.2011.5881298.
[63] R. Khan, D. Phillips, D. Fernando, J. Fowles, and R. Lea, ‘Environmental Health Indicators in New Zealand: Drinking Water—A Case Study’, *EcoHealth*, vol. 4, no. 1, pp. 63–71, Apr. 2007, doi: 10.1007/s10393-007-0089-1.
[64] K. Kerr, ‘The Development of a Data Quality Framework and Strategy for the New Zealand Ministry of Health’, Jan. 2000.

[65] B. Otto, K. Wende, A. Schmidt, and P. Osl, 'Towards a Framework for Corporate Data Quality Management', *Institute of Information Management*, 2007.
[66] '9223 Enzyme Substrate Coliform Test'. Standard Methods For the Examination of Water and Wastewater, 23rd, Jan. 2018.
[67] *Climate Change (Stationary Energy and Industrial Processes) Regulations 2009*. 2009. [Online]. Available: https://www.legislation.govt.nz/regulation/public/2009/0285/latest/DLM2394207.html
[68] *Resource Management (Measurement and Reporting of Water Takes) Regulations 2010*. 2010. [Online]. Available: https://www.legislation.govt.nz/regulation/public/2010/0267/latest/DLM3174201.html
[69] *Resource Management (National Environmental Standards for Sources of Human Drinking Water) Regulations 2007*. 2007. [Online]. Available: https://www.legislation.govt.nz/regulation/public/2007/0396/latest/DLM1106901.html
[70] Environment Canterbury, 'Resource Consent CRC051724'. May 16, 2006. [Online]. Available: https://www.ecan.govt.nz/data/consent-search/consentdetails/CRC051724/CRC051724
[71] Environment Canterbury, 'Resource Consent CRC030999.1'. Oct. 18, 2004. [Online]. Available: https://www.ecan.govt.nz/data/consent-search/consentdetails/CRC030999.1/CRC030999.1
[72] *Climate Change Response Act 2002*. 2002. [Online]. Available: https://www.legislation.govt.nz/act/public/2002/0040/latest/DLM158584.html
[73] 'Method 3A - Determination of Oxygen and Carbon Dioxide Concentrations in Emissions from Stationary Sources (Instrumental Analyzer Procedure)'. United States Environmental Protection Agency, Aug. 03, 2017.
[74] 'Method 7E - Determination of Nitrogen Oxides Emissions from Stationary Sources (Instrumental Analyzer Procedure)'. United States Environmental Protection Agency, Oct. 07, 2020.
[75] NSW Deparment of Health, 'Guide for Submitting Water Samples to the Division of Analytical Laboratories for Analysis'. Sydney West Area Health Service, Jun. 2010.
[76] National Association of Testing Authorities, 'General Accreditation Criteria: Equipment assurance, in-house calibration and equipment verification'. May 2019.
[77] National Association of Testing Authorities, 'General Accreditation Criteria: Proficiency Testing Policy'. Dec. 2021.
[78] 'National Greenhouse and Energy Reporting (Measurement) Determination 2008'. Jul. 01, 2021. [Online]. Available: https://www.legislation.gov.au/Details/F2021C00740
[79] 'Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)'. OJ L 108, Apr. 25, 2007.
[80] 'Directive 20003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information and repealing Council Directive 90/313/EEC'. OJ L 41/26, Feb. 14, 2003.
[81] 'Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy'. OJ L 327/1, Dec. 22, 2000.
[82] J. Goodwin, M. Woodfield, M. Ibnoaf, M. Koch, and H. Yan, 'Chapter 2: Approaches to Data Collection', in *2006 IPCC Guidelines for National Greenhouse Gas Inventories*, vol. 1: General Guidance and Reporting, IGES, Japan: Prepared by the National Greenhouse Gas Inventories Programme, 2006.
[83] M. R. J. Doorn *et al.*, 'Chapter 6: Wastewater Treatment and Discharge', in *2006 IPCC Guidelines for National Greenhouse Gas Inventories*, vol. 5: Waste, IGES, Japan: Prepared by the National Greenhouse Gas Inventories Programme, 2006.
[84] M. J. S. Sánchez, S. Bhattacharya, and K. Mareckova, 'Chapter 8: Reporting Guidance and Tables', in *2006 IPCC Guidelines for National Greenhouse Gas Inventories*, vol. 1: General Guidance and Reporting, IGES, Japan: Prepared by the National Greenhouse Gas Inventories Programme, 2006.