# NHS Scotland Data Quality Framework for Climate and Sustainability

# 1. Contents


1. Introduction 3
2. What is included in this data quality framework? 4
3. What is meant by ‘good quality’ sustainability data? 5
4. Why does sustainability data quality matter in practice? 6
5. Who is this sustainability data quality framework for? 7
6. Data quality principles 8
   - Commit to data quality 8
   - Know your users and their needs 9
   - Assess quality throughout the data lifecycle 10
   - Communicate data quality clearly and effectively 11
   - Anticipate changes affecting data quality 12
7. Data quality across the data lifecycle 13
8. Sustainability data quality checklists 21
9. Data quality dimensions 22
   - Accuracy 22
   - Validity 23
   - Completeness 24
   - Consistency 25
   - Relevance 26
   - Timeliness 27
   - Uniqueness 28
10. Role-based guidance - how to get started 29
11. Glossary 31
12. Appendices 32
    - Appendix A - Sustainability Data Quality Assessment 32
    - Appendix B - Sustainability Data Quality Action Plan 33
    - Appendix C - Root Cause Analysis 34

# 1. Introduction

## What is the data quality framework for?

The Sustainability Data Quality Framework provides practical guidance on what good quality sustainability data looks like and how to manage it across the full data lifecycle.

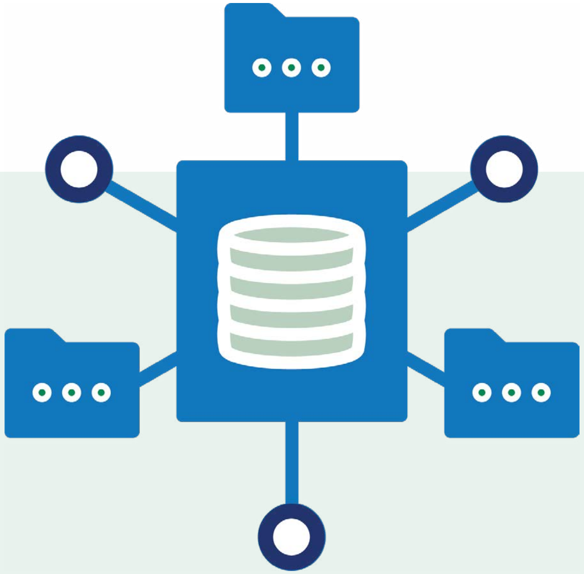
It focuses on:

- Shared principles and expectations for data quality
- What ‘fit for purpose’ means in different contexts
- Common data quality dimensions (such as accuracy, completeness and timeliness)
- Practical checks, examples and role-based guidance

## Use this document if you want to…

- Improve the quality of sustainability data you work with
- Understand what checks are proportionate and why
- Apply consistent approaches in day-to-day data handling, analysis or reporting

This framework can be used now and does not depend on the implementation of recommendations in the Data Quality Strategy for Climate and Sustainability.



# 2. What is included in this data quality framework?

This Sustainability Data Quality Framework sets out the **core components needed for boards to manage the quality of sustainability data** across NHS Scotland.

Sustainability **data is a strategic asset** to NHS Scotland and having **data with poor or unknown quality impacts the organisation** and **hinders ability to become a greener, more sustainable NHS.**

**This Sustainability Data Quality Framework covers:**

- **Data Quality Principles** that describe behaviours and expectations for managing sustainability data
- **The Sustainability Data Lifecycle**, showing how quality should be considered at each stage, from planning what data is required and how it will be designed and collected, through to reporting and archiving
- **Data Quality Dimensions** that explain how to assess whether data is appropriate and reliable for its intended purpose
- **Typical User Group** examples to illustrate how different colleagues interact with sustainability data across NHS Scotland
- A high-level **Approach to Assessing and Improving Data Quality**, including an explanation of Data Quality Assessments and Data Quality Action Plans, supported by simple templates

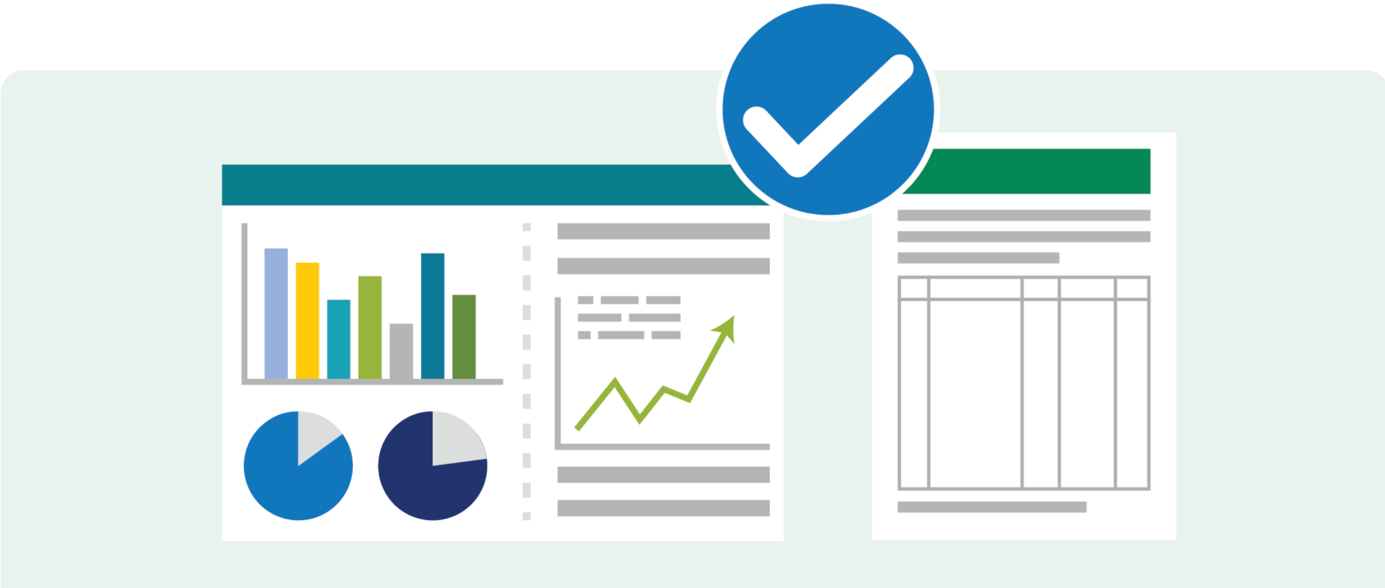
# 3. What is meant by 'good quality' sustainability data?

**'Good quality sustainability data'** is data that is **fit for purpose**.

It doesn't have to be perfect, but it must be **reliable enough to support the decisions, actions or reporting it informs**.

In practice, this means the sustainability data should be:

- **Complete enough to answer the stated question** (for example, all expected meter readings are present for every meter in the period)
- **Valid and accurate** against agreed formats, ranges, and units (for example, data recorded in kWh rather than $m^3$ or in kg rather than tonnes)
- **Consistent** across sources and over time (for example, the site code used consistently refers to the same site)
- **Timely** for the decision-making or reporting cycle (for example, monthly data is available by the 10th working day)
- **Traceable** with clear lineage, source information and version control (for example, the emission factor version is logged)
- **Transparent about limitations** with clear communication of constraints, assumptions, data gaps or risks to inform users of the data (for example, identifying and communicating where gas data has been estimated due to supplier delays)



# 4. Why does sustainability data quality matter in practice?

Sustainability data underpins every decision NHS Scotland makes about its environmental impact. It supports effective decision-making, credible reporting and progress toward net-zero commitments.

This framework supports the Sustainability Data Quality Strategy by providing practical guidance on how sustainability data quality should be managed in day-to-day work.

Data quality is about ensuring that the data is reliable and suitable for its intended purpose. The level of quality needed depends on the decision being made using the data. However, without consistent standards and processes, sustainability data varies between teams, systems and suppliers.

A unified approach to managing sustainability data quality is needed across NHS Scotland where much of the data relies on manual collection, inconsistent templates and local practices. This framework provides a clear and practical way forward and sets out the principles and expectations needed to build a culture of quality, reduction in duplication and targets improvement where it matters most.

By adopting this framework, NHS Scotland can deliver more reliable insights, stronger evidence and better outcomes for people, finances and the planet.

# 5. Who is this sustainability data quality framework for?

This Data Quality Framework provides **guidance for all individuals involved in the creation, handling, management or use of sustainability data** across its full lifecycle.

It supports the consistent protection and maintenance of data quality from **initial creation and transformation through to sharing, analysis and decision making**.

It supports the people and processes that:

- Generate raw data
- Prepare, transform and refine data
- Produce insights, reporting or evidence
- Make operational decisions
- Develop policies

The framework will help to ensure that data is:

- Captured accurately at the source
- Maintained consistently as it flows between systems
- Interpreted correctly when used
- Trustworthy, comparable and fit-for-purpose for those that rely on it

# 6. Data quality principles

The data quality principles (or behaviours and expectations) required to manage the quality of sustainability data are:

## Commit to data quality

**Create a sense of accountability for data quality for those that work with sustainability data and make a commitment to the ongoing assessment, improvement and reporting of data quality.**

### Ensure effective data management and governance:

- Ensure that accountability for data quality is present at all levels:
  - Data leaders should guide an organisation or team's strategic direction by ensuring awareness and improvement of data quality.
  - Data practitioners should ensure that measuring, communicating and improving data quality is at the forefront of activities relating to sustainability data.

### Build data quality capability:

- Dedicate time and resource to building capability in assessing, improving and communicating data quality through training and sharing best practice.
- Include best practice in data quality management (such as **9. Data Quality Dimensions**) as part of training materials.

### Focus on continuous improvement through monitoring data quality to identify and define where efforts should be prioritised:

- Benchmark and regularly assess levels of data quality over time to track changes in quality.
- Prioritise and iterate effective improvements to achieve fit for purpose data.
- Use **Data Quality Assessments** and **Data Quality Action Plans** to identify and define where efforts should be prioritised.

## Know your users and their needs

**Understanding user needs is essential to ensuring that sustainability data is fit for purpose. Data quality should be shaped by how the data will be used, the decisions it informs, and the risks associated with poor quality data.**

**Taking time to understand users and their needs helps teams prioritise effort on the data that matters most, rather than attempting to improve everything equally.**

### Research your users and understand their quality needs:

- Identify the different user groups that rely on sustainability data, such as operational teams, data analysts, policy leads, sustainability leads and senior decision makers.
- Understand how the data is used, including the decisions it supports, reporting requirements and who relies on the data next.
- Recognise that different users may have different quality expectations for the same dataset (for example, completeness for reporting versus timeliness for operational monitoring).
- Use user needs to define what ‘fit for purpose’ means and which data quality dimensions are most critical.
- Prioritise data quality effort where poor quality would have the greatest impact on outcomes, confidence, or decision making.
- Review user needs periodically to reflect changes in policy, reporting requirements, systems, or sustainability priorities.

# Assess quality throughout the data lifecycle

**Data should be managed across its lifecycle, paying close attention to quality measures and assurance at each stage.**

## Assess data quality across the lifecycle:

- Consider data quality at each stage of the data lifecycle, from initial planning through to archive or destruction.
- Avoid relying solely on end-stage checks by embedding quality considerations as early as possible.
- Apply proportionate checks at points where data is created, transformed, integrated or reused.
- Ensure quality issues identified at later stages are traced back to earlier lifecycle stages to identify and resolve root causes rather than simply treating symptoms.

## Communicate with users and stakeholders across the lifecycle:

- Develop effective communication channels with and between stakeholders to ensure a broad understanding of data quality.
- Communicate any changes in data quality to sustainability data users at all stages of the lifecycle.
- Proactively engage with data providers to ensure a clear understanding of data quality requirements.

# Communicate data quality clearly and effectively

**Communicate quality to users regularly and clearly to ensure data is used appropriately.**

## Communicate data quality to users:

- Provide clear data quality information and describe its impact on use of the data.
- Communicate trade-offs in data quality clearly to aid understanding of the data’s strengths and weaknesses.
- Be transparent about the quality assurance approach taken and communicate data quality issues clearly to users.
- Build strong relationships with suppliers of external data to identify data quality problems at source.
- Inform users in advance about changes made to data processes which could impact on quality.
- Communicate clearly and in plain language, following relevant style guides for published materials.
- Provide clear definitions of terminology used and not presume a high-level of user understanding of data quality.

## Provide effective documentation and metadata:

- Document and share metadata to minimise ambiguity and enhance opportunity for data access and reuse.
- Document and report data quality issues and be transparent about steps being taken to address them.

## Anticipate changes affecting data quality

**Not all future problems can be predicted. Where possible, anticipate and prevent future data quality issues through good communication, effective management of change and addressing quality issues at source.**

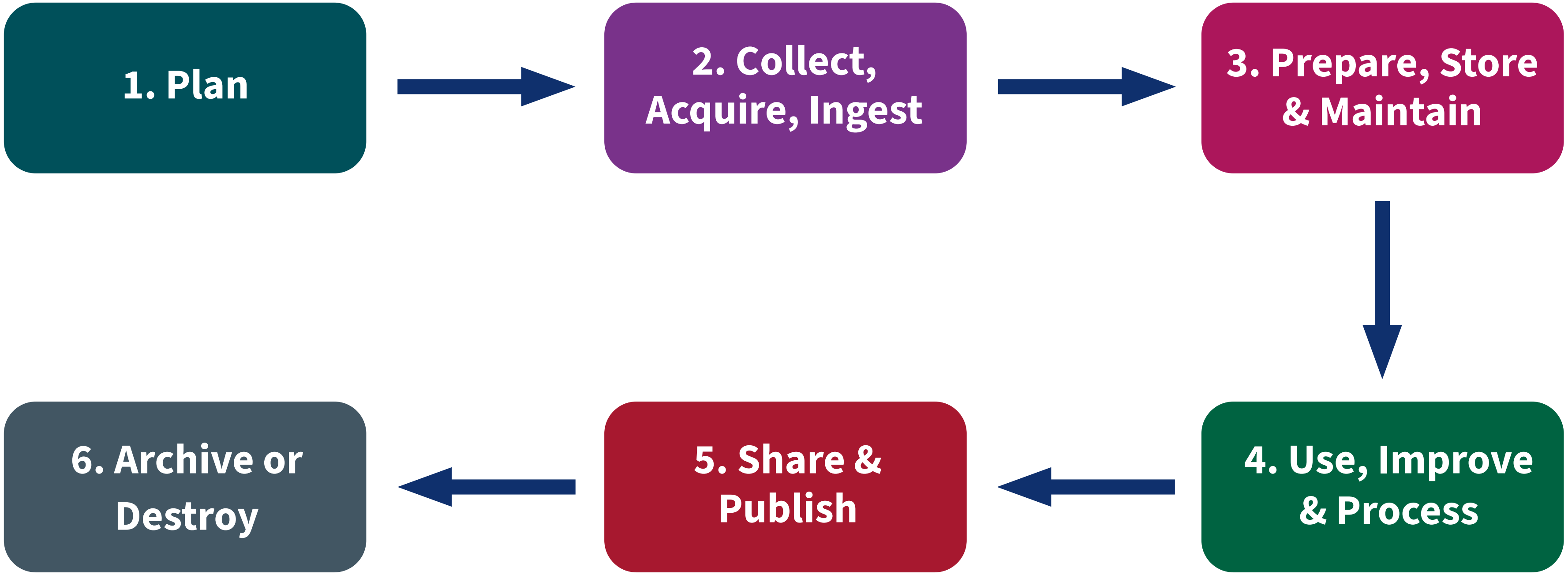
### Plan for the future:

- Use **root cause analysis** to solve data quality issues at source, rather than apply temporary fixes.
- Regularly communicate with users to keep up with changing data and quality requirements.
- Proactively consider the impact of changes in systems on data quality.
- Integrate quality processes into the design of new data systems.
- Ensure metadata and other supporting documentation is thorough and up to date.

# 7. Data quality across the data lifecycle

**Sustainability data moves through a few stages over time, from planning and collection through to use, sharing, and eventual archiving or destruction.**

**The diagram below illustrates these key stages and provides a common reference point for understanding where data quality needs to be considered.**



# The data lifecycle explained

## 1. Plan

The Plan Stage is where an organisation or team intending to collect, store and use data must plan their processes and data storage.

The planning stage involves determining business needs, identifying what data exists already and what needs to be collected or acquired. It also involves designing how this data will be collected and managed.

---

### Typical Activities

- Define the purpose, source and scope of the data
- Confirm units, identifiers, collection cadence
- Identify validation rules
- Agree data governance roles

### Potential Data Quality Problems

- Poor design of data collection
- Lack of validation rules
- Failure to specify use of master or reference data
- Lack of data standards
- No data owner
- Undefined units and identifiers
- No timeline
- Design does not consider 'upstream' data use

### Suggested Considerations

- Have business needs for data been clearly defined?
- Is there an inventory/catalogue of existing data and identification of new data requirements?
- Are data collection and management processes designed with quality in mind?
- Are data standards and validation rules specified?
- Is the use of master or reference data considered?
- Does the design account for upstream and downstream data use?

### What 'good' looks like

Good planning means:

- The purpose and intended use of the data are clearly documented and agreed
- Data requirements (including units, frequency, and identifiers) are defined upfront
- Ownership and governance responsibilities for the data are clear
- Known dependencies on other data sources or downstream uses are understood
- Data quality expectations are considered as part of design, not added later

## 2. Collect, acquire, ingest

During the collect, acquire and ingest stage, an organisation or team will acquire data based on user needs. They can improve the quality at source through validation rules and capturing appropriate metadata.

### Typical Activities

- Capture meter reads
- Receive supplier files
- Collect waste weights
- Retrieve travel data

### Potential Data Quality Problems

- Errors in manual data entry
- Incomplete data
- Duplicated data
- Inconsistent formats
- Insufficient information about the data that has been received (for example from producers of administrative data)
- Insufficient or poor-quality metadata
- A failure to ensure effective feedback loops between users of data and the collection/acquisition of data (particularly where it is outsourced to third parties)

### Suggested Considerations

- Is data being captured or received in the expected format, units, and structure?
- Are validation checks applied at the point of entry or receipt, and are missing, duplicate or inconsistent records identified and resolved?
- Is sufficient metadata captured (for example, source, date received, method of collection) to support understanding and traceability?
- Are there clear processes and responsibilities for monitoring and resolving data quality issues during collection or ingestion?
- Are effective feedback loops in place with data providers or suppliers to address issues at source?

### What 'good' looks like

Good collection and ingestion means:

- Data is received in the expected format, units, and structure
- Basic validation checks are applied at source to catch obvious issues early
- Metadata is captured to explain where the data came from and how it was produced
- Duplicate or incomplete submissions are identified and resolved promptly
- Clear feedback routes exist with data providers or suppliers

## 3. Prepare, store & maintain

At this stage data is prepared for storage, formatted for use at further stages in the data lifecycle and maintained for use within the organisation.

Consistent standards should be applied to the data and where necessary, the data should be anonymised. Where possible, data should also be cleaned and linked with other records in organisational data stores. This can help to reduce quality problems such as duplication and issues of consistency.

Data may then be integrated into the organisational data stores. Practitioners ensure the data is stored appropriately and provide the access necessary to business users. Any data that is subject to change should be regularly monitored for its data quality to ensure it continues to be fit for purpose.

### Typical Activities

- Cleanse data
- Standardise units
- Deduplicate
- Link datasets
- Store with metadata

### Potential Data Quality Problems

- Lack of adequate data preparation
- Inaccuracies and corruption resulting from integration of data sources and feeds
- Lack of informative metadata
- Lack of documentation
- Incorrect data linking
- Inconsistent standards applied to the data, which can lead to problems when linking
- Data accuracy decaying over time
- System changes causing inconsistencies
- Data not being actively managed

### Suggested Considerations

- Is data prepared and formatted according to standards?
- Is data cleaned, anonymized and linked where appropriate?
- Are consistency checks performed across data sets?
- Are metadata and documentation maintained and updated?
- Is data actively managed and monitored for accuracy over time?
- Are system changes tracked for potential impact on data quality?

### What 'good' looks like

Good preparation and maintenance means:

- Data is stored in a consistent, standardised format
- Units, identifiers, and reference data are applied consistently
- Metadata and documentation are kept up to date alongside the data
- Data quality is monitored over time to identify deterioration or change
- System or process changes are assessed for impact on data quality

## 4. Use, improve & process

At this stage of the data life cycle, data is processed and used for the specified business needs. This may involve exploration and analysis of the data, as well as production of outputs.

### Typical Activities

- Calculate carbon dioxide equivalent
- Analyse trends
- Produce dashboards
- Create KPIs

### Potential Data Quality Problems

- Lack of adequate data preparation
- Inaccuracies and corruption resulting from integration of data sources and feeds
- Lack of informative metadata
- Lack of documentation
- Incorrect data linking
- Inconsistent standards applied to the data, which can lead to problems when linking
- Data accuracy declining over time
- System changes causing inconsistencies
- Data not being actively managed

### Suggested Considerations

- Is data prepared and formatted according to standards?
- Is data cleaned, anonymized and linked where appropriate?
- Are consistency checks performed across data sets?
- Are metadata and documentation maintained and updated?
- Is data actively managed and monitored for accuracy over time?
- Are system changes tracked for potential impact on data quality?

### What 'good' looks like

Good use and processing means:

- Data is fit for the stated analytical or reporting purpose
- Assumptions, estimates, and transformations are documented
- Known data quality issues are understood by those using the data
- Results can be traced back to source data if needed
- Outputs are reviewed to ensure they align with underlying data quality

## 5. Share & publish

Data is shared where it is appropriate for processing for secondary purposes. Where the data is suitable for publication, data should be quality assured, anonymised and made available with appropriate documentation including details on its quality.

Open data published by public authorities should be released in consistent and accessible formats, to improve its utility.

### Typical Activities

- Distribute reports
- Publish open data where appropriate
- Submit returns

### Potential Data Quality Problems

- Failure to adhere to organisational data management practices and principles
- Failure to identify and log errors
- Failure to understand and address known quality issues
- Failure to carry out risk-based assessment on whether to use data because of poor understanding of data quality
- Human error in manual production of analysis and outputs
- Misinterpretation by downstream users
- Missing caveats
- Poor accessibility

### Suggested Considerations

- Are data principles being followed?
- Are errors identified, logged and addressed?
- Are known quality issues documented and communicated?
- Is risk-based assessment performed before using data?
- Are outputs reviewed for human error?

### What 'good' looks like

Good sharing and publication means:

- Data is shared with appropriate context, caveats, and quality information
- Known limitations are clearly communicated to users
- Data is published in accessible and consistent formats where appropriate
- Quality risks are assessed before reuse or publication
- Users are supported to interpret the data correctly

## 6. Archive & destroy

Once data is no longer in active use the data owner should determine whether it should be archived (available and secure) or destroyed. Information about the quality should be stored with the data.

### Typical Activities

- Retain data with context (such as sufficient metadata, documentation, data quality information and known limitations) to allow users to correctly interpret, assess and reuse the data across its lifecycle
- Decommission data when no longer required

### Potential Data Quality Problems

- Integrity of data compromised by changes made to it after it is archived
- Loss of organisational knowledge about the data and its quality
- Keeping data longer than needed

### Suggested Considerations

- Is the decision to archive or destroy data documented?
- Is information about data quality stored with archived data?
- Are controls in place to prevent changes to archived data?
- Is organisational knowledge about archived data maintained?

### What 'good' looks like

Good archiving and destruction means:

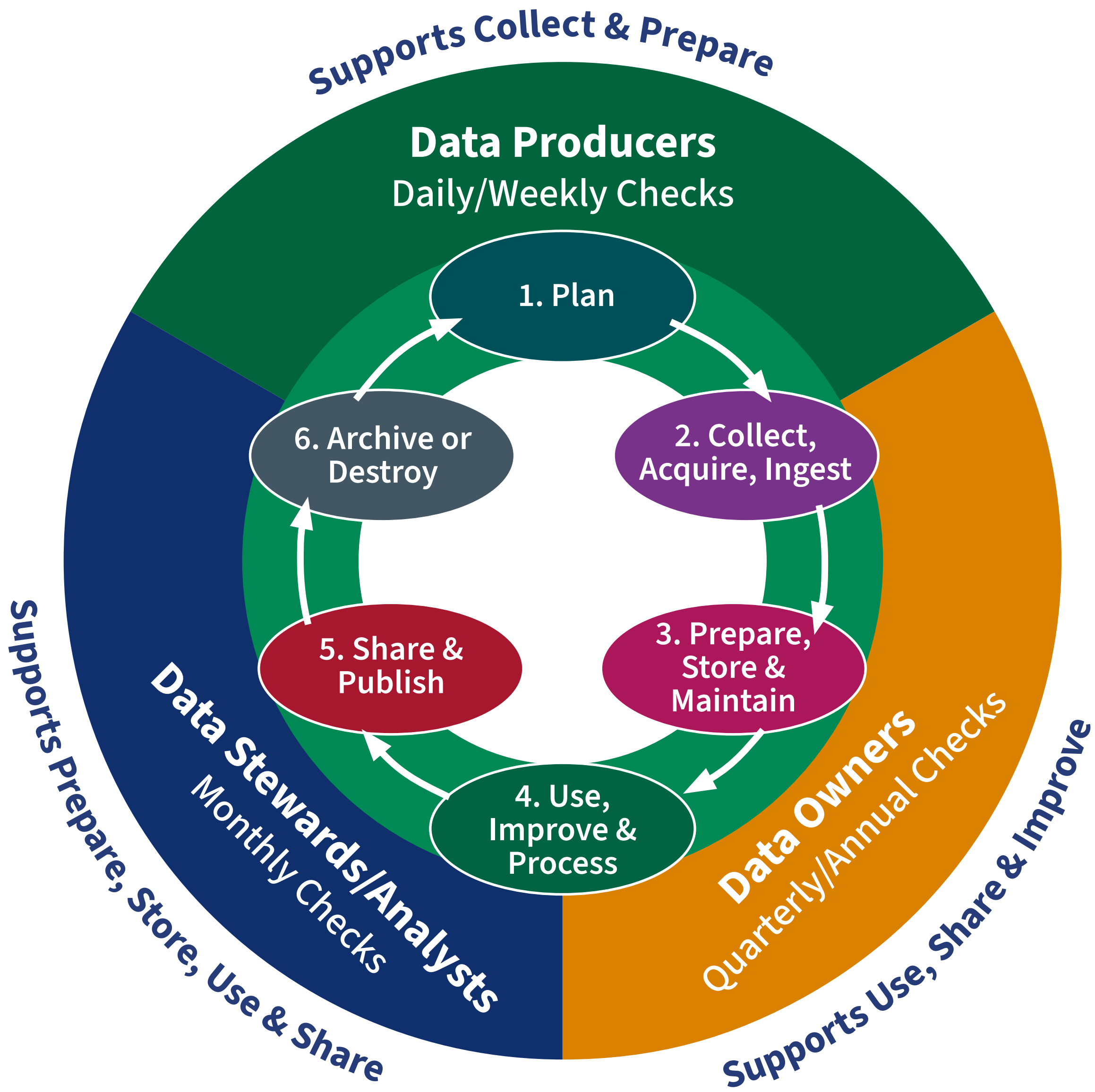
- Decisions to archive or destroy data are documented and justified
- Archived data retains sufficient context, metadata, and quality information
- Controls prevent unintended changes to archived data
- Organisational knowledge about the data and its quality is preserved
- Data is not retained longer than necessary

## Data quality checks throughout the data lifecycle

Sustainability data quality is supported through proportionate checks applied at different points in the data lifecycle.

Responsibilities and frequency vary by data role and risk and are intended as prompts rather than procedures.

The diagram highlights where different roles contribute most to data quality across the lifecycle, aligning to formal data governance roles. It does not imply exclusive ownership of any lifecycle stage, or that data quality is only considered at these points. The cadence shown reflects when different roles typically apply quality checks, rather than a prescribed sequence of activities through the lifecycle.



## 8. Sustainability data quality checklists

The sustainability data quality checklists below provide practical reminders that support the data lifecycle and the role-based responsibilities illustrated in the above section. They are intended to help teams apply proportionate data quality checks in day-to-day work, rather than act as exhaustive or mandatory procedures. They are intended to be used flexibly, reflecting the importance and risk of the data involved.

### Daily/weekly (Data producers)

- Correct template used.
- Units and date formats correct.
- All mandatory fields complete (or reason code).
- Obvious outliers investigated or flagged.

### Monthly (Data stewards/analysts)

- Coverage check: all sites/streams present.
- Duplicate scan run and resolved.
- Emission factor version confirmed.
- Reconciliation with prior period within tolerance.
- Data Quality Note updated.

### Quarterly/annually (Data owners)

- Review thresholds and SLAs.
- Confirm reference data (sites/meters) current.
- Sample audit against source (bills, tickets).
- Review open issues and close root cause actions.

# 9. Data quality dimensions

The data quality dimensions provide a shared way of describing and understanding data quality issues once they have been identified.

They are not intended to be applied universally or measured for every dataset. Instead, relevant dimensions should be selected based on the purpose of the data and how the data is used.

The dimensions support data quality assessments, action planning and clear communication of data quality to users.

The data quality dimensions are:

## Accuracy

**Data will be accurate and free from errors. Procedures will be in place to validate and correct data as needed to maintain its accuracy.**

**Definition**
The degree to which data correctly describes the 'real world' object or event being described.

**Reference**
Ideally the 'real world' truth is established through primary research. However, as this is often not practical, it is common to use 3rd party reference data from sources which are deemed trustworthy and of the same chronology.

**Measure**
The degree to which the data mirrors the characteristics of the 'real world' object or objects it represents.

**Scope**
Any 'real world' object or objects that may be characterised or described by data, held as data item, record, data set or database.

**Unit of Measure**
The percentage of data entries that pass the data accuracy rules.

**Related Dimension(s)**
Validity is a related dimension because, to be accurate, values must be valid, the right value and in the correct representation.

**Optionality**
Mandatory (as inaccurate data may not be fit for use).

**Suggested Considerations**

- Are there data accuracy checks in place?
- Are they manual or automated?
- Are they supported by a documented process?
- Where is this held?

**Example Data Quality issue**

- Gas consumption recorded with a questionably high value due to an incorrect meter reading.
- Electricity usage is entered against the wrong site, even though the value itself is realistic.

# Validity

**Data will be valid, conforming to defined formats, types and ranges. Procedures will be in place throughout the data lifecycle to ensure validity.**

### Definition
Data are valid if it conforms to the syntax (format, type, range) of its definition.

### Reference
Database, metadata or documentation rules as to the allowable types (string, integer, floating point etc.), the format (length, number of digits etc.) and range (minimum, maximum or contained within a set of allowable values).

### Measure
Comparison between the data and the metadata or documentation for the data item.

### Scope
All data can typically be measured for Validity. Validity applies at the data item level and record level (for combinations of valid values).

### Unit of Measure
Percentage of data items deemed Valid to Invalid.

### Related Dimension(s)
- Accuracy
- Completeness
- Consistency
- Uniqueness

### Optionality
Mandatory (invalid data cannot be meaningfully processed).

### Suggested Considerations
- Are there data validity checks in place?
- Are they manual or automated?
- Are they supported by a documented process?
- Where is this held?

### Example Data Quality issue
- Electricity consumption submitted with negative values where only positive values are valid.
- A reporting period recorded as ‘2025 13’ where the month value is invalid.

# Completeness

**Data will be complete, ensuring that all required information is available. Data validation checks will be conducted to identify and address missing or incomplete data.**

### Definition
The proportion of stored data against the potential of '100% complete'.

### Reference
Business rules which define what '100% complete' represents.

### Measure
A measure of the absence of blank (null or empty string) values or the presence of non-blank values.

### Scope
0-100% of critical data to be measured in any data item, record, data set or database.

### Unit of Measure
Percentage

### Related Dimension(s)
- Validity
- Accuracy

### Optionality
Conditional:
- If a data item is mandatory - 100% completeness should be achieved
- If a data item is considered a critical data asset - 100% completeness should be aimed for
- Validity and Accuracy checks would need to be performed to determine if the data item has been completed correctly
- Incompleteness in non-critical data may not matter to the business.

### Suggested Considerations
- Are there data validity checks in place?
- Are they manual or automated?
- Are they supported by a documented process?
- Where is this held?

### Example Data Quality issue
- Missing monthly water consumption data for one or more sites.
- Waste data submitted without a waste stream classification.

# Consistency

**Data will be consistent across systems and processes, ensuring that it presents a coherent and unified view. Data integration and reconciliation processes will be employed to maintain consistency.**

### Definition
The absence of difference, when comparing two or more representations of a thing against a definition.

### Reference
Data item measured against itself or its counterpart in another data set or database.

### Measure
Analysis of pattern and/or value frequency.

### Scope
Assessment of things across multiple data sets and/or assessment of values or formats across data items, records, data sets and databases. Processes including people based, automated, electronic or paper.

### Unit of Measure
Percentage

### Related Dimension(s)
- Validity
- Accuracy
- Uniqueness

### Optionality
Conditional (dependent upon whether data is stored across multiple systems and processes).

### Suggested Considerations
- Is the data integrated across systems?
- How is the data integrated?
- Are consistency checks in place to monitor and measure data across systems?
- Are they manual or automated?
- Are they supported by a documented process?
- Where is this held?

### Example Data Quality issue
- The same site reported under different site codes in energy, waste and travel datasets.
- F-Gas usage recorded differently between the contractor service notes and the F-Gas register for the same period.

# Relevance

**Data will be relevant to the intended business purpose. Data stewards will regularly review data to ensure continued relevance and alignment with organisational goals.**

### Definition
The degree to which the data provides an insight into to the problem or meets the intended business purpose.

### Reference
The predefined criteria, business rules, or contextual standards used to assess whether data is meaningful, appropriate and aligned with the intended purpose or decision-making need.

### Measure
- Data items measured against the predefined criteria, business rules, or contextual standards.
- Data Usage statistics.
- Feedback from users.

### Scope
Any data item, record, data set or database.

### Unit of Measure
Percentage, Number of views of data, CSAT.

### Related Dimension(s)
- Accuracy
- Validity
- Completeness
- Consistency
- Timeliness
- Uniqueness

### Optionality
Conditional (dependent on the purpose and intended use of the data).

### Suggested Considerations
- Are regular reviews in place to capture data relevance?
- Are they supported by a documented process?
- Where is this held?

### Example Data Quality issue
- Sustainability data collected at a level of detail that does not meet the needs of its intended reporting or decision-making purpose.
- Data continues to be collected using categories or structures that do not align with current sustainability reporting requirements.

# Timeliness

**Data will be timely, meeting the needs of users and business processes. Data will be updated, archived or deleted as required to maintain its timeliness.**

### Definition
The degree to which data represent reality from the required point in time.

### Reference
The time the real-world event being recorded occurred.

### Measure
Time difference

### Scope
Any data item, record, data set or database.

### Unit of Measure
Time

### Related Dimension(s)
Accuracy (as it inevitably decays with time).

### Optionality
Conditional (dependent upon the needs of the business).

### Suggested Considerations
- Are regular reviews in place to capture stakeholder needs around data timeliness?
- Are they supported by a documented process?
- Where is this held?

### Example Data Quality issue
- Energy data received after the reporting deadline, requiring estimation.
- Sustainability data not available in time to support planned review meetings or decision points.

# Uniqueness

**Data will be unique, preventing duplicate or redundant datasets. Data processes will be in place to maintain uniqueness.**

### Definition
No thing will be recorded more than once based upon how that thing is identified.

### Reference
Data item measured against itself or its counterpart in another data set or database.

### Measure
Analysis of the number of things as assessed in the 'real world' compared to the number of records of things in the data set. The 'real world' number of things could be either determined from a different and perhaps more reliable data set or a relevant external comparator.

### Scope
Measured against all records within a single data set.

### Unit of Measure
Percentage

### Related Dimension(s)
Consistency

### Optionality
Conditional (dependent on circumstances/context).

### Suggested Considerations
- Do you have quality controls to manage duplicate or redundant datasets?
- Are they supported by a documented process?
- Where is this held?

### Example Data Quality issue
- Duplicate records created for the same emission source because different identifiers are used in different systems.
- The same reporting period submitted twice following corrections, resulting in double counting.

# 10. Role-based guidance - how to get started

Everyone involved in sustainability data plays a role in its quality, whether by creating, handling, analysing or using data.

The guidance below is intended to help individuals understand how they can support good quality data in a practical and proportionate way. You may recognise yourself in more than one role, depending on the task or timing.

Pick the option(s) that best match your data responsibilities:

## Data producer

Data producers are often best placed to spot issues early, before they affect reporting or decision making:

- If templates have been agreed, ensure they are used and do not add your own columns or formats
- Carry out simple checks before submitting data:
  - Units of measure are correct (for example, kWh / $m^3$ / kg / miles)
  - Dates or reporting periods are correct
  - All mandatory fields are populated
- If a value looks wrong (for example, an unexpected jump), flag it early rather than correcting it in isolation

## Data Steward

Data stewards ensure sustainability data remains consistent, understandable, and fit for purpose as it moves between systems, teams and reporting outputs.

### Manage and track data sources

- Create and maintain a Source Register, this is a simple record of where sustainability data comes from and how it should be handled, as a minimum, record key information such as:
  - Data source
  - Main data items
  - Units of measure
  - Who is responsible for the data
  - How often it is updated
- Record additional information where helpful (such as known limitations, validation checks or links to guidance).

### Maintain auditability

- Keep an audit trail of the data (for example, file name, date received, and key processing steps).
- Store raw files securely and consistently.

## Assess and improve data quality

- Complete a Sustainability Data Quality Assessment and Sustainability Data Quality Action Plan for your dataset.
- Set thresholds for readiness for the data to be used for analysis or reporting purposes (for example, ≥98% completeness or no invalid units).

## Communicate data quality

- Include a short Data Quality Note with every published output (suggested template below).

| **Data Quality Note** | |
|---|---|
| **Data quality description** | 3 to 5 sentences for example:<br>[This dataset represents electricity consumption data collected from supplier files and internal meter readings.<br>Basic validation checks have been applied to confirm units, reporting periods, and site identifiers.<br>Data quality is considered sufficient for Public Bodies Climate Change Duties Reporting and high-level trend analysis, though some estimates have been used where supplier data was unavailable.] |
| **Scope and period** | For example:<br>[Electricity consumption - All acute and community sites.<br>Reporting period – April 2024 to March 2025.] |
| **Known limitations** | For example:<br>[Electricity data for two sites was estimated for one month due to late supplier submissions.<br>Estimated values have been clearly flagged and will be replaced when final figures are received.<br>Comparisons at individual site level should be interpreted with caution for the affected period.] |
| **Quality status** | For example:<br>[**Amber** – because a small proportion of data has been estimated due to supplier delays, but overall completeness and consistency are high and the dataset remains fit for its intended reporting purpose.] |
| **Contact** | Provide team name and email address |

## Data owner

The Data Lead or Data owner sets the conditions that enable others to manage data quality effectively.

- Endorse the minimum standards (for example, timeliness SLAs, emission conversion factor version control)
- Ask for a monthly data quality snapshot (completeness, timeliness, issues, actions)
- Resource the most cost effective **root cause fixes**

# 11. Glossary

| Abbreviation / Term | Definition |
|---|---|
| **Accuracy** | How close the data is to the real-world value it is meant to represent. |
| **Completeness** | Whether all the data needed for the intended purpose is present. |
| **Consistency** | Whether data matches across systems and across time when it should. |
| **CSAT** | Customer Satisfaction Score – A measure of user satisfaction, sometimes used to assess whether data meets user needs. |
| **DQAP** | Data Quality Action Plan - A structured plan to address identified data quality issues, focusing on root causes and proportionate improvement actions. |
| **DQF** | Data Quality Framework - Guidance setting out principles, lifecycle stages and dimensions for managing sustainability data quality. |
| **Emission factor** | A value that converts an activity (such as kWh, litres, miles, or tonnes of waste) into an estimate of greenhouse gas emissions ($CO_2e$). |
| **KPI** | Key Performance Indicator - A measurable value used to assess performance against objectives. |
| **Lineage** | A record of where the data came from and what happened to it along the way (such as data being reformatted, updated emission factor being applied, categories being consolidated for standardisation). |
| **RCA** | Root Cause Analysis - A technique used to identify the underlying cause of data quality issues, rather than treating symptoms. |
| **Reference data** | Agreed lists or standardised values used across systems (such as site lists, meter lists). |
| **Timeliness** | Whether data is available at the right time and for the correct reporting period. |
| **Uniqueness** | Ensuring there are no unintended duplicate records. |
| **Validity** | Whether the data is in the correct format, range and units. |

# 12. Appendices

## Appendix A - Sustainability Data Quality Assessment

### What do we mean by a Data Quality Assessment?

A data quality assessment is a structured way of:

- Reflecting on how data is used
- Identifying which aspects of the data matter most for that use
- Recognising where quality issues may exist and why

It does **not** require data to be perfect, nor does it mean measuring everything. Proportionate assessment focuses attention on **what matters most**, based on risk and impact.

### Starting with purpose

Data quality thinking should begin with its purpose. A simple starting point is to complete the sentence:

*'This data exists so that we can…'*

Examples might include:

- Support decision making, meet reporting requirements, monitor performance, or enable operational activity.

Different purposes place **different demands on data quality**. The same dataset may need to meet different quality expectations depending on how it is being used.

### Identifying what really matters

Rather than starting at individual data fields, it is often more helpful to think about **priority parts of the data** that are critical to achieving its purpose.

A useful prompt is:

*'In order to [purpose], this part of the data needs to be correct.'*

This helps teams focus on the most important themes or components of the data, rather than attempting to assess everything at once.

### Using data quality dimensions proportionately

Once priorities are clear, data quality dimensions can be used to describe **what 'good' looks like** for that purpose.

Not all dimensions will be relevant in every case. For example:

- Some uses may rely heavily on completeness or timeliness
- Others may depend more on consistency or validity
- Some datasets may only need to meet quality expectations at an aggregate level

Selecting only the dimensions that genuinely matter helps ensure effort is targeted and meaningful.

### Early signals of data quality issues

Teams often recognise data quality issues before they are formally assessed. Common signals include:

- Reliance on manual checks or local workarounds
- Recurring questions about whether data can be trusted
- Unexplained differences between sources
- Delays that affect downstream processes

Capturing these observations can provide valuable input to future assessments.

## Linking assessments to action planning

Over time, structured data quality assessments will support the development of **data quality action plans**. These help teams:

- Understand the root causes of quality issues
- Prioritise improvements based on impact
- Track progress over time

Including assessment thinking early helps ensure that action planning is **purpose driven, proportionate and consistent** across NHS Scotland.

## Actions that can be taken now

Ahead of formal assessment methods being introduced, teams can:

- Document known data quality issues and their impacts
- Be explicit about limitations when using or sharing data
- Note where manual effort is relied upon
- Share learning and examples across teams

These steps build shared understanding and reduce the risk of repeated issues.

# Appendix B - Sustainability Data Quality Action Plan

## What is a Data Quality Action Plan?

A Data Quality Action Plan (DQAP) is a structured way to manage and improve data quality by focusing on whether data is fit for its intended purpose. It helps teams identify what matters most, decide how to measure quality proportionately, and prioritise actions that address the root causes of data quality issues.

## Key principles of a Data Quality Action Plan

- **Purpose comes first** - Every action plan is built around why the data exists and how it is used. Different purposes may require different quality expectations.
- **Focus on what matters most** - Rather than assessing everything, DQAPs concentrate on the priority parts of the data that are critical to achieving the stated purpose.
- **Use data quality dimensions selectively** - Relevant data quality dimensions are chosen to define what 'good' looks like for that purpose. Not all dimensions apply in every case.
- **Measure proportionately** - Simple, achievable metrics are used to understand whether data meets expectations. Measures may include percentages, counts or high-level checks, depending on risk and complexity.
- **Set realistic thresholds** - Thresholds reflect when data may no longer be fit-for-purpose, not when it is perfect. They vary by dataset, use, and impact.
- **Turn insight into action** - Assessment results are used to identify improvement actions, prioritised by risk and impact, and reviewed over time to track progress.
- **Review and adapt** - Data quality action plans are living artefacts. They should be revisited regularly to reflect changes in data, systems, or user needs.

# Appendix C - Root Cause Analysis

There are different approaches to root cause analysis (RCA). One helpful technique is the five whys.

Keep using the answer to the question 'why?', as the basis for the next question and continue asking 'why?' for five times (or more if you need to!). This should lead you to identify the root cause of the problem.

For example, inconsistent waste recycling rates between boards.

**Problem:** Recycling rates vary widely between boards and cannot be compared confidently

- **Why?** Boards are reporting different recycling figures for similar services
- **Why?** Waste categories are interpreted differently across boards
- **Why?** Guidance on waste classification is unclear or inconsistently applied
- **Why?** There are no shared definition set or practical examples
- **Why?** Governance for waste data standards is informal and not actively maintained

**Root cause:** Lack of consistent definitions and guidance for waste data classification.

**Action:** Agree and publish standard definitions with examples, embed them in reporting guidance and review through existing governance forums.

Some options for tackling root causes include:

- Improved training of staff
- Changing systems to make data entry easier
- Better guidance
- Adjusting automated processes
- Adding validation to fields
- Introducing data standards

It is important to note that cleaning bad data is **only** cost effective once you have tackled the root cause.

# NHS Scotland Data Quality Framework for Climate and Sustainability

With thanks to contributors from Public Services Delivery Scotland Data Office for their valued insight and collaboration.