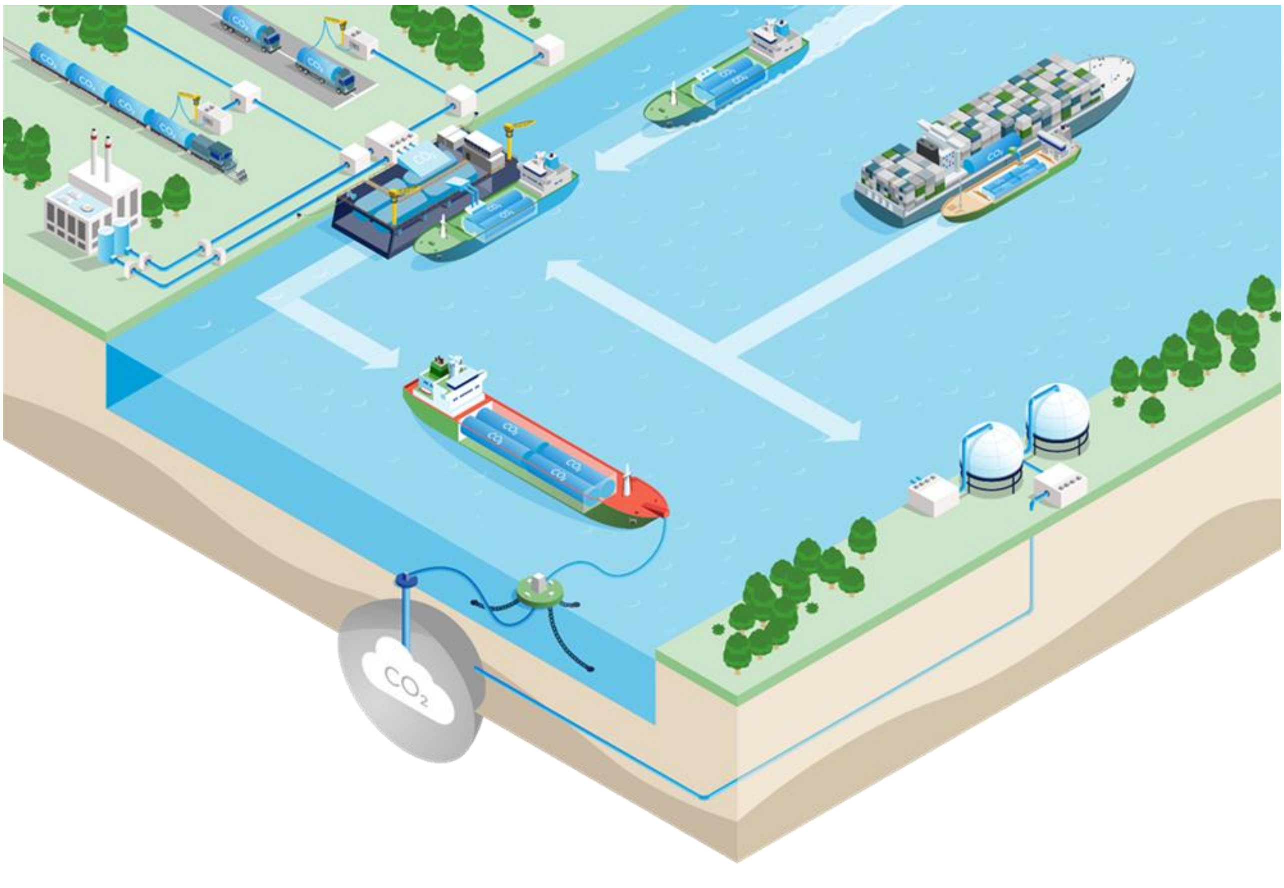
LONDON REGISTER OF SUBSURFACE CO2 STORAGE

Data quality assurance

Imperial College London Ltd.

Report no.: 10591504-02, Rev. 1
Document no.: 2789106
Date: 2025-12-16



DNV

Project name:	London Register of Subsurface CO2 Storage	DNV Services UK Limited Energy Systems
Report title:	Data quality assurance	Digital Energy
Customer:	Imperial College London Ltd., Exhibition Road Sherfield Building SW7 2AZ London United Kingdom	5th Floor Vivo Building Vivo Building 30 Stamford Street SE1 9LQ London United Kingdom GB 440 60 13 95
Customer contact:	Xiaowei Gao	
Date of issue:	2025-12-16	
Project no.:	10591504	
Organisation unit:	Digital Energy	
Report no.:	10591504-02, Rev. 1	
Document no.:	2789106	

Prepared by:	Verified by:	Approved by:
Simpson, Dave — Digitally signed by Simpson, Dave Date: 2025.12.18 08:34:33 Z	Digitally signed by Daoud, Ali Date: 2025.12.18 08:35:32 Z	Faiz, Graham — Digitally signed by Faiz, Graham Date: 2025.12.18 08:55:55 Z
Dave Simpson Principal Consultant	Ali Daoud CCUS Segment Lead – UK & Ireland	Graham Faiz Head of Digital Energy

Internally in DNV, the information in this document is classified as:

	Can the document be distributed internally within DNV after a specific date? No	Yes
☒ Open ☐ DNV Restricted	--	--
☐ DNV Confidential ☐ DNV Secret	☐	☐

Rev. no.	Date	Reason for issue	Prepared by	Verified by	Approved by
0	2025-12-04	First issue	D. Simpson	A. Daoud	G. Faiz
1	2025-12-16	Updated for publication	D. Simpson	A. Daoud	G. Faiz

Table of contents

1 EXECUTIVE SUMMARY

The London Register of Subsurface CO_2 Storage (“the register”) is a pioneering initiative that consolidates project-level global CO_2 storage data from operational carbon capture and storage (CCS) projects. This register aims to address the critical need for accessible, trustworthy, and comprehensive data to support scientific research, policy-making, investment decisions, and public transparency in the rapidly evolving CCS sector.

This assurance review, conducted by DNV, evaluates the register’s data quality and management practices with reference to internationally recognised standards, including DNV-RP-0497 and ISO 8000. The review process involved collaborative workshops to define the register’s core uses, requirements, and risks, and to establish the process followed for the register’s development and maintenance. Prioritised data quality principles have been established—completeness, understandability, traceability, and accuracy—as the foundation for assessing the register’s value and reliability. DNV has reviewed the register against the prioritised principles, providing recommendations for future development and enhancement.

The register demonstrates strong coverage and transparency, being available for peer review and benchmarking. However, limitations within the register include inconsistent metadata, lack of a formal methodology document, and reliance on manual data processing. These limitations introduce risks of error; and reduce scalability. While the register benefits from diverse data sources and public availability, improvements have been identified in standardising formats, documenting methodologies, and ensuring machine readability.

To enhance data quality and governance, the report recommends transitioning from spreadsheet-based storage to a relational database, formalizing data processing workflows, introducing robust quality control measures, and developing a definitive methodology document. These steps will improve traceability, usability, and long-term reliability.

The London Register of Subsurface CO_2 Storage provides a valuable foundation for global CCS data reporting. While the project is at an early stage of data quality maturity, the commissioning of this assurance review marks a significant step toward establishing best practices and robust governance. Implementing the recommended improvements will further strengthen the register’s role as a trusted resource for the CCS community and its stakeholders

2 INTRODUCTION

The London Register of Subsurface CO_2 Storage (hereafter “the register”) compiles project-level global CO_2 storage data from operational CCS projects on an annual basis. Storage data is sought for many purposes by the scientific community, policy makers, technology providers, and the public. Such purposes include tracking industry progress, project performance and global trends; which in turn inform insight, investment and policy decisions. Access to trustworthy data is valuable to support these activities.

Most operating storage projects report their storage rates publicly; however, these sources vary widely in format, publication type, and reporting methods. This leads to challenges in accessing and analysing global data; hence the register seeks to alleviate access and usability challenges by presenting a consolidated source of global data. It is worth noting at the outset that the ambitions of the register to provide publicly available, high value data are commendable.

This review has established prioritised data quality principles which embody the objectives and value of this dataset. The register has subsequently been reviewed against these principles. This has been carried out based on DNV-RP-0497 “Assurance of data quality management” /1/. DNV issued an interim report previously for inclusion in the publication of the register’s annual report /2/. This report supersedes the interim report and completes this assurance review.

2.1 Objectives

Data assurance provides confidence that data meets a specific need, and that its requirements have been or will be achieved. DNV has carried out an assurance review of the register to evaluate the extent to which it fulfils its requirements and make recommendations for enhancement.

The register is at an early stage, with initial publication in November 2025 /3/. Rather than carry out a critical claims-based assessment, this review establishes prioritised data quality principles as the foundation for assessing the register’s value and reliability. In this report we review the data gathering process, format and contents of the register against the data quality principles established. Reference is made to internationally recognised data quality standards, including DNV-RP-0497 and ISO 8000 /4/.

3 INDUSTRY CONTEXT

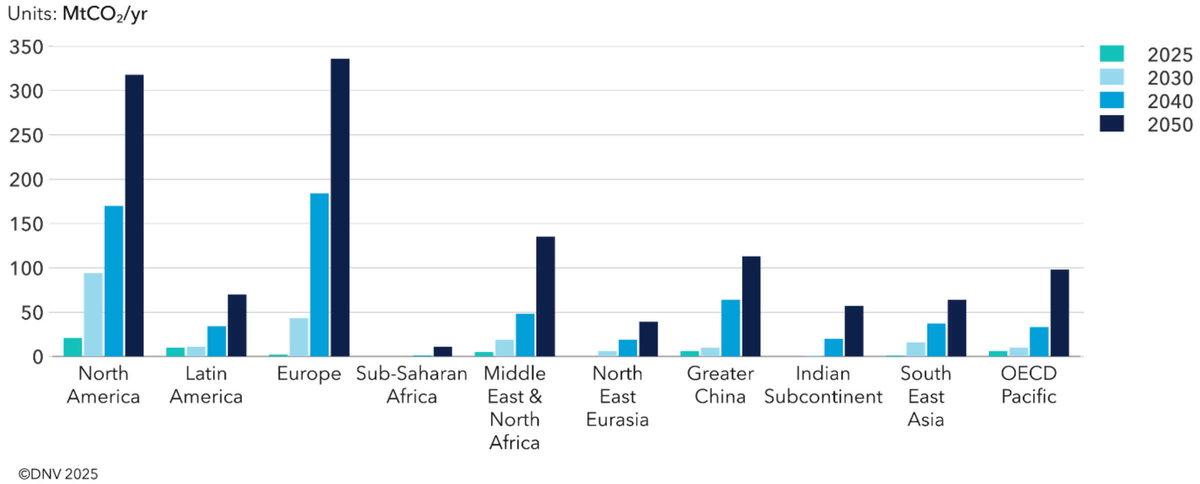
3.1 DNV CCS insights

DNV's Energy Transition Outlook (ETO) predicts that global carbon capture and storage (CCS) capacity is expected to reach approximately 1,600 million tonnes (Mt) of CO_2 annually by 2050, representing 6% of global emissions /5/.

While this represents significant growth from today's levels, this falls significantly short of net zero targets. Achieving this trajectory requires a quadrupling of CO_2 storage by 2030, reaching 6,040 Mt CO_2 annually by 2050, as outlined in the International Energy Agency's (IEA) "Net Zero Emissions by 2050 Scenario" /6/. Under this scenario, CCS is anticipated to contribute approximately 8% of total emissions reductions.

DNV's recent ETO report "CCS to 2050" /7/ discusses in detail the outlook for CCS including noting the recent acceleration in deployment. This expansion has been primarily driven by regulatory and fiscal incentives, including the 45Q tax credit in the United States, Japan's Act on Carbon Dioxide Storage Business and the European Union's Emissions Trading System (ETS) and Net Zero Industry Act. Industrial sectors such as cement, steel, refining, and chemicals remain the principal drivers of CCS adoption, while Europe, North America and the Middle East & North Africa are identified as leading regions for implementation. Figure 1 shows DNV's projections to 2050 across regions.

Figure 1 – Regional carbon capture and storage /7/



3.2 The role of data

DNV's ETO "CCS to 2050" report states:

"a consistent approach to reporting operational performance and transparency regarding the data could offer significant benefits to the CCS industry. Such data could enable more accurate quantification of CO_2 avoided and provide the basis for bench-marking and performance improvements"

In CCS, as for all components of the energy transition, achieving net zero requires clear strategy, policy alignment, investment, and global coordination: all of which are built upon a bedrock of quality data. The efforts of the London Register of Subsurface CO_2 storage in seeking to provide a consolidated view of global storage data are an important contribution to this critical foundation.

With more CCS projects coming into operation there is a need for transparent reporting of project performance, particularly to understand the impact of CCS as a climate change mitigation tool. As well as this, trust in reported volumes of CO_2 storage is critical when considering commercial implications associated with compliance and voluntary carbon markets, subsidy support schemes and tax credits.

Global CO_2 storage data can be used to track and compare industry progress across regions, influencing policy and technology development, and informing models and predictions in the effort to tackle climate change. Without high quality data these aims are inhibited, and risks are introduced, including:

- Reputational damage, to operators, reporting bodies, and the industry as a whole.
- Misinformed and suboptimal decision making.
- Loss of confidence in CO_2 storage technology, operators and industry.
- An unrealistic picture of the role of CO_2 storage, an inability to forecast and track the contribution to climate change.

For data to have impact it must be trustworthy, and it must be trusted. This is represented in the data's quality characteristics, reflecting the data sources, gathering and transformation processes, presentation format, and means of access.

3.3 Reporting standards

The register consolidates data from a variety of sources. Some of the data sources reference reporting standards, which give context to the data reported. These standards are summarised in Table 1.

Table 1 – reporting standards referenced by the register

Standard	**Description**	**Ref.**
IPIECA CCE-3	The International Petroleum Industry Environmental Conservation Association (IPIECA) provides sustainability reporting guidance tailored to the oil and gas sector, emphasising environmental, social, and governance (ESG) performance. Its framework promotes consistency in reporting greenhouse gas (GHG) emissions, energy use, and operational sustainability metrics, making it highly relevant for accurate CO_2 accounting and disclosure. CCE-3 is the indicator for lower-carbon technology, within module 3: climate change and energy. Reporting under IPIECA standards requires CO_2 storage data to be consistent, verifiable, and aligned with industry best practices, ensuring comparability across operators and compliance with global climate objectives.	/8/
GRI	The Global Reporting Initiative (GRI) offers a comprehensive sustainability reporting framework applicable across industries. Its primary focus is on transparent disclosure of environmental impacts, including climate mitigation measures and emissions accounting. For CO_2 storage, GRI supports reporting on the effectiveness of carbon management strategies and their contribution to broader sustainability objectives. Under GRI, data must adhere to principles of accuracy, completeness, and reliability, enabling stakeholders to evaluate the effectiveness of CO_2 storage within broader sustainability strategies.	/9/
SASB Standards EM-EP	The Sustainability Accounting Standards Board (SASB) develops sector-specific standards that integrate ESG considerations into financial reporting. Its emphasis on financially material issues, such as climate risk and carbon management, ensures that sustainability disclosures are aligned with investor expectations and enterprise value considerations.	/10/

	Compliance with SASB standards ensures that CO_2 storage data is precise, auditable, and integrated into financial reporting frameworks, positioning technical performance within the context of enterprise value and risk mitigation.	
HKEX-ESG standards KPIs	The Hong Kong Exchanges and Clearing Limited (HKEX) ESG Reporting Guide mandates ESG disclosure requirements for listed entities, focusing on governance, environmental responsibility, and stakeholder engagement. Its framework requires detailed reporting on emissions management and climate resilience strategies, making it particularly relevant for CO_2 storage initiatives in Asian markets. For HKEX-ESG, reporting involves embedding CO_2 storage data into corporate governance structures, ensuring transparency and traceability. This approach enhances investor confidence and supports regulatory compliance in Asian markets.	/11/
U.S. Environmental Protection Agency Greenhouse Gas Reporting Programme (EPA GHGRP)	The EPA Greenhouse Gas Reporting Program (GHGRP) mandates annual reporting of GHG emissions from large facilities, supporting national climate policy through standardised emissions accounting. For CO_2 storage, GHGRP specifies methodologies for quantifying captured and stored CO_2, along with verification and recordkeeping protocols. Reporting under GHGRP requires data to be accurate, complete, and supported by rigorous measurement systems, enabling credible assessment of emissions reductions and reinforcing transparency in climate mitigation efforts.	/12/

The specifications given in the standards are varied, with some giving more detailed definitions or requirements than others. Some standards relate to wider GHG reporting at a company level e.g. IPIECA CCE-3, compared to EPA GHGRP Subpart RR which focuses on CO_2 stored for an individual site. It is relevant to note that the referenced standards include both voluntary and compliance standards. For example, the EPA GHGRP Subpart RR is a mandatory reporting scheme for projects injecting CO_2 in the United States and the HKEX-ESG standard is mandatory for companies listed on the Hong Kong Stock Exchange. The GRI, IPIECA and SASB standards/guidance are voluntary in nature.

There is not a straightforward relationship between data being reported under these standards and confidence which can be placed in the data itself. Requirements for verification or external assurance vary, as do the levels of transparency required. EPA GHGRP Subpart RR is rigorous in its reporting requirements, definitions, and governance processes - and relates specifically to subsurface CO_2 storage. This is tied to the administration of the 45Q tax credit. The other standards listed are not comparable in their level of rigour because either:

1. They only indirectly relate to subsurface CO_2 storage, being primarily related to emissions targets or overall environmental and social factors; or,
2. They constitute *guidance* which isn't verified or mandated.

The reporting standards therefore constitute useful project level context, but don't necessarily bolster the confidence in the values presented in the register.

The register would benefit from an accompanying outline of the relevance of the reporting standards to the data presented; see recommendations in section 7. This would be particularly beneficial when considering potential differences in how the CO_2 storage rate is reported across different standards and programs, and the implication for the confidence in the reported figures.

DNV

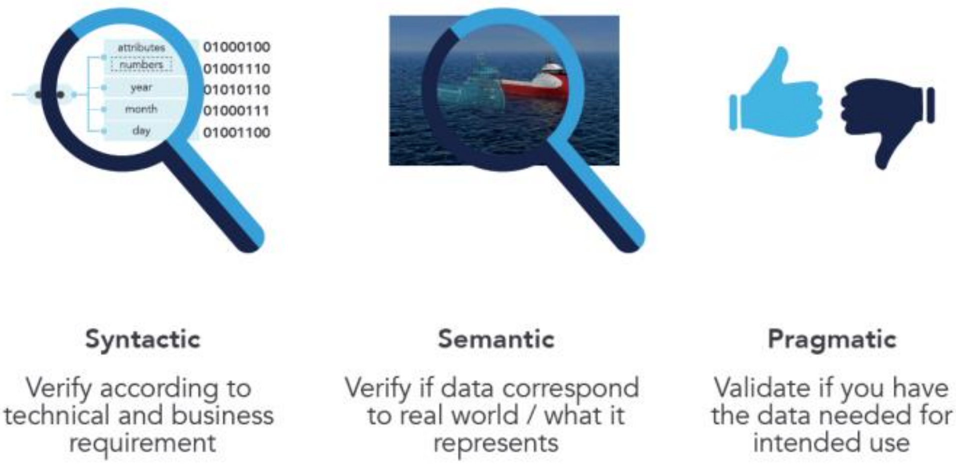
4 ASSESSMENT METHODOLOGY

DNV's recommended practice DNV-RP-0497 "Assurance of data quality management", provides a framework for the assessment of data quality and data quality management. This is firmly rooted in ISO 8000, providing a practical application of data quality principles detailed in this ISO series.

ISO 8000 defines fundamental concepts of data quality and their relationship to quality management systems. The standard defines three categories for data quality:

- **Syntactic:** The degree to which data conforms to their specified syntax.
- **Semantic:** The degree to which data corresponds with that which it represents.
- **Pragmatic:** The degree to which data is appropriate and useful for a particular purpose.

Figure 2: ISO 8000-8 data quality assessment model



4.1 Syntactic quality

Syntactic quality measures the extent to which a dataset conforms to its declared structure — including schema, metadata, and pattern requirements. As defined by DNV-RP-0497 and ISO 8000, this involves checking syntactic and integrity constraints against a defined schema and rule base, supported by data profiling to identify illegal values, outliers, or structural anomalies before semantic or pragmatic assessment. The approach is bottom-up and often tool-assisted, and may include steps to:

1. Establish or confirm the reference schema and conceptual model.
2. Profile the data to reconstruct missing metadata and identify inconsistencies.
3. Express syntactic requirements as executable rules (e.g. for completeness, uniqueness, type/format, domain lists, allowed ranges, precision/scale, and temporal bounds); and
4. Perform automated checks to derive metrics (e.g., pass/fail counts, defect rates) and produce flagged-record lists for remediation.

To carry out these quality checks, syntactic requirements should be clearly defined and measurable. The register currently does not include syntactic definitions. It is recommended that such a specification be developed, to aid understanding and enable definition of quality checks and rules, see recommendations in section 7.

4.2 Semantic quality

Semantic quality assesses how accurately the data reflects the real-world concepts it represents. While syntactic checks confirm structure, semantic checks confirm meaning. This ensure that values are consistent, logical, and aligned with reference sources.

The assessment follows a model-driven approach:

1. Define the semantic model and relevant reference sources.
2. Profile the data to identify inconsistencies or implausible values with respect to physical reality.
3. Implement executable rules to test accuracy, consistency, and coherence; and
4. Generate metrics and exception lists for review and remediation.

The above-specified activities implement the semantic data quality considerations from the DNV-RP-0497 framework, ensuring that both meaning and relationships are consistent, traceable, and verifiable throughout the dataset.

4.3 Pragmatic quality

The pragmatic quality reflects the degree to which the data is considered fit for its intended use — being complete, timely, and reliable enough to support decisions or analyses. While *syntactic* checks confirm structure and *semantic* checks confirm meaning, pragmatic checks evaluate usefulness within the application context.

The assessment takes into account both the operational and decision-making context of the dataset. It examines whether the available information is sufficient to meet the stated purpose, whether appropriate confidence levels are assigned, and whether any known limitations affecting its use are documented. Typical considerations include data coverage, temporal consistency, stability, and reproducibility of derived values.

Pragmatic checks confirm that the dataset can be used with an agreed level of confidence for its objectives, highlighting where supplementary data or clarification may be required.

4.4 Data Profiling

Data profiling is a valuable and commonly used method for exploring the characteristics of datasets, particularly where metadata is limited or absent. It enables the empirical investigation of statistical patterns and the identification of potential quality issues such as illegal values, outliers, and other anomalies. Through this process, metadata can be re-engineered, providing deeper insights into the dataset's composition and integrity.

Data profiling is often used in the early stages of data activities, such as migrations, integrations, and analytics to anticipate whether actions might be necessary to address the quality issues. For the register data, the results of the profiling can form the basis of syntactic quality requirements. Profiling has also informed generation of indicative data quality rules, see Appendix B.

Section 6 presents the outputs of the data profiling activities conducted. These included structural checks such as data types, maximum length and precision, as well as key integrity indicators like primary key flags.

5 ASSURANCE BASIS

An assurance basis consists of an agreed framework within which an assurance review can be carried out. This is critical to the establishment of the "claims" against which the solution can be reviewed, with reference to evidence items. The assurance approach is outlined in DNV-RP-0497.

While the objective of all organisations, projects and activities may be to "maximise data quality", it is important to take a deliberate and explicit approach to defining requirements. ISO/IEC 25030 /13/ provides a framework for quality requirements, noting:

"Quality requirements can have conflicts with one other. These conflicts inevitably arise as a result of the inter-relationships between quality characteristics. In case that some quality characteristic can negatively influence on another, trade-offs should be conducted to resolve the conflict to find the right balance of them."

ISO 25012 /14/ defines a general data quality model, and notes:

"Data quality characteristics will be of varying importance and priority to different stakeholders."

Collaborative definition of data quality requirements, and transparency in quality objectives allow prioritisation of activities to maximise the value of data and promote trust in data users. This is the purpose of establishing an assurance basis.

DNV held a collaborative workshop with the register team on 30th October 2025 to establish and consolidate the core purpose and requirements of the register, incorporating:

1. Uses: both current and potential future uses. This serves to define the substance of the register – its specification; as well as forming the foundation for prioritised requirements.

2. Requirements: based on anticipated uses. This identifies the characteristics and properties of the register which should be in place for effective use. Data requirements inform key data quality metrics which can be used for assurance evaluation, and quality monitoring.

3. Risks of poor quality. Identification of key risks further clarifies the priority of aims for the register, and requirements for high quality data in this context.

The assurance basis is a live document and can be updated throughout the life of the register as its purpose, users and requirements evolve. Any formal assurance statement must be linked to a static version of the assurance basis.

5.1 Purpose

The London Register of Subsurface CO_2 Storage is a centralised register of annual rates of CO_2 stored underground from operational projects worldwide /15/. Most operating storage projects report their storage rates publicly; however, these sources vary widely in format, publication type, and reporting method. This leads to challenges in accessing and analysing global data. The register's primary purpose therefore is to present a consolidated source of global data.

5.2 Core requirements

The outputs of the workshop have been consolidated to capture the intended use and therefore requirements of the register at this stage. The raw outputs gathered during the workshop are presented in Appendix A.

5.2.1 Uses: functions

The register enables the following activities:

1. Track, monitor and compare: industry progress, project performance, trends across projects, geographies and time.

2. Gather global data into a single consolidated source.

3. Provide simple and direct metrics indicating scale, performance and overall storage levels.
4. Serves as a focal point for global subsurface CO_2 storage, as a prominent, consolidated source.

5.2.2 Uses: purpose

The register enables the activities in section 5.2.1 to:

1. Promote transparency and trust: countering misinformation, building public confidence and ensuring accountability.
2. Provide policy and investment decision support: informing incentives, directing investor confidence and commercial viability.
3. Pursue equity and fairness: identifying global disparities and guiding targeted climate actions.
4. Provoke scientific and societal insight and action: providing empirical data for model development and data-driven insight. This includes the assessment and quantification of storage impacts on climate change.

5.2.3 Requirements

The foundational motivation for establishing the database was the difficulty in gathering data from disparate sources for any holistic view of global CO_2 storage. The originating requirement of the register is therefore that it is as comprehensive as possible: covering all operational CO_2 storage projects globally; and that it is sufficiently transparent to promote trust and therefore use. The following general requirements apply:

1. Completeness: global coverage, enriched with layered project data and accompanying metadata.
2. Usability and understandability: users are able to retrieve and understand the data sufficiently for their use; this includes a structured favouring flexibility and future scalability.
3. Transparency and traceability: full data lineage readily available promoting trust and enabling use-case decision making (including limitations).
4. Data integrity and accuracy: 'trueness' enabling robust and quality decision making (includes considerations for uncertainty quantification).

These requirements are supported by definitions in ISO/IEC 25012:

- Completeness; the degree to which subject data associated with an entity has values for all expected attributes and related entity instances in a specific context of use.
- Understandability: the degree to which data has attributes that enable it to be read and interpreted by users, and are expressed in appropriate languages, symbols and units in a specific context of use.
- Traceability: the degree to which data has attributes that provide an audit trail of access to the data and of any changes made to the data in a specific context of use.
- Accuracy: The degree to which data has attributes that correctly represent the true value of the intended attributes of a concept or event in a specific context of use.

5.3 Data quality principles

The core requirements presented above can be represented in the following principles:

1. The data is sufficiently comprehensive, covering all major operational subsurface CO_2 storage projects globally.
2. The data is sufficiently well documented and structured so that a target users can utilise it easily, and without additional support.

3. The data sources and processing methodology is sufficiently well documented to inspire trust in the broad community, and enable understanding of limitations for individual use cases.
4. The data reflects the truest available representation of injection rates for each project.

5.4 Assurance plan

The register is at an early stage, with the initial publication in November 2025. Rather than carry out a critical claims-based assessment, this assurance review establishes the prioritised principles which form the core of the register's value. In this report we review the data gathering process, format and contents of the register against the data quality principles outlined above.

These principles could constitute top-level claims and form the basis of a detailed assurance assessment in the future. This could be further substantiated by structured evidence and accepted through argument-based assessment. Through the course of the register being used by a broad, global community, the objective of the register may be further refined or expanded. This could lead to stricter quality requirements; and the refinement of the assurance basis to specific provable claims.

A narrative assurance review is considered appropriate at this stage given the relative maturity of the register. DNV has therefore reviewed the register against the prioritised principles, providing recommendations for future development and enhancement.

6 EVALUATION

DNV evaluated the register against the agreed data quality principles described in section 5.3, with recommendations for where quality can be improved in these areas.

6.1 Data profiling

The main body of the register is contained in a key sheet in the published spreadsheet file – this is the subject of the profiling review in this section /15/. There two additional sheets:

1. A readme, containing narrative instructions, references, and definitions.
2. Pivot table of the main register sheet, consolidating the values into a strict tabular form.

The register data is stored in a tabular structure, but split by geographic region, with national totals. Merged cells indicate a new geographic section. Overall totals and counts are presented at the bottom of the table. Colours are used to indicate:

1. Where a cumulative total in the source has been averaged across multiple years.
2. The source which has been selected as primary, where multiple exist.

The datasheet also includes annotations such as units and other footnotes. A basic representation of the format of the data is shown in Figure 3. The register's column identifiers can be split into broad functions. These are shown in Table 2, along with data types for each .

While this data profiling activity consists primarily of exploration rather than evaluation, a few observations can be made:

- The dataset can be considered 'sparse', i.e. there are a high number of empty cells. This is largely due to the broad time period captured, and also reflects the structural format of the register. Figure 4 shows the relative proportions of null values in each column.

- A primary key is a unique identifier for each row of data. The register data has a composite primary key, being the project name and data source, columns "CO_2 Storage Project" and "Data Source. In the register data these values are always complete, with no duplicates, meaning these function as a usable primary key – uniquely identifying each row.
- While Table 2 indicates that the storage values are expected to be numeric, there exist anomalies including symbols "/" and "*". In addition to null values (empty cells) and the presence of zero (0) values, the variety of ways to indicate a null value may benefit from standardisation, if all of these are equivalent, or from commentary giving clarity to the user.
- There is duplication between sheets, including data source references in the "readme" and the data values repeated in the pivot table.
- The overall presentation of the register requires informed interpretation for use, particularly by reference to the various annotations and instructions. This requires time invested by users, and risks misinterpretation or possible inaccessibility e.g. through use of colours to convey key information.
- Due to this multi-media storage format – using colours as well as annotations, and inclusion of merged cells for titles and subtotals, the machine readability of the register is currently very limited.

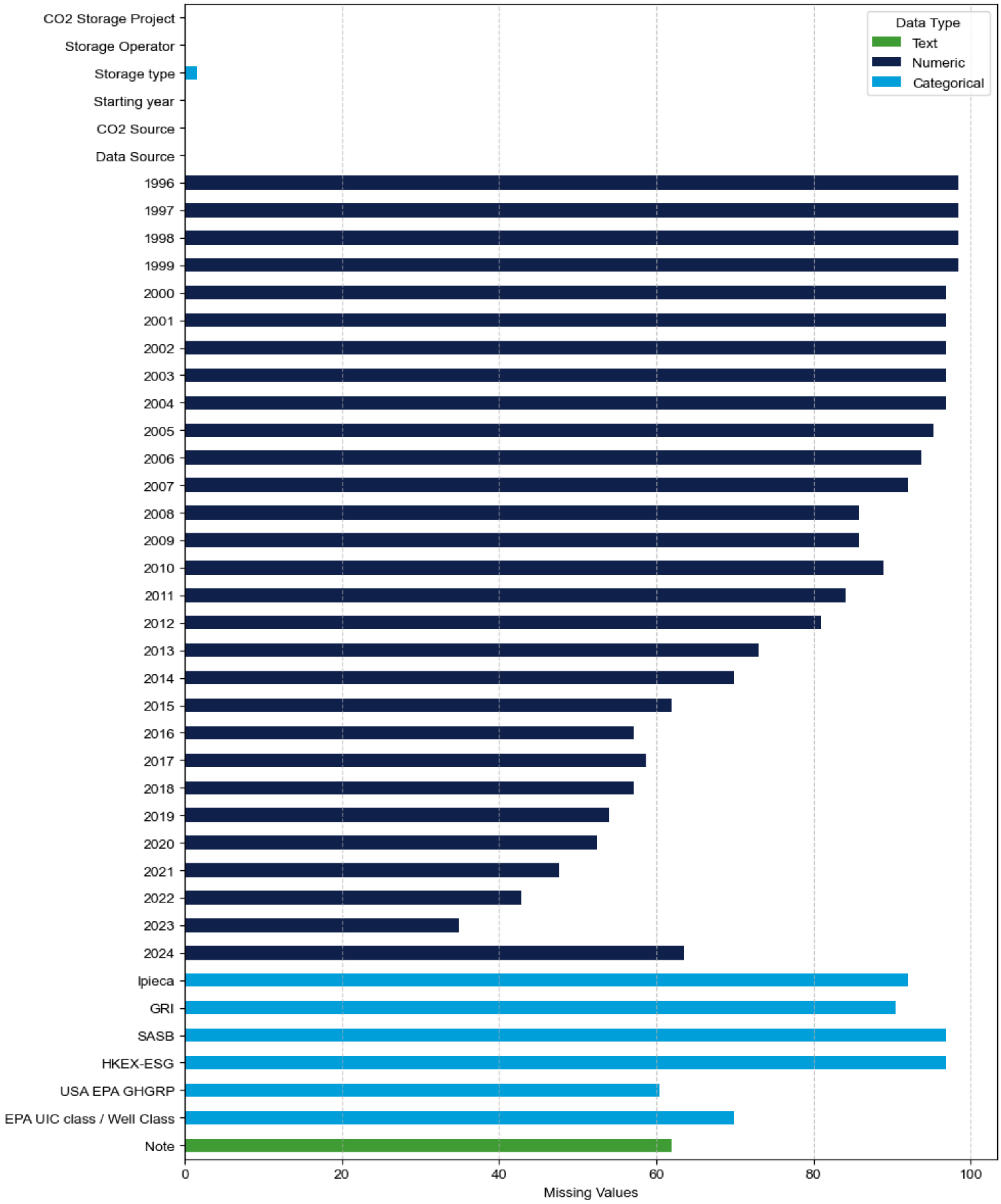
Figure 3 – Basic representation of the register's structure

Column identifier 1	Column identifier 2	…
National identifier		
Data values 1	Data values 2	…
National subtotal 1	**National subtotal 2**	…
National identifier		
Data values 1	Data values 2	…
National subtotal 1	**National subtotal 2**	…
…	…	…
Global total 1	**Global total 2**	…

Table 2 – CO_2 dataset column summary

Column	Data Type	Function
CO_2 Storage Project	Text	Project identifier
Storage Operator	Text	Project metadata
Storage type	Categorical	
Starting year	Numeric	
CO2 Source	Text	
Data Source	Text	Reference for given storage values
Year [1996-2024]	Numeric	Storage values
Ipieca	Categorical	Project metadata
GRI	Categorical	
SASB	Categorical	
HKEX-ESG	Categorical	
USA EPA GHGRP	Categorical	
EPA UIC class / Well	Categorical	
Note	Text	Additional commentary

Figure 4 – Percentage of null values in each column in the register



6.2 Project data gathering process

The register consists of operational subsurface CO_2 storage projects worldwide from 1996 up to the present day. Projects are identified using global CO_2 storage project databases and register consortium input. The focus is strictly on subsurface injection volumes – that is, CO_2 stored geologically.

Annual CO_2 storage amounts are gathered from multiple source types. Primary sources include data reported directly by project operators (for example, values published in company sustainability or ESG reports, or figures provided through storage operator surveys) and official data from government national inventory reports or regulator databases (such as the U.S. EPA Underground Injection Control program reports). Secondary sources – for instance, industry reports, research publications, or environmental assessments – are included particularly when primary (operator or official) data are unavailable.

The data gathering and presentation process followed to compile the register has been provided by the register team. Figure 5 summarises the simplified process.

Figure 5: Data gathering process

These steps are broken down into more detail below:

1. Identify relevant projects from global reports of operational CO_2 storage projects, via input from the project consortium, or under known operators.
2. Search for storage data sources, including project disclosure reports, government databases. press releases, emissions reporting, etc.
3. Sources are archived (downloaded and saved) in the register file storage.
4. Locate storage values in reports via keyword search.
5. Copy values out of the reports.
6. Convert units to $MtCO_2$/yr where necessary.
7. Average multi-year values: where a cumulative value is given over multiple years, an average yearly rate is assumed – equally divided among the years.
8. Record values in the register spreadsheet (Microsoft Excel; xlsx) file
9. Prioritise data sources where multiple exist, through discussion amongst register team.
10. Add the data source reference to the register file, where publicly available. Maintain maximum precision from the data source.
11. Update formatting used in the register spreadsheet to indicate:
    a. Where cumulative values have been averaged.
    b. Which source has been prioritised, where multiple exist.
12. When an update is published: upload xlsx file to project website, using a unique indicator in the filename to differentiate releases.

All the compiled data are stored in the master xlsx file which serves as the single source of truth for the annual CO_2 storage figures.

The intention is that an updated version of the register is published on an annual cycle and archived with version identifiers (for example, by date or release number) to ensure that any changes or revisions over time are fully traceable. The register is available as a downloadable spreadsheet on the project's website and contains source references for project datasets, where available. A thorough description of the data presented is given in section 6.1.

6.3 Sample data review

A sample of projects' data has been reviewed to evaluate the data lineage and clarity of presentation in the register. Some discrepancies were identified, including:

1. Differences in precision compared to source data, impacting calculation of annual rates from cumulative values.
2. An instance of double counting, where a single year's rate is attributed to two subsequent years.
3. An instance of a mistake in one of the notes annotating the data.
4. Some uncertainty regarding total cumulative injection vs yearly injection rates.
5. Referenced URLs not working, having been removed or redirected.
6. An instance of reported values not being evident in the source, perhaps indicating additional processing or assumptions which aren't indicated in the register notes.
7. In some case the data source is given as "Storage Operator Response to Survey" when in fact the data has been sourced through an intermediary.

DNV considers these discrepancies to be the result of manual errors, compounded by the lack of an audit log showing data processing steps carried out between the source and the register. Recommendations for improvement are given in section 7.

6.4 Data quality evaluation

The register has been evaluated against the data quality principles established in section 5.3. These are grouped into overarching data quality characteristics defined in ISO 25012. The evaluation outcomes can be seen in Appendix C, and summarised in the following points:

- Completeness is generally robust, supported by diverse sources and the publication for peer review and benchmarking. Challenges exist in capturing historical changes and reconciling conflicting published information, however this is largely beyond the control of the project team.
- Understandability is aided by contextual notes and consortium review processes, but the absence of a formal methodology, inconsistent formatting, and limited metadata reduce clarity and usability, particularly for automated analysis.
- Traceability benefits from the inclusion of public URLs and the maintenance of a project archive storing copies of the data sources. The lack of storage of extracted raw values, transformation records, and detailed metadata limits auditability and reproducibility. Intrinsic volatility in carbon storage project reporting is reflected in inaccessible source links.
- Accuracy must be framed against how representative the data is of the available sources, and presentation of source and assumptions. This is somewhat enabled by the transparency in publication, but the discrepancies identified during sample checks and the absence of embedded quality controls pose risks. Use of a spreadsheet format poses additional risks and limits scalability, with challenges including limited metadata layering, limited version control, and vulnerability to accidental overwrites, making it unsuitable for complex, evolving datasets.

Overall, while the dataset provides a valuable foundation for global CO_2 storage reporting, systematic improvements in documentation, metadata, and process would enhance trust, usability, and scalability. Recommendations are detailed in section 7.

It is important to note the significant value represented by the explicit prioritisation of these quality characteristics by the register. Trust in published data is often presumed by association with an organisation or recognisable name, with methods and sources being unclear or inaccessible. Here the register seeks to provide trust which is self-evident. This goes beyond comparable datasets but also represents an effort to maximise the value of data.

A series of draft data quality rules have been developed and can be found in Appendix B. These have contributed to the evaluation, however, it is recommended that the project consider establishing data quality rules as the register matures. These rules can then be monitored as the register is updated, and used for review before each publication.

7 FINDINGS AND RECOMMENDATIONS

The register demonstrates a strong commitment to transparency and broad coverage of global subsurface CO_2 storage projects.

The register has made good initial steps in aligning with its core principles and fulfilling its key data quality requirements. The project is at a relatively low maturity in data quality management. DNV-RP-0497 defines organisational capability levels, which are further detailed in DNV-RP-A204 /16/. In these terms the project is at Level 1: “Initial” with limited formal governance allowing quality monitoring and measurement. The completion of this assurance review is a good initial step in establishing this governance.

The recommendations here are aimed at increasing this maturity towards Level 2: “Repeatable”, incorporating some basic governance and processes and measurement; and Level 3: “Defined” which embeds governance and best practice. The recommendations are consolidated in Table 3, and referenced in the following sections grouped by theme.

DNV have recommended a priority ranking for the recommendations based on anticipated impact.

Table 3 - Recommendations

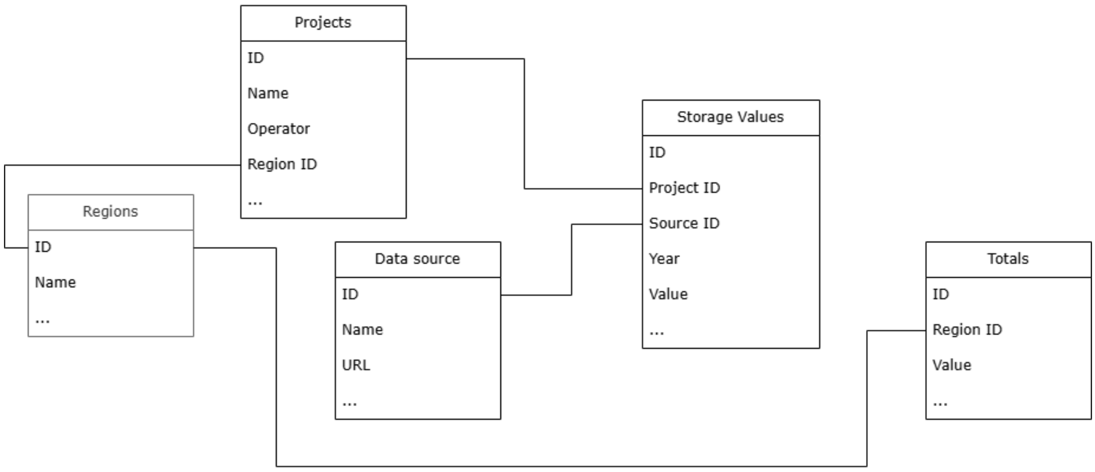
ID	Recommendation	Priority
1.	Store the register data in a relational database, rather than spreadsheet format.	Should
2.	Store extracted raw values and processed data separately.	Must
3.	Include a QC step in the workflow, using the “four eyes” principle.	Must
4.	Keep a data processing log, showing transformation steps taken from data source to register.	Should
5.	Document the methodology for publication alongside the register.	Should
6.	Standardise presentation conventions, with clear definitions.	Should
7.	Develop a metadata specification, describing the register data.	Should
8.	Establish data quality rules, against which the register can be evaluated prior to publication.	Should

7.1 Data storage method

Transitioning from a spreadsheet format to a relational database is strongly recommended to improve scalability, traceability, and data integrity [recommendation 1]. A relational structure allows for constraints that enforce data validity (e.g., mandatory fields, standardised units), reducing errors and inconsistencies. This would also support robust archiving of historical versions, enabling audit trails and long-term reference. Relational models also facilitate annotations, multiple relationships, and categorization, which are essential for linking projects, sources, and metadata without cluttering the main dataset. This would allow elimination of the use of colours as a primary means of **storing** useful project data, though colours could still be used in presentation. Figure 6 shows a simple example of a relational model for the data, for illustrative purposes.

Separately storing both raw extracted values and processed data supports transparency and facilitates quality checks [recommendation 2]. Use of a relational database model would support this separation of raw data from the dataset presented to users, and give flexibility in the presentation format and reporting available. Current duplication seen in the register spreadsheet file could be eliminated, where currently any updates made to these elements – including the storage values – must be updated in multiple places. Automated processing based on a single source of truth would enhance repeatable and trustworthy publications.

Figure 6 – Example of relational data design



7.2 Methodology and data processing

Data gathering and transformation currently relies on manual steps with limited oversight, introducing risks of inconsistency and error. Implementing a documented workflow with any required manual steps clearly defined, will improve reliability.

Significant benefit would be gained by introduction of a quality control (QC) process using the "four-eyes principle", where every update is reviewed by a second person before publication [recommendation 3]. This would significantly reduce the risk of manual errors.

Separation of extracted values and processing, from the published register output is discussed in section 7.1. This would support the use of a processing log, which is critical for traceability: capturing decisions, assumptions, and transformation steps. This log should be linked to the dataset for transparency and audits, and reviewed as part of the embedded QC process [recommendation 4].

Development of a definitive methodology document would be beneficial [recommendation 5]. This should include a detailed and formalised update process to ensure consistency and accountability. The methodology document should outline identification of data sources, processing rules, assumptions, and also source prioritisation criteria. This would reduce reliance on ad hoc annotations within the dataset and provide a clear reference for users. This should also include description of the significance of the referenced reporting standards, discussed in section 3.3. This structured process will enhance traceability, accuracy, and trust in the dataset.

7.3 Data presentation

Presentation of the register is discussed in section 6.1. The register is commendable for targeting clarity and consistency as a priority, and the separation of 'back-end' and 'front-end' data would position the register for significant enhancements in this regard, where currently the register is performing both functions. Complex relationships and metadata can be robustly stored in a back-end database, allowing the published output to prioritise user experience.

Standardised formatting conventions would eliminate ambiguity, for example in the use of different characters to indicate null values [recommendation 6]. Colour coding or visual indicators can enhance user experience, particularly in making a published report aesthetically pleasing, though this comes at a potential accessibility cost.

7.4 Governance

An important part of the register documentation should be the metadata specification [recommendation 7], which defines the structure, format and constraints applicable to the register data, ensuring standardisation across the register team. Additionally this specification allows for technical constraints to be applied to the storage method – for example enforcing data types – as well as the assessment of syntactic data quality.

As the register team seeks to improve the project's maturity in data quality management, establishing a set of fundamental data quality rules would be beneficial. This allows clarity on the desired data quality, and the assessment of the register against this level [recommendation 8].

8 REFERENCES

/1/ DNV RP 0497, Assurance of data quality management. October 2023

/2/ DNV, Data assurance summary, Doc No. 2789106, Report number 10591504-01, Rev. 0, 12/11/2025

/3/ Gao, X., & Krevor, S. (2025). The London Register of Subsurface CO_2 Storage: 2025 Annual Report. Imperial College London.

/4/ ISO 8000. Data quality

/5/ DNV (2025), Energy Transition Outlook 2025

/6/ IEA (2023). Net Zero Roadmap: A Global Pathway to Keep the 1.5 °C Goal in Reach. IEA, Paris. https://www.iea.org/reports/net-zero-roadmap-aglobal-pathway-to-keep-the-15-0c-goal-in-reach

/7/ DNV (2025). Energy Transition Outlook "CCS to 2050"

/8/ IPIECA (2025) Sustainability reporting guidance for the oil and gas industry. https://www.ipieca.org/resources/sustainability-reporting-guidance

/9/ GRI (2025). GRI Standards English Language. https://www.globalreporting.org/how-to-use-the-gri-standards/gri-standards-english-language/

/10/ SASB (2025). SASB Standards overview. https://sasb.ifrs.org/standards/

/11/ HKEX (2024). How to prepare an ESG Report. https://www.hkex.com.hk/-/media/hkex-market/listing/rules and-guidance/environmental-social-and-governance/exchanges-guidance-materials-on esg/app2_envirokpis

/12/ EPA (2025). Subpart RR – Geologic Sequestration of Carbon Dioxide. https://www.epa.gov/ghgreporting/subpart-rr-geologic-sequestration-carbon-dioxide

/13/ ISO 25030:2019 Systems and software engineering — Systems and software quality requirements and evaluation (SQuaRE) — Quality requirements framework

/14/ ISO 25012:2008 Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model

/15/ "20251105_LondonRegisterofSubsurfaceCO2Storage.xlsx" provided by Xiaowei Gao, 5th November 2025

/16/ DNV-RP-A204. Assurance of Digital Twins. October 2023

APPENDIX A – ASSURANCE WORKSHOP OUTPUTS

This section contains the **raw** outputs from the assurance basis workshop.

A.1 Data uses

This section includes ideas gathered via interactive whiteboard for uses - known current, and possible future - of the London register of subsurface CO_2.

ID	Use
1	Track industry progress
2	Track historical CO2 injection
3	Track project performance
4	Tracking the project-level annual performance.
5	Track injection volumes to understand regional pressure increases/effects
6	Track venting/EOR recycling losses/leaks - potential net storage calculation
7	Track carbon storage contribution to climate change
8	Track the geographic trends in deployment, within countries/regions
9	Have a way of tracking global storage into the future as it scales (expecting dramatically)
10	Track permanent vs EOR
11	Understanding of performance
12	Match performance with finite storage resources
13	Understand effectiveness of CCS as a climate mitigation tool
14	Estimate climate change mitigation benefit
15	Assess scalability/trajectory of storage technology
16	Provide a simple metric to demonstrate the scale of deployment
17	Provide data-driven merits for the CCS trajectory research
18	Provide a scientific basis on which academics/public/policy makers can assess the potential and challenges for using CCS in climate change mitigation
19	A roadmap for transparency and trust in global carbon storage
20	Independent assessment of performance
21	Provide a factual basis to counter misinformation about CCS/CDR
22	Provide a focal point for more detailed data - via total storage numbers with references to detailed reports etc.
23	Confidence in CCUS for the public as a climate change measure?
24	Good publicity for storage operators/industry?
25	Accountability for industry
26	Determine if projects have net reduction of CO2
27	Influence policy for carbon storage (including incentives)
28	Watch gaps between plans and operations.
29	Delivering measurable carbon mitigation activities and measuring the achievements of projects.
30	Identify inequality in Carbon Storage among Countries, thereby proposing targeted and practical climate actions in the future.
31	Confidence in commercial applications e.g. CDR sales, Allowances, tax credits
32	Stimulate confidence for CCUS investors
33	Understand how business case affects injection volumes

34	If production history + up/mid/downstream and scope 3 emissions etc. for a field then EOR benefit could be quantified
35	Enable comparison across regions/projects
36	Identify most and least successful project chains, e.g., CO_2 from natural gas processing and storage in oil fields
37	Evaluate gaps between project plans and implementation
38	Parameterise and constrain models of projections of future deployment with empirical data, e.g., growth models, integrated assessment models

A.2 Data Requirements

This section includes ideas gathered via interactive whiteboard for requirements for the dataset, based on anticipated uses.

ID	Requirement
1	Data structure should be relational
2	Machine readable? So that it can be e.g. programmatically analysed
3	Flexible - in that it can scale as data grows - e.g. 10x projects
4	Multiple sources of data for the same project for the same time period should be compiled
5	Original data should be stored and built on with calculations analysis etc.
6	Access to un-reconciled data alongside description of how heterogeneities are reconciled
7	Traceable - needs to be clear where it came from
8	Trust in the data source - basis for reporting
9	Traceability - offline collation of sources
10	Full referencing and documentation on conversions
11	Download hard copies of everything
12	Data sources are archived
13	Public data is preferable for transparency
14	Accurate - data needs to be 'correct'
15	Source types and quality assessments are key
16	Self-consistency in the data
17	Scientifically choose the decimals to mitigate uncertainties
18	Consistency around the variable reported
19	Consistent (definitions) so that data is comparable across projects/regions
20	Needs to be comprehensive - covering global
21	Clear reporting standards to be referenced
22	Clear business case logging
23	Source of CO_2 and if possible biogenic content
24	Information on reservoir would also be very useful
25	Project targets should be included where available
26	Boundaries of the data need to be clear
27	Identifiability - over the longer term
28	Understandable - so users can pick it up and use it
29	Collecting PVT of reported numbers is important
30	Acknowledgement of missing data
31	Also have qualitative data associated with projects
32	It would be valuable to have understanding of the uncertainty in the data

33	Data can be updated or back-revised
34	Accurately indicate the missingness of data
35	It requires the prompt support and engagement of stakeholders to sharing
36	Business cases need to be clearly logged e.g. EOR vs pilot gov funded vs 45Q etc.
37	Documented standards by which numbers have been generated such that definitions, e.g, of storage, and assurance in the project operation can be understood

A.3 Data quality risks

This section includes initial ideas gathered via interactive whiteboard for risks of poor quality data, including causes, events or impacts (where all three constitute a risk).

ID	Risk
1	Loss of data due to public sources being removed
2	Data with lower quality used as the basis for narratives
3	Reported data misunderstood/misinterpreted
4	Discrepancies between sources for the same project
5	Manual errors in gathering/processing
6	Numbers adjusted retrospectively by operators
7	Data in register is misattributed in time
8	Data requires reprocessing for a particular use
9	Reporting standards used by operators do not provide sufficient detail
10	Reporting standards are missing
11	Injection numbers reported, but if a leak occurs the boundary becomes important
12	Some operators report capture capacity not injected volumes
13	Operator does not differentiate between permanent storage and EOR
14	Notable project absences from the dataset
15	Missing data for incomprehensive annual performances
16	Project reporting stops becoming publicly available
17	Disaster caused the sudden closure of the basin or injection sites
18	Acknowledgement of missing data (non-reporting) and also projects that only report via personal communications
19	Data used by 3rd parties to make statements not supported by the data
20	In acid gas removal projects - sometimes the non-CO_2 component is not quantified clearly
21	Over reporting etc.
22	Reputation damage

APPENDIX B – DRAFT DATA QUALITY RULES

This section includes draft data quality rules, to be considered for expansion and adoption into the project process, where the register is reviewed against these rules prior to publishing register updates. These are indicative and would require refining before use – including the prerequisite of a metadata specification for syntactic rules, and detailing of pragmatic rules so they can be assessed.

Rule ID	Rule Description	Category
S-01	Primary key (e.g. CO_2 Storage Project, Data source) is unique across the dataset.	Syntactic
S-02	Primary key (e.g. CO_2 Storage Project, Data source) contains no not null or empty entries.	Syntactic
S-03	All year columns (1996-2024) contain only numeric values or null.	Syntactic
S-04	Valid country names: "Country" values match a predefined list of valid country names (e.g. ISO 3166).	Syntactic
S-05	There are no rows where all storage values are null.	Syntactic
S-06	All categorical fields are from a valid defined set.	Syntactic
S-07	Declared precision/scale: all CO_2 storage values reflect the maximum precision in the stated source.	Syntactic
S-08	Data should match expected, defined formats.	Syntactic
Se-01	There are no non-zero storage values before stated “Starting Year”, unless a reason is stated.	Semantic
Se-02	All storage values are zero, positive or null.	Semantic
Se-03	There is a realistic change in CO_2 storage value year to year.	Semantic
Se-04	When a source reports multi-year cumulative CO_2 storage, the dataset records the correct annualised average and documents the reporting period.	Semantic
Se-05	All project-year records include a valid data source reference.	Semantic
Se-06	Project metadata (starting year, CO_2 source, storage type etc) has a valid data source reference.	Semantic
P-01	There is a single documented methodology which reflects the latest data gathering and transformation for the register.	Pragmatic
P-02	Referenced data sources are accessible to users.	Pragmatic
P-03	Any non-trivial narrative is explained in the project notes.	Pragmatic

APPENDIX C – DATA EVALUATION

Characteristic /14/ and principle [section 5.3]	Strengths	Limitations
Completeness *The degree to which subject data associated with an entity has values for all expected attributes and related entity instances in a specific context of use.* **Principle 1: The data is sufficiently comprehensive, covering all major operational subsurface CO_2 storage projects globally.**	1. Leveraging a project consortium ensures multiple inputs and perspectives, enhancing coverage and reducing blind spots.	1. Some projects or regions do not publish openly, making reliable data acquisition difficult.
	2. Public availability of data enables ongoing updates and gap-filling. Publication fosters informal peer review, improving credibility.	2. Capturing all relevant information is challenging, particularly for historical changes or when multiple sources provide conflicting data.
	3. Aggregated totals can be compared against external estimates for validation and benchmarking.	
	4. Integration of a wide range of sources contributes to a more complete and balanced representation of global CO_2 storage activities.	
	5. Inclusion of survey data adds depth and provides additional context, often accompanied by supporting datasets.	

Characteristic /14/ and principle [section 5.3]	Strengths	Limitations
Understandability *The degree to which data has attributes that enable it to be read and interpreted by users, and are expressed in appropriate languages, symbols and units in a specific context of use* **Principle 2: The data is sufficiently well documented and structured so that a target users can utilise it easily, and without additional support.**	1. The collaborative approach via the project consortium, combined with explicit circulation prior to published update, provides a screening mechanism for clarity and interpretability.	1. In several cases, additional notes would improve clarity, particularly where assumptions or prioritisation decisions influence data interpretation.
	2. The existence of a data column for explanatory notes against data sources can add valuable context, clarifying assumptions and aiding user understanding.	2. A formal methodology document is lacking, leaving decision-making processes—such as handling assumptions or prioritization—opaque to users. There are different sources of commentary, including the register website, the published annual report, and referenced academic papers. The register would benefit from a definitive methodology document, accompanying the dataset.
		3. The use of a spreadsheet limits the ability to embed layered metadata, reducing opportunities for richer contextualization. The use of colours for key metadata (indicating an averaged value, or the prioritised source) also limits the useability, including for machine-read application.
		4. Variations for null values include blanks, asterisks (*), and forward slashes (/) which may cause confusion and hinder consistent interpretation.
		5. The current data structure is not machine-readable, restricting automated processing and integration with analytical tools.

Characteristic /14/ and principle [section 5.3]	**Strengths**	**Limitations**
Traceability *The degree to which data has attributes that provide an audit trail of access to the data and of any changes made to the data in a specific context of use.* **Principle 3: The data sources and processing methodology is sufficiently well documented to inspire trust in the broad community and enable understanding of limitations for individual use cases.**	1. Inclusion of public URLs for data sources is highly beneficial, providing transparency and enabling users to verify original materials.	1. Without a published methodology, it is difficult for users to replicate processes or understand decision-making steps.
	2. The inclusion of annotations to add some context to the processing aids user understanding where the register does not directly match source data.	2. Lack of auditable processing steps reduces traceability and inhibits validation. This could include noting where in a given report values were taken from, and intermediate processing steps – for example conversion of cumulative storage to annual rates.
	3. The maintenance of an archive of the data sources is highly beneficial, allowing reproduction of the register values from the raw data source (assuming any transformation steps are provided).	4. Failure to document data transformation steps limits scalability and makes quality checks challenging.
		5. Reliance on self-reported data and unpublished sources introduces uncertainty and reduces confidence.
		6. More granular metadata would improve clarity, especially where data is compiled from multiple reports but referenced under a single source.
		7. Referenced URLs can change or be removed, as observed during sample checks, undermining long-term traceability. In addition, many of the references are to webpages hosting multiple documents, including multiple different years. It is not always evident which report produced which value.
		8. Users cannot assess data reliability due to the absence of documented confidence rules or assurance evidence for each project. Additionally, reference standards lack clarity, with varying definitions of storage and measurement methods.

Characteristic /14/ and principle [section 5.3]	Strengths	Limitations
Accuracy *The degree to which data has attributes that correctly represent the true value of the intended attributes of a concept or event in a specific context of use.* **Principle 4: The data reflects the truest available representation of injection rates for each project.**	1. Multiple data sources are often included for the same projects, which contributes to a picture of the "true" storage value, where different accounts are published.	1. Sample reviews revealed inconsistencies in reported values, highlighting the need for embedded quality checks within the register team during data collection and updates.
	2. Inclusion of operator surveys contributes to this picture, giving an additional, high-value source.	2. Many entries lack the precision of the source, reducing confidence in the reported figures.
	3. The register indicates where assumptions or calculations (such as averaging) have been applied, providing some visibility into data handling.	3. No formal quality assurance process is integrated into the workflow, increasing the risk of manual errors.
		4. Beyond project references, there is limited clarity on what the reported variable represents, which may lead to misinterpretation.

About DNV

DNV is the independent expert in risk management and assurance, operating in more than 100 countries. Through its broad experience and deep expertise DNV advances safety and sustainable performance, sets industry benchmarks, and inspires and invents solutions.

Whether assessing a new ship design, optimizing the performance of a wind farm, analyzing sensor data from a gas pipeline or certifying a food company’s supply chain, DNV enables its customers and their stakeholders to make critical decisions with confidence.

Driven by its purpose, to safeguard life, property, and the environment, DNV helps tackle the challenges and global transformations facing its customers and the world today and is a trusted voice for many of the world’s most successful and forward-thinking companies.