

## National Underground Asset Register (NUAR)

### NUAR Data Model Publication

## A Data Quality Maturity Assessment Toolkit for data in the NUAR Harmonised Data Model

### Publication History

Date	Version	Description
13/06/2025	0.1	First draft of a NUAR Data Quality Maturity Assessment Toolkit

### Purpose of this document

This document summarises how elements of data quality could be measured in terms of conformance to the NUAR Harmonised Data Model.

This example toolkit is provided for data professionals in asset owner organisations eligible to participate in NUAR to support their work on the measurement and improvement of quality of their data, for those engaged in digital transformation projects involving variable quality data, and for those with a general interest in data quality assessment and improvement.

The purpose of any NUAR Data Maturity Model is ultimately to support the consistent, frequent supply of good, and improving, quality data to the NUAR service, and not to replace or replicate other maturity models which may be employed, of which this represents a small subset focused on data quality in the NUAR ecosystem.

If data quality assessments are already in place in an organisation, these should easily adapt and map onto the subset described here for the supply of data to NUAR.

## Background and General Principles

- The purpose of a NUAR Data Quality Maturity Assessment toolkit is to provide a benchmark based on the NUAR Harmonised Data Model for asset owner organisations, to support efforts to identify areas where data quality improvement efforts could be focused.
- Any Data Quality Maturity score that an organisation may calculate using the criteria outlined below reflects elements of quality for data relevant to NUAR. Aspects of data quality which may be objectively measured include:
  - Geometric conformance
  - Use of unique and/or persistent identifiers
  - Completeness (of datasets and attribution)
  - Domain Consistency (use of codelist values)
  - Currency
  - Ability to detect change relative to a baseline date

In the future, some of these measures may be automatically measured and reported on as part of the value of NUAR participation, but for now these are presented as measures which organisations may be able to measure themselves if a data quality maturity assessment would be useful.

## **A Data Quality Maturity Model based on conformance to the NUAR Harmonised Data Model**

Organisations can assign a score for Data Quality Maturity based on the criteria outlined below.

The score equates to a Data Quality Maturity level as follows (in descending order of maturity):

- **Level 5:** Champion
- **Level 4:** Exemplar
- **Level 3:** Good Citizen
- **Level 2:** Improver
- **Level 1:** Participant

### **Data Quality Maturity**

#### **Level 5: Champion**

- Availability of comprehensive attributes describing the physical properties of their assets and asset lifecycle, currency and quality
- Unique persistent identifiers are assigned to all features in all datasets
- Close to 100% conformance to relevant codelist values in the NUAR Harmonised Data Model
- All qualifying datasets are digitised and could be provided to NUAR
- Changes to data relative to a baseline date can be identified at the dataset, area or feature level

#### **Level 4: Exemplar**

- Availability of comprehensive attributes describing the physical properties of their assets and some attributes describing asset lifecycle, currency and quality
- Unique identifiers are assigned to all features in all datasets but may not be persistent in all cases
- Strong, and improving, conformance to relevant codelist values in the NUAR Harmonised Data Model
- All qualifying datasets are digitised and could be provided to NUAR
- Changes to data relative to a baseline date can be identified at the dataset level and, for datasets with high rates of change, at the area or feature level

### **Level 3: Good Citizen**

- Availability of comprehensive attributes describing the physical properties of their assets
- Unique identifiers are assigned to features in most datasets but may not be persistent
- Reasonable conformance to relevant codelist values in the NUAR Harmonised Data Model, with some improvement observed over time
- All qualifying datasets are digitised, and the vast majority could be provided to NUAR
- Changes to data relative to a baseline date can be identified at the dataset level

### **Level 2: Improver**

- Availability of some attributes describing the physical properties of their assets
- Identifiers are assigned to features in some datasets but may not be unique and unlikely to be persistent
- Some conformance of attribute values to relevant codelist values in the NUAR Harmonised Data Model
- The vast majority of qualifying datasets are digitised, but not all are in a state to be provided to NUAR for direct inclusion on the combined map (although at least indicative geo-referenced locations of assets or sites could be provided)
- Changes to data relative to a baseline date can be identified at the dataset level for some datasets

### **Level 1: Participant**

- Availability of the minimum level of information to NUAR, about the location and ownership of their assets
- Identifiers are not assigned to features
- Minimal or no conformance of any attribute values to relevant codelist values in the NUAR Harmonised Data Model
- Not all qualifying datasets are digitised, or in a suitable form to provide to NUAR for direct inclusion on the combined map (although at least indicative geo-referenced locations of assets or sites could be provided)
- Changes to data relative to a baseline date cannot easily be identified

## Data Quality Maturity Measures

Organisations must adhere to at least **8** of the measures specified for a level in order to achieve that level, with the relevant Digitisation level being a mandatory element for all maturity levels. As per the requirements for **Level 1: Participant** above, the minimum data requirement is the location and ownership of assets, meaning that at least indicative geo-referenced locations of assets or sites must be provided, even if not all data is vectorised.

Ref	Measure	Level 1	Level 2	Level 3	Level 4	Level 5
1	Digitisation level <sup>†</sup>	> 50%; < 80%	80%	100%	100%	100%
2	Vectorisation level <sup>‡</sup>	< 75%	75%	100%	100%	100%
3	Completeness of category/type attribution (asset type, voltage level etc)	< 80%	80%	95%	100%	100%
4	Completeness of physical description attribution (material, diameter etc)	< 50%	50%*	80%	95%	100%
5	Completeness of depth/elevation attribution <sup>‡</sup>	< 5%	5%*	25%	50%	>= 75%
6	Completeness of temporal attribution	< 25%	25%*	50%	80%	>= 95%
7	Completeness of data quality attribution	< 5%	5%*	25%	50%	>= 75%
8	Completeness of lifecycle attribution	< 5%	5%*	25%	50%	>= 75%
9	% of compliant geometries (basic geometry validation checks)	< 75%	75%*	85%	95%	>= 99%
10	% of datasets with asset owner supplied unique identifiers	< 50%	50%	80%	100%	100%
11	% of eligible features with unique identifiers populated	< 50%	50%*	75%	90%	100%
12	% of eligible attributes with conformant codelist values	< 25%	25%*	50%	75%	>= 90%
13	Changes to data relative to a baseline date can be established at the dataset level	✗	✗	✓	✓	✓
14	Changes to data relative to a baseline date can be established at the area or feature level	✗	✗	✗	✗	✓

<sup>†</sup> Digitisation level is measured as the percentage of relevant datasets that are in a digital format that could be shared with NUAR and is mandatory for attainment of all maturity levels.

<sup>‡</sup> Vectorisation level is measured as the percentage of qualifying “core” datasets (i.e. datasets that would be represented directly on the NUAR combined map as opposed to supplementary or supporting data) that are in a vector format as opposed to CAD drawings, raster files or other digital document formats

\* Or greater than 10% annual improvement

‡ It is likely that a more sophisticated measure will initially be required than simply measuring population of “Depth” attribution for linear features. Many organisations indicate depth of linear assets via point features indicating depth at a location, so possibly the density of these features per linear length, plus population of the Depth attribute of relevant point features, should be used to measure this element of maturity. Use of network descriptions to indicate depth should only allow a maximum of Level 3 or 4 to be achieved.

### ***What do these measures mean?***

The table below outlines what the practical implications are of supplying good quality data to NUAR.

Ref	Measure	What this means to you and NUAR users
1	Digitisation level	Only data which is available in a digital form can be loaded into NUAR, so any data that isn't in a digital form will not be available to NUAR users to support safe excavation.
2	Vectorisation level	Data in a vectorised form can be displayed in the combined map alongside the other datasets in NUAR, and users will be able to interact with the data to view important details about the data. Any data in a non-vectorised digital format may be available in NUAR but will not be seamlessly integrated with the other datasets.
3	Completeness of category/type attribution (asset type, voltage level etc)	The type of buried asset, and categorisations of voltage level, pressure tier, temperature etc. are very important pieces of information which might lead to different precautions and ways of working to be adopted by those carrying out excavations.
4	Completeness of physical description attribution (material, diameter etc)	Like categorisations of buried assets, their physical characteristics, such as material and diameter, are crucial pieces of information which may dictate the level of risk associated with an excavation, and the necessary precautions to circumvent those risks.
5	Completeness of depth/elevation attribution	Knowledge of the depth below the surface of a buried asset is invaluable to those planning and executing excavations.
6	Completeness of temporal attribution	Information about support for date and time related metadata in the NUAR Harmonised Data Model can be found <a href="#">here</a> . Information about when an asset was installed, surveyed, uploaded etc. can provide useful context to those locating that asset, and inform users about how up to date the data they are viewing is.



7	Completeness of data quality attribution	Information about support for data quality and accuracy metadata in the NUAR Harmonised Data Model can be found <a href="#">here</a> . Information about how the location of an asset was measured, and the accuracy of measurements helps to inform users of NUAR about the level of confidence they may have in the data being presented.
8	Completeness of lifecycle attribution	The operational status of an asset provides important context, and useful information when faced with ambiguous and complex conditions on site.
9	% of compliant geometries (basic geometry validation checks)	This information is useful to providers of data to NUAR to inform them about the geometric conformance of their data relative to established standards. Non-conformances in geometry can lead to mis-leading visualisation of data in all software, not just NUAR.
10	% of datasets with asset owner supplied unique identifiers	Having unique identifiers for features in a dataset supports all sorts of useful capabilities in NUAR and beyond. For example, it may help to track changes to a given asset feature over time and allows Observations about an asset feature to persist across any updates to that feature.
11	% of eligible features with unique identifiers populated	See above for the benefits of unique identifiers.
12	% of eligible attributes with conformant codelist values	Conformance to values in the NUAR codelist register gives an indication of how consistent descriptions for various aspects of buried assets are. Where descriptions and categorisations are represented as free text, measuring domain conformance can highlight potential spelling mistakes and inconsistencies of terminology. It can also indicate where the NUAR codelist register is deficient and lead to improvements.
13	Changes to data relative to a baseline date can be established at the dataset level	Being able to determine when and in what way a dataset has changed can be useful in measuring data currency and informing processes for updating data in NUAR and other systems. It may also help outside the NUAR context, for example to measure the rate of change of data in order to report on the effectiveness of operational processes.
14	Changes to data relative to a baseline date can be established at the area or feature level	Being able to determine when and in what way individual features, or subsets of features have changed allows physical changes to be tracked over time, or improvements in data quality to be measured, as well as potentially allowing smarter approaches to data updates to be explored.

### ***Calculation of Data Quality Maturity Measures***

The table below outlines the ways in which the listed data quality measures could be calculated using information stored in the NUAR Harmonised Data Model.

Ref	Measure	Method of Calculation
1	Digitisation level	% of relevant datasets that are maintained in digital form
2	Vectorisation level	% of relevant datasets that are in vector format
3	Completeness of category/type attribution (asset type, voltage level etc)	% of features in relevant datasets that have a clear indication of the type of asset, including pressure tier/voltage level as appropriate
4	Completeness of physical description attribution (material, diameter etc)	% of features in relevant datasets that have clear indications of Material and physical dimensions
5	Completeness of depth/elevation attribution	% of features in relevant datasets that have depth attribution
6	Completeness of temporal attribution	% of features in relevant datasets that have information recorded about date of installation and last survey
7	Completeness of data quality attribution	% of features in relevant datasets that have information recorded about the method of capture, accuracy and/or quality level
8	Completeness of lifecycle attribution	% of features in relevant datasets that have a valid Operational Status value
9	% of compliant geometries (basic geometry validation checks)	% of features in relevant datasets with a valid geometry
10	% of datasets with asset owner supplied unique identifiers	% of relevant source datasets that have a unique identifier attribute
11	% of eligible features with unique identifiers populated	% of features in relevant datasets that have a unique identifier assigned
12	% of eligible attributes with conformant codelist values	% of codelist-driven attributes in features in relevant datasets which have values present in, or easily mapped to, NUAR Codelist values. Eligible attributes are those identified above for indicating type/category, physical description and data quality
13	Changes to data relative to a baseline date can be established at the dataset level	Relevant datasets can be identified as containing changes relative to a baseline date.
14	Changes to data relative to a baseline date can be established at the area or feature level	Features in relevant datasets can be individually identified as having changed relative to a baseline date, or features aggregated to a geographical area can be identified as having change relative to that baseline.





### Example Assessment

Ref	Measure	Score	Level	To achieve next level
1	Digitisation level	100%	5	N/A
2	Vectorisation level	100%	5	N/A
3	Completeness of category/type attribution (asset type, voltage level etc)	95%	3	100%
4	Completeness of physical description attribution (material, diameter etc)	85%	3	95%
5	Completeness of depth/elevation attribution	30%	3	50%
6	Completeness of temporal attribution	75%	3	80%
7	Completeness of data quality attribution	10%	2	25%
8	Completeness of lifecycle attribution	25%	3	50%
9	% of compliant geometries (basic geometry validation checks)	82%	2	85%
10	% of datasets with asset owner supplied unique identifiers	80%	3	100%
11	% of eligible features with unique identifiers populated	85%	3	90%
12	% of eligible attributes with conformant codelist values	75%	4	90%
13	Changes to data relative to a baseline date can be established at the dataset level	✗	2	✓
14	Changes to data relative to a baseline date can be established at the area or feature level	✗	4	✓
OVERALL				Level 3