Appendix M – Data Specification Outline

Description and Purpose:

This document contains an outline proposal for the structure and content of a technical Data Specification that will support a national roll-out of the NUAR platform by standardising how data is provided to it. Alongside the NUAR Harmonised Data Model, it will give clarity to Asset Owners about how to supply data to the NUAR platform.

A technical Data Specification based on this outline will be produced during the Build Phase and is intended to assist and guide organisations in supplying data to the NUAR Platform in a repeatable and sustainable manner. It describes the different methods of supplying data to NUAR, and the relevant data formats, schemas, supply methods and frequency that will assist organisations as they create the necessary processes, transformations, and services to contribute their data to NUAR.

This document represents a proposed outline for the areas that a technical data specification should cover for a national roll-out of the NUAR Platform, and will be refined by the Supplier during the Build Phase.

Contents

1. Glossary of Terms

Term	Definition
Active Region(s)	A region that is “live” in the NUAR Platform, i.e. a Region in which users are able to access the data on the platform to identify underground assets in support of the planning or execution of excavations
BNG	British National Grid. A coordinate reference system that uses the OSGB36 (EPSG 27700) geodetic datum and a single Transverse Mercator projection for the whole of Great Britain. Positions on this projection are described using easting and northing coordinates in units of metres
CAD	Computer-aided Design. A type of software to create precision drawings or technical illustrations. CAD software can be used to create two-dimensional (2-D) drawings or three-dimensional (3-D) models, and is typically used by architects, engineers, and drafters
Data Asset Owner	An organisation that manages Assets that are to be displayed through the NUAR UI
Data Model	An abstract model that organises elements of data and standardises how they relate to one another and to the properties of real-world entities
Data Provider	An organisation that is supplying data to the NUAR platform – a Data Asset Owner or a service provider engaged by them.
DDA	Data Distribution Agreement
GeoJSON	A geospatial data interchange format based on JavaScript Object Notation (JSON). It defines several types of JSON objects and the manner in which they are combined to represent data about geographic features, their properties, and their spatial extents. GeoJSON uses a geographic coordinate reference system, World Geodetic System 1984, and units of decimal degrees.
Geospatial Commission	The Geospatial Commission is an expert committee that sets the UK’s geospatial strategy and promotes the best use of geospatial data
GeoTiff	Geographic Tagged Image File Format. An image file format that includes additional spatial (georeferencing) information embedded in the metadata of the tiff file as tags
GIS	Geographic Information System. A software system used to capture, store, manage, analyse and visualise geospatial data
GML	Geography Mark-up Language. An open standard for expressing geographical features in XML grammar
MUDDI	Model for Underground Data Definition and Integration. Conceptual model developed by the OGC for representation of subsurface features

Term	Definition
NUAR Regions	The target geography of the NUAR platform is divided into regions, which enable the functionality of the NUAR platform to be rolled-out on a region by region basis as data for each region reaches an acceptable level of completeness. The regions also enable granular reporting on aspects such as usage, quality and value released, as well as the targeting of activities such as data quality improvement drives
Observation	In the context of NUAR, this refers to the capturing of valuable information that will increase the accuracy and currency of data on the platform, making it better able to meet the requirements of end users
OGC	Open Geospatial Consortium. An international standards organisation that collaborates on open standards for geospatial content, services, IoT, and data processing and sharing. In the context of this specification document the OGC are responsible for the creation and maintenance of the GML Standard
PNG	Portable Network Graphics. A compressed raster image format
Raster Data	A data structure that represents a grid of pixels and stored in image files in a variety of formats
Schema	Schema refers to the structure of a file. A schema definition file is an expression of the allowable or valid structure of, in this case, a GML file.
Vector Data	In the context of spatial data, Vector data is comprised of vertices and paths, allowing suitable software to render the geometric elements, the three basic elements being points, lines and polygons. Attributes are typically stored alongside geometric elements
xsd	XML Schema Definition. See Schema definition

2. Introduction

This document contains an outline proposal for the structure and content of a technical Data Specification that will support a national roll-out of the NUAR platform by standardising how data is provided to it. Alongside the NUAR Harmonised Data Model, it will give clarity to Asset Owners about how to supply data to the NUAR platform.

A technical Data Specification based on this outline will be produced during the Build Phase and is intended to assist and guide organisations in supplying data to the NUAR Platform in a repeatable and sustainable manner. It describes the different methods of supplying data to NUAR, and the relevant data formats, schemas, supply methods and frequency that will assist organisations as they create the necessary processes, transformations, and services to contribute their data to NUAR.

This document represents a proposed outline for the areas that a technical data specification should cover for a national roll-out of the NUAR Platform, and will be refined by the Supplier during the Build Phase.

The challenge presented by the broad range of data storage formats, types of data, and range of levels of data management maturity amongst Data Asset Owners, will be met by striking a balance between “guidance” to enable data to be delivered to the platform with minimal barriers, and “specification” required to provide clarity for data providers that will result in a platform that can scale nationally.

This document describes the way data quality may be considered in the NUAR context, the options for data supply and the requirements around each of the options.

A proposal for a set of conformance classes for Data Asset Owners is outlined. These classes detail a series of requirements that build on one another that are intended to drive the NUAR platform from the minimum viable level of providing data that meets the safe excavation use case, to a richer more complete user experience that will support scaling to a national solution and improve the data quality, at source, of underground asset data as it does so.

Ultimately the specific metrics associated with each of these conformance classes will be jointly “owned” by the Geospatial Commission, the Supplier during the build phase and by the Data Provider and User communities more broadly.

As the NUAR program develops, there are areas that will be developed and further explored that will impact the content and approach described in this document. These can be summarised as follows:

- Data Model Management Processes, in particular how Codelist extension requests are made, considered and then implemented. This process will be defined as part of Data Model governance during the build phase.
- Detailed requirements for the supply of Organisation metadata – currently outlined in this version as ‘Organisation Data Supply’.

- Establishing bespoke data ingestion specifications in collaboration with asset owners that will document a repeatable supply of data, including source data formats and structures, transformation mappings, conformance levels and supply frequency. This will impact the Data Quality section of this document.
- The manner in which data updates can be supplied to the NUAR platform. This will impact the Supply of Change section of this document.

2.1 Key Concepts

Model for Underground Data Definition and Integration (MUDDI): The Open Geospatial Consortium (OGC) are developing a standard for modelling underground assets. This is a conceptual model that has been designed with the intention of being the basis for the creation of interchangeable logical and physical implementations of underground assets. NUAR has adopted MUDDI as the core of its data model in anticipation of MUDDI's adoption as an OGC open standard.

NUAR Harmonised Data Model (NHDM): This is the data model that describes the data stored in and made available through the NUAR platform. It gives the NUAR platform a way of presenting data in a consistent, clear, and comprehensive way, despite the fact that it is sourced from multiple different asset owners, each managing data in different ways, with a variety of software.

The data model has been extended from the MUDDI model through collaboration with asset owner organisations throughout the first phase of the NUAR programme, and over time is likely to go through further evolution. This Data Specification will therefore be impacted by changes to the NUAR Harmonised Data Model.

A comprehensive definition of the NHDM is included in the supplied documentation.

2.2 The Function of the NUAR Data Specification

A NUAR Data Specification should:

1. Give data owners **definitive guidance on data expected** (types, content and structures) from a data provider for inclusion into NUAR.
2. Give data users **detailed information about the data that is being ingested or made available** through the NUAR platform.
3. State the **minimum compliance conditions** for data to be accepted to the Data Model
4. State **conformance requirements** and define criteria for dealing with **non-conformant** data (e.g. accepting data but feeding back to data suppliers to enable data quality improvement). These requirements and criteria may vary depending on the Data Maturity Level of the supplier and the type of data.
5. Provide definitions and explain the **purpose** of the various elements of the Data Model in terms of supporting the functionality of the platform.

2.3 Requirements

The purpose of this document is to guide and make clear the means and details of providing data to the NUAR platform. As such, where clear requirements are expressed in this document, they will be highlighted in the following way:

Requirement Identifier

Requirement Title

The requirement will be described here

2.4 Normative References

Table 1 below lists the normative documents that are referenced in this specification.

Table 1: Normative References

Document	Version	Date	Description
Geography Markup Language (GML) Encoding Standard	3.3		The specification for expressing geographical features in XML grammar
GeoJSON specification	RFC 7946	Aug 2016	The specification for expressing geographical features in JSON syntax
ISO19157:2013 – Geographic information – Data quality	Edition 1	2013	The standard establishing the principles for describing the quality of geographic data
PAS128:2014		2014	Specification for underground utility detection, verification and location
NUAR Harmonised Data Model	1.2		Logical UML model of the NUAR data. Included in the supplied documentation
NUAR_HarmonisedModel_v1-2.xsd	1.2		XML schema definition for the data to be supplied in a NUAR Harmonised Data Model compliant way. Included in the supplied documentation.

2.5 Target Audience for the Data Specification

A technical Data Specification based on this outline is a reference for organisations that supply data to the NUAR platform.

Figure 1, below, shows that, for a Data Asset Owner, there are a number of potential methods of supplying data to the NUAR platform. The following flow-chart will help Data Asset Owners to understand which is the most appropriate for them. The section on Supply Methods then outlines each method and guides the reader through the appropriate sections of this Specification.

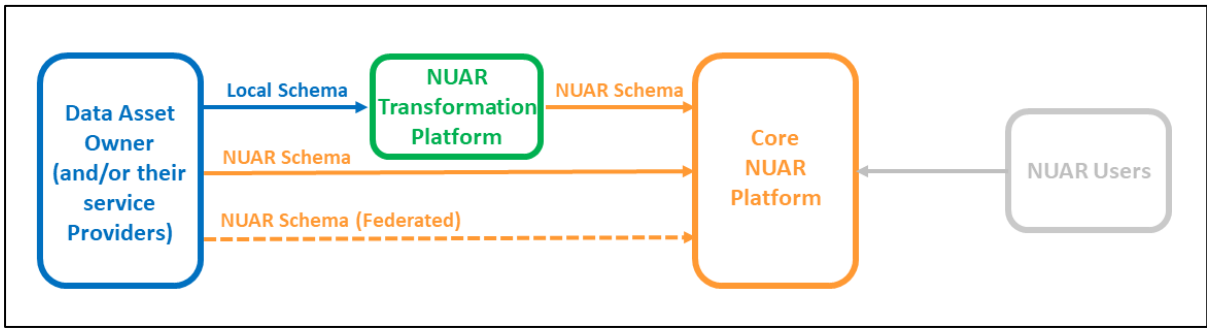

Figure 1: Supply Methods

2.6 Supply Methods

There are a number of options available to asset owners when considering how to provide their data to the NUAR platform. Ultimately, all data ingested into the core NUAR platform must be transformed into the target NHDM schema, but there are a number of routes via which that transformation may take place.

Local Schema Data Supply: This refers to the direct supply of data to the NUAR Transformation Platform developed by the NUAR Supplier that translates the inbound data into the target NUAR data model schema suitable for ingestion into the NUAR platform. The data schema and data format supplied will reflect the way the data is stored locally, and must be provided in a consistent, repeatable form defined during the onboarding process and codified in a bespoke data ingestion specification documented for, and agreed by, that asset owner. Transformation to the NUAR Harmonised Data Model compatible schema will be performed by the NUAR Transformation Platform which is conceptually part of the “NUAR System” as far as the asset owner is concerned. Significant collaboration will be required between the Data Asset Owner and the team responsible for data transformation to understand how the data can be transformed from local to NUAR schema. Refer to the **Local Schema Data Supply** section.

NUAR Schema Data Supply direct from the asset owner: Some asset owners may perform the transformation to the target NHDM schema “on-premise” (or via a service provider engaged by them) and supply this transformed data directly into the Core NUAR Platform. These asset owners will be able to keep close control over the processing of their data, and will be able to avail themselves of validation services exposed by the platform.

All data supplied will have been transformed from local schema into a consistent schema and format specified in the **NUAR Schema Data Supply** section.

Federated Data Supply: This is the supply of data through an OGC compliant web service. Data Providers may also supply periodic updates of data to be hosted by NUAR that will be served through the UI in the case that the API is unavailable. The key sections for this type of data supplier are:

- **Federated Data Supply** which explains the requirements for the API.
- **NUAR Schema Data Supply** which explains the schema to which the data must conform.

3. Data Quality

The quality of the data presented to end users is critical to the success of the NUAR platform. This section explains the NUAR approach to data quality, which includes definitions of the terms and how they apply to features, datasets and data providers. It also explains the breakdown of data conformance, from the minimum viable contribution of data to the NUAR platform that supports the safe excavation use case, to the highest level contribution that will provide the richest experience to the users.

The following is a list of the data quality dimensions used in NUAR, which are then defined and explained further on in this section:

- Positional Accuracy
- Completeness
- Domain Consistency
- Currency

Where appropriate ISO19157 terms are used, otherwise the dimensions are created specifically for the NUAR context.

This section on data quality includes an overview of the NUAR approach to:

- Identifiers
- Geometry

3.1 Positional Accuracy

ISO19157: Accuracy of the position of features within a spatial reference system.

This breaks down into two data quality elements with which NUAR is concerned:

- Absolute positional accuracy: closeness of reported coordinate values to values accepted as or being true.

- Relative positional accuracy – closeness of the relative positions of features in a data set to their respective relative positions accepted as or being true.

A NUAR approach to accuracy metadata needs to be agreed with the asset owner community, and while it is desirable to understand the accuracy of all of the data that the NUAR platform displays, this may not be realistic. For instance, accuracy could be measured and expected only for assets that are laid or resurveyed beyond a certain date.

The NUAR Harmonised Data Model includes several metadata attributes that indicate how closely the geometric data representations align with the real-world assets.

- **MUDDIObject:qualityLevel:** This is the quality level in accordance with PAS128:2014. This field should only be populated if the **dateDataCollected** is populated with the data of a PAS128-compliant utility survey, and dataQuality should reflect the Quality Level assigned during the survey performed at that time. Values for qualityLevel are defined in the <QualityLevelValue> codelist (see the **Feature Types and Attributes** section for details). If the last identified survey was not conducted according to the PAS128 standard this value must be left unpopulated.

- **MUDDIObject:dateDataCollected:** Date when a utility feature was last surveyed in the field. If this survey was carried out according to the PAS128 standard, the Quality Level assigned to the feature should be recorded in the **qualityLevel** field. If the survey on this date was not carried out according to PAS128, the **qualityLevel** field must be left unpopulated.

- **MUDDIObject:horizontalAccuracy:** If the last survey of the feature was carried out according to PAS128, the **horizontalAccuracy** field value should represent the horizontal 'Location accuracy' associated with the Quality Level assigned to this feature, expressed in metres. If the last survey was not carried out according to PAS128, this field should represent the worst-case 'margin of error' reported by the surveying equipment used at the time, expressed in metres. This field should only be populated if objective data about horizontal accuracy is available.

- **MUDDIObject:verticalAccuracy:** If the last survey of the feature was carried out according to PAS128, the **verticalAccuracy** field value should represent the vertical 'Location accuracy' associated with the Quality Level assigned to this feature, expressed in metres. If the last survey was not carried out according to PAS128, this field should represent the worst-case 'margin of error' reported by the surveying equipment used at the time, expressed in metres. This field should only be populated if objective data about vertical accuracy is available.

- **MUDDIObject:horizontalMeasurementMethod:** This field is intended to distinguish between features whose horizontal position was captured via measurement relative to an above ground feature and those whose absolute coordinates were captured using surveying equipment. A value of "Unknown" should be assumed to represent a relative measurement as per the traditional means of capturing horizontal positions.

3.2 Completeness

ISO19157: Completeness is defined as the presence and absence of features, their attributes and relationships. It consists of two data quality elements:

- commission: excess data present in a dataset;
- omission: data absent from a data set.

For NUAR, Completeness measures “omission” in the following contexts:

1. Completeness of Attribution: Within the data supply from a Data Asset Owner, how well populated are attributes with acceptable values?

2. Completeness of Features: The proportion of features from a Data Asset Owner that are successfully loaded into the NUAR platform

3. Completeness of Data Supply: For a given area in which NUAR data is held, the proportion of data that has been received from known asset owners

We will need asset owners to provide information about coverage, which is effectively a statement to the effect that “We have provided all the data that we hold that falls within this area of coverage”. This expression of coverage may be presented in terms of their own operational boundaries, in terms of the coverage of the NUAR platform, or in other terms (e.g. a region or local authority boundary).

3.3 Domain Consistency

ISO19157: Adherence of values to the value domains.

In the NUAR context this refers to how comprehensively codelist terms have been adopted in the transformation from the local schema to the NUAR Harmonised Data Model.

As a key NUAR data principle, the addition of new required values to Codelists should be as responsive and frictionless as possible. However, there may still be occasions where it is not possible to map a source value onto a standardised codelist value. In this case it may be appropriate to assign a value of “Other”.

This should generally be a last resort option to be taken after consideration of:

- A correction at source for values which are clearly incorrect (e.g. incorrectly spelt material values)
- The time it will take for the addition of a new codelist value for valid source values which are not currently represented in the data model

Any mapping to “Other” values will be fed back as part of conformance reporting against the Consistency category (see above).

“Other” values may not be accepted unless a Codelist extension request has been made. The validity of “Other” values may be time-limited, to ensure that once a Codelist has been extended as requested, the data from a Data Asset Owner conforms to the new Codelist value(s).

3.4 Currency

Currency is captured against features in a series of metadata attributes that indicate when a real-world asset was surveyed or captured in an asset management system.

- The **originalDateDataCollected** field value should represent the date upon which the feature was originally captured in the source asset owner systems, which may provide useful context. This may be explicitly represented in asset owner data, or may be inferred from a “Date of Installation” or “Date Commissioned” field. The earliest date should be used if in doubt - we are looking to represent the “era” during which the data for the asset was originally captured.

- The **dateDataCollected** attribute represents the date that data was surveyed in the field, as already described in the **Positional Accuracy** section.

An indication of data currency is also maintained against Data Asset Owners and the data that they supply as:

- The frequency of data supply. This will be established and agreed with Data Asset Owners during the on-boarding process

- Change in data assets supplied within an agreed timeframe. The timeframe will be established and agreed with Data Asset Owners during the on-boarding process

- The last supply of data to the platform. This will be recorded as data is loaded onto the platform

- Responsiveness to reported “observations”. The definition of data currency could be extended to include this

NOTE: These Data Asset Owner currency measures will also apply to organisations that provide data through an in-bound API, but the method of assessing currency may differ.

3.5 Identifiers

One of the most significant attributes for the successful maintenance of NUAR data is a unique and persistent identifier for a feature. The benefits of supplying the Data Asset Owner’s unique, persistent feature identifiers include:

- The ability to improve data quality at source through targeted feedback from the data ingestion process
- Gaining the full benefits of the Observations feedback loop
- The possibility of supporting change-only updates
- Ease of interoperability with other systems in the future

However, it is understood that the Asset ID will not always be maintained by Data Asset Owners or be made available to the NUAR platform. Given the benefits listed above, the provision of unique, persistent Asset IDs will be strongly encouraged.

3.6 Geometry

The following table lists example geometry checks which may be carried out on input data. A data validation tool will identify and report invalid geometries with a warning or a rejection. The NUAR Transformation Platform or data ingestion subsystem may be able to automatically fix an invalid geometry in a way that doesn't introduce undesirable or unexpected outcomes in the data. This should always be reported back to the data provider at least as a warning. If it is not possible to repair the geometry in a reliable and acceptable way, the entire feature would not be loaded into the platform, and this should be reported as an ingestion error to the data provider.

Check	**Description**	**Possible Action if invalid**
Contains -0, Nan, or Infinity	The geometry value cannot be NaN (not a number), or infinity	Reject and report
Contains Null Geometry Parts	A geometry is made up of 1 or many parts. None of the parts can be Null	Repair and warn
Degenerate or Corrupt Geometries	A degenerate geometry is one whose geometry type can be simplified. For example, a polygon that has 0 area is degenerate, and can be simplified to a line, point, or Null. A corrupt geometry in contrast contains conflicts in the geometry definition, such as an arc whose angles and endpoints disagree	Reject and report
Duplicate Consecutive Points	The geometry must not contain duplicate consecutive points – This is checked in 2D (x and y) only	Repair and warn
Self-intersections in 2D	The geometry must not contain duplicate points – This is checked in 2D (x and y) only	Reject and report
OGC Simple/Valid Compliant (OGC Version 1.2.0)	The geometry will be evaluated according to OGC standards. See http://www.opengeospatial.org/standards/sfa for more information	Reject and report

4. Conformance

Conformance describes the level to which a feature, dataset, or Data Asset Owner meets the data quality criteria of the NUAR platform.

The NUAR platform defines four classes of conformance, which range from the base level, whereby Data Asset Owners are in a position to submit data to the NUAR platform and have provided data which includes valid mandatory attributes, up to the top level of conformance where data is frequently updated, supplied with persistent, unique identifiers and features include attribution that add richness to the NUAR platform beyond the minimum set of mandatory attributes. Identifying conformance classes and categorising Data Asset Owners in this way enables NUAR to gain a granular understanding of data quality on the platform, and to provide targeted support to Data Asset Owners where required.

Each conformance class builds on the previous one, thus conformance class B meets all the requirements of conformance class A. As such, the requirements stated in this section relate explicitly to the supply of data at each conformance level. An organisation that is supplying data at Conformance Class A does not need to meet the requirements stated in Conformance Class C for example.

An overview of the conformance classes is presented in a matrix below, and the rest of this section describes the detail of the conformance classes.

4.1 Data Asset Owner Conformance Matrix

The table below is a proposal for a matrix that could be used to illustrate the conformance classes that will have to be discussed and agreed with data providers.

The creation and evolution of this matrix should be carried out in close collaboration with the NUAR user community to ensure buy-in and a focus on the metrics that add value from a user’s perspective.

It will also be possible to define a sector-specific breakdown of the proposed data quality measures that is guided by, for example, criticality of infrastructure or data maturity in the sector. Similarly, it may be appropriate to apply some data quality measures only to newly laid or surveyed assets. Where specific values are provided in the table below, these are for illustration only at this stage.

It must be noted, that whether the data from a Data Asset Owner is to be hosted on the NUAR platform, or to be served to the NUAR platform via an API, the same data quality measures should apply.

The concept of “NUAR Standard” in the matrix below refers to the data requirements that do not fall into typical data quality dimensions, but are described alongside data quality dimensions in the matrix. This category covers, for example, the provision of organisation metadata, or the provision of persistent unique identifiers.

Table 2: Example Data Asset Owner Conformance Matrix

Data Quality Category	Measure	A	B	C	D
NUAR Standard	Organisation Metadata is supplied	✓	✓	✓	✓
Currency	Data is updated and supplied to the NUAR platform at least every 3 months	✓	✓	✓	✓
Currency	Full refresh updates are supplied at an agreed frequency of monthly or more frequently	✗	✓	✓	✓
Currency	Change only updates are supplied at the agreed frequency of monthly or more frequently	✗	✗	✓	✓
NUAR Standard	Unique, persistent identifiers for assets are provided by the Data Asset Owner	✗	✗	✓	✓
NUAR Standard	Data is supplied in the NUAR Harmonised Data Model Schema	✗	✗	✗	✓
Currency	Response time for acknowledging Observations	✗	3 months	1 month	2 weeks
Completeness	Completeness of data supplied in Active Regions	99%	99%	99.5%	100%
Completeness	Material attribute completeness	50%	95%	97.5%	99%
Completeness	Asset Type and Size completeness	75%	90%	95%	100%
Completeness	Depth/Elevation attribute completeness	0%	25%	50%	75%
Completeness	Vectorisation level	0%	90%	95%	100%
Completeness	Quality Level attribute completeness	0%	50%	90%	100%
Domain Consistency	Level of codelist value adoption	50%	90%	95%	100%
Positional Accuracy	Range of Horizontal Accuracy values	tbc	tbc	tbc	tbc
Positional Accuracy	Range of Vertical Accuracy values	tbc	tbc	tbc	tbc

4.2 Conformance Class A

Conformance class A deliberately sets the bar for participation in NUAR at a low level to ensure that organisations are able to contribute data and the benefits of the NUAR platform can be realised as early as possible.

Conformance Class A requires the supply of:

- **Organisation Data**
- Feature records containing **mandatory attributes**
- Minimum level of **dataset completeness**
- **Updates at the agreed frequency**
- Data in the **local schema**
- **Cross-section diagrams** where appropriate
- Data following the required **file naming conventions**

4.2.1 Organisation Data

See the section on **Organisation Data Supply** which describes the type and format of organisation data that must be supplied.

4.2.2 Mandatory Attributes

A feature will be accepted into the NUAR platform provided that, as an absolute minimum, it can be spatially located and plotted within a defined area of validity (possibly as broad as the UK), and its owner can be identified. The NUAR Harmonised Data Model specifies these characteristics as mandatory attributes for each of the feature types that will be supplied by the Data Asset Owner. Full details of all the attributes (mandatory and optional) can be found in the **Feature Types and Attributes** section.

Requirement A01

Mandatory Attributes

In order for a feature to be accepted into the NUAR platform it must include valid values for the attributes described in this section

Mandatory Attributes of the MUDDI object
The parent class of underground infrastructure or other underground environment features is the MUDDI object. Its mandatory and optional attributes are inherited by all more specialised classes in the MUDDI core model and the derived NUAR Harmonised Model. The following table details the mandatory attributes that the data provider shall supply for all underground assets.

Table 3: Mandatory Attributes for NUAR features

Attribute: assetID	
Description: This is an identifier assigned uniquely to a feature by an asset owner. It should be as stable and persistent as possible relative to that particular physical asset or defined region. If a persistent unique identifier cannot be assigned by the Data Asset Owner, a value of “Not Supplied” must be supplied.	
Type: CharacterString	Multiplicity: 1..1
Attribute: dataAgentAssetID	
Description: This is an identifier assigned uniquely to a feature during the transformation process, which may be carried out by a service provider appointed by the Asset Owner. This ID may differ from the source Asset ID to ensure uniqueness within the transformation process. It is mandatory where the Data Asset Owner is not able to provide their own unique, persistent source identifier for a feature (i.e. where a unique, persistent source Asset ID is not available, one should be assigned during transformation if possible).	
Type: CharacterString	Multiplicity: 0..1
Attribute: businessContact	
Description: Contact information for the asset responsible party.	
Type: CharacterString	Multiplicity: 1..1
Attribute: emergencyContact	
Description: Contact information (telephone number, etc.) to reach the responsible party in the event of an emergency involving this asset.	
Type: CharacterString	Multiplicity: 1..1
Attribute: horizontalCRS	
Description: Coordinate system, datum, and epoch date (if applicable) associated with the X and Y coordinates.	
Type: CharacterString	Multiplicity: 1..1
Attribute: geometry	
Description: Geometry of the feature.	
Type: CharacterString	Multiplicity: 1..1

4.2.3 Dataset Completeness

In order for the NUAR platform to meet the safe dig use case, reaching a level of confidence that all available asset data is presented to the user is of fundamental importance. A proposal for the Conformance Class A level of “Completeness of data supplied in Active Regions” can be found in the **Data Asset Owner Conformance Matrix**.

4.2.4 Updates at agreed frequency

During the on-boarding process a data ingestion specification will be agreed to underpin the regular and sustainable supply of data. The agreed frequency will be established based on the type of data and the update frequencies of existing channels of distribution.

Requirement A02

Data Supply Frequency

Data must be supplied to the NUAR platform at the frequency agreed during the on-boarding process

4.2.5 Local Schema

In Conformance Class A Data Asset Owners may supply their data in their local schema to the NUAR Transformation Platform. This includes the supply of feature and raster data. Refer to the **Local Schema Data Supply** section for details.

4.2.6 Cross-section Diagrams

If a Data Asset Owner holds cross-section diagrams, these can be supplied according to the requirements specified in the Cross-section Diagrams section.

4.2.7 File Naming Conventions

All data must be supplied in accordance with the file naming specifications.

4.3 Conformance Class B

Conformance Class B is where Data Asset Owners are reaching a level of completeness and domain consistency that will ensure an improved experience for the users of the NUAR platform.

Proposed levels for the bullet points below can be found in the **Data Asset Owner Conformance Matrix**. Once agreed, a requirement for Conformance Class B can be specified.

Able to supply:

- A full refresh of data at an agreed frequency of at least a month
- Valid feature records with increased **attribute completeness**
- Data that reaches a set level of **Domain consistency**
- **Response to Observations** within a set time frame

4.4 Conformance Class C

In this conformance class Data Asset Owners are providing persistent unique identifiers. This has the potential to open up additional functionality in the NUAR platform, such as the provision of Change only updates and the full benefit of the Observations feedback loop.

Able to supply:

- Persistent, unique identifiers

Requirement C01

Persistent Unique Identifiers

Data must be supplied to the NUAR platform with identifiers that are unique and persistent.

4.5 Conformance Class D

This is where Data Asset Owners are able to supply all of their assets as vector data in the NUAR schema. This increases the efficiency and scalability of the platform overall as there are fewer bespoke processes to set up and maintain to handle data from a range of local schemas and data formats.

Once Data Asset Owners are able to supply data that conforms to the NUAR schema, it will also be possible to provide data via a web service hosted by a Data Asset Owner or a service provider engaged by them. Details and requirements for this can be found in the **Federated Data Supply – Features** section.

Able to supply:

- Data in the NUAR Schema
- Data in vector format

Requirement D01

Data Supply in NUAR Schema

All asset data must be supplied to the NUAR platform in accordance with the requirements described in the **NUAR Schema Data Supply** section.

5. Organisation Data Supply

Organisation Data refers to the data that the NUAR platform needs to know about an organisation in order to process data, make statements to users about levels of data coverage, and ultimately to meet the requirements of the safe excavation use case.

Requirement A03

Organisation Data Supply

The data described in this section must be provided by all Data Asset Owners during the onboarding phase.

Updates to this data may also be required as the Data Asset Owner's situation changes.

The full scope of the Organisation Data Supply is yet to be agreed, however, key attributes will include a unique identifier for the organisation, their operational area(s), and the data supply area.

A unique Organisation Identifier: A shortened or abbreviated name or code for an organisation that can be consistently used across the platform to uniquely identify Data Asset Owners.

Data Update Frequency: Data Asset Owners will be asked to commit to an agreed frequency of providing updates to their data assets. When compared with the date of the last supply, the data update frequency value will enable users to understand how current the data from a particular Data Asset Owner is.

Supply of Change: For Data Asset Owners with slowly changing datasets, it may be more appropriate to record a stated commitment to supply change within a certain time frame. This will enable the NUAR platform to indicate to users the level of data currency and validity even when data is not being supplied due to there being no updates to the data assets.

Operational Areas: The area in which a Data Asset Owner operates. Operational Areas will exist at different levels: Franchise, Network and Patch. Knowing the operational areas will enable NUAR, for any given point, to answer the question: Who do I need to get records from to know what is buried here?

Data Supply Area(s): The area(s) for which a Data Asset Owner can confirm that they have supplied all known data.

The Operational Area information, combined with the Data Supply Area information, intersected against the NUAR regions is critical to understanding when it is sensible to make a NUAR region active, and to confidently provide information to users about known gaps in data coverage.

The supply method, format and structure in which this data must be supplied will be determined during the build phase.

6. Local Schema Data Supply

6.1 Process

The key to loading data into the NUAR platform for data providers is collaboration with the NUAR data transformation team, ensuring that they understand the complexities and nuances in the local data such that they are able to extract all relevant data. To this end, an example process that covers the essential steps is outlined below:

Step 1 – Read Guidance
The Data Asset Owner will be supplied with outline information on preferences for data supply and sample mappings relevant to the Data Asset Owner's sector; Gas, Electric, etc. See the **File Formats** section for details.

Step 2 – Data Review Session
A session with the NUAR Data Team to capture information such as:

- GIS software in use (as this is likely to influence file format supply)
- An overview of what data they capture e.g. cables, structures, ducts, sites, etc.
- An overview of how the data is stored i.e. vector or raster data

Step 3 – Sign Data Exploration and Distribution Agreements
A Data Exploration Agreement is a lightweight legal contract that allows the NUAR team to internally analyse the asset owner data.

The Data Distribution Agreement is a legal contract that permits the sharing of Data with all relevant NUAR organisations.

Full acceptance in the NUAR process has a pre-condition for the NUAR Data Distribution Agreement to be signed by the Data Asset Owner and the NUAR platform operator.

Step 4 – Supply Sample Data
Once a Data Exploration Agreement is signed as a minimum, sample data that includes all features types will then need to be provided so that the NUAR Data Team can begin work on the transformation from the local schema to the NUAR Harmonised Data Model schema.

Step 5 – Data Workshop
Once the sample data has been received and inspected by the team, it will be possible to arrange a workshop to collaboratively define the mappings for feature types, attributes and codelist values.

Step 6 – Agree and document data ingestion specification
The process to extract the data from the supplier's system must be repeatable such that the data is output in the same format, with the same feature types and same attribute names each time the process is run. The terms of a repeatable and sustainable process for supplying data in an agreed format at an agreed frequency

will be agreed and documented with the Asset Owner. This will form the basis for continued supply of the data.

Step 7 – Supply Full Dataset
The full data set must be supplied, and will go through a test load process to iron out any issues that may not have been identified in the work on the sample data. Once testing is successfully completed the data will be loaded into the live system and conformance reports will be generated.

6.1.1 Subsequent Data Supply and Self-upload

It will be possible for Data Asset Owners to upload their own data once the transformation process has been implemented, and this could include the provision of updated data. For details on this refer to **Supply of Change**, and the **Data Submission** sections.

6.2 File Formats

Requirement A04

Local Schema File Formats

Data supplied in a local schema must be provided in one of the formats listed in Table 4: Acceptable Local Schema File Formats

Table 4: Acceptable Local Schema File Formats

Format	NUAR Status	Notes
GML	Preferred	Mature open standard. Uses XML grammar so validation is possible in a broad range of software tools and libraries.
Geopackage	Preferred	Open standard. Based on SQLite, which is mature. Broad tooling support. Supports very large data volumes.
ESRI Geodatabase	Supported	Proprietary standard, but still has broad tooling support.
ESRI Shapefile	Tolerated	Proprietary but has broad tooling support. Attribute names are limited to 10 characters. Also has a more restricted file size than the supported or preferred formats.
MapInfo (tab, dat, ind)	Tolerated	Proprietary but has broad tooling support.

Format	NUAR Status	Notes
CAD (DWG, DXF, DGN)	Tolerated but not advised	Due to the time taken to convert CAD files to vectorised asset data, it is preferable that this conversion be done upfront by the asset owner, and may require assistance from a specialised service provider.

6.3 Geo-referenced Raster Data Supply

Geo-referenced Raster data refers to the representation of underground assets in an image format that has been aligned to a coordinate reference system. This type of data typically exists where underground assets have not been fully digitised by the Data Asset Owner.

Requirement A05

Geo-referenced Raster File Formats

Data supplied in raster format must be provided in one of the formats listed in Table 5: Acceptable Geo-referenced Raster File Formats

6.4 File Formats

Table 5: Acceptable Geo-referenced Raster File Formats

Format	NUAR Status	Notes
GeoTiff	Preferred	This format is preferred because the required metadata is inline, i.e. within the data file itself
TIFF (with accompanying world file - .tfw)	Tolerated	Widely supported but requires external metadata file
JPG (with accompanying world file - .jgw)	Tolerated	Widely supported but requires external metadata file
PNG (with accompanying world file - .pgw)	Tolerated	Widely supported but requires external metadata file

6.4.1 Data Submission

For details on how to supply data, and subsequent updates, refer to the **Data Submission** and **Supply of Change** sections.

6.5 Supplemental Drawings

These are typically an enlargement or annotated representation of a location. These files represent a real-world location at a larger scale than the main map view or with further detail included in the image. These files will typically be linked to an area in the main map view rather than a single point as for cross-sections (see below). It is also worth noting that some supplemental drawings may contain both enlarged map views and cross-section drawings. As Asset Owners increasingly digitise their data, it is anticipated that the need to manage supplemental drawings will decrease over time.

The management of these types of files may need to be considered on an organisation by organisation basis. However, one way that has successfully been demonstrated to supply, store and display them is as follows:

1. Ensure that each supplemental drawing file is uniquely named within the organisation's data supply (also, follow the guidance in the **File Naming** section).

2. Associate each file to a spatial extent, which can be done by creating a GeoJSON feature collection, where each feature consists of a uniquely named spatial extent, and its properties consist of a name, and an array of the names of all the related supplemental drawings. An example of such a GeoJSON file is shown in Figure 3:

```
{
	"type": "FeatureCollection",
	"features": [
		{
			"type": "Feature",
			"geometry": {
				"type": "Polygon",
				"coordinates": [
					[
						[
							REDACTED,
							REDACTED
						],
						[
							REDACTED,
							REDACTED
						],
						[
							REDACTED,
							REDACTED
						],
						[
							REDACTED,
							REDACTED
						],
						[
							REDACTED,
							REDACTED
						],
						[
							REDACTED,
							REDACTED
						]
					]
				]
			},
			"properties": {
				"filename": "[REDACTED]",
				"supplementaries": [
					"[REDACTED]_supplemental1",
					"[REDACTED]_supplemental2",
					"[REDACTED]_supplemental3"
				]
			}
		}
	}
```


Figure 2: Sample GeoJSON supporting Supplemental Diagrams

6.5.1 Anticipated Formats

The platform will determine the most appropriate formats for the display of Supplemental drawings over time. This is likely to include formats primarily aimed at web graphics such as PNG, or formats that work at multiple resolutions that also support hard copy print outs such as PDF.

6.5.2 Data Submission

Refer to the **Data Submission** section.

6.5.3 Supply of Change

Refer to the **Supply of Change** section.

7. NUAR Schema Data Supply

This section describes the structure and format of data that is to be supplied in accordance with the NUAR Harmonised Data Model schema.

7.1 Data Types

Type	**Description**
Boolean	Value of ‘true’ or ‘false’
Date	Specifies a day within the Gregorian calendar in the format YYYY-MM-DD
Double	A 64-bit IEEE 754 double precision Floating Point Number. A double has 15 decimal digits of precision
Integer	Any positive or negative whole number or zero
String	An ordered set of characters

7.1.1 Geometry

The following are the types of valid geometries that may be supplied to NUAR:

Point
A Point is a 0-dimensional geometric object and represents a single location in coordinate space. A Point has an x-coordinate value and a y-coordinate value.

Example: A point defined in the BNG reference system has easting and northing ordinates in units of metres, where the easting is in the range 0 to 700000 and the northing is in the range 0 to 1300000. Ordinates may be supplied to mm precision.

LineString
A LineString is an ordered set of points that are interpolated linearly. Each consecutive pair of Points defines a Line segment.

MultilineString
A MultiLineString is an array of LineString geometries. There are no specific semantics or rules applied to the collection.

Polygon
A polygon is a single closed region on the spatial reference system projection plane, defined by a set of geometric rings that represent the boundaries. A polygon must have one outer boundary and zero or more inner boundaries (holes in the polygon). The boundaries must not intersect themselves, and the inner boundaries must not cross each other or contain other inner boundaries. Coordinates in outer boundaries are oriented in an anticlockwise direction; coordinates in inner boundaries are oriented in a clockwise direction.

Multipolygon
A multipolygon is an array of Polygon coordinate arrays.

British National Grid (BNG) - for data in Great Britain
All geometric data supplied to NUAR within Great Britain must use the British National Grid (BNG) spatial reference system. BNG uses the OSGB36® (EPSG 27700) geodetic datum and a single Transverse Mercator projection for the whole of Great Britain. Positions on this projection are described using easting and northing coordinates in units of metres.

Where geometries need to be transformed to BNG, the OSTN15 model must be used. A guide to coordinate systems in Great Britain that includes details of OSTN15 is available at: http://www.ordnancesurvey.co.uk/docs/support/guide-coordinate-systems-great-britain.pdf

Introductory material on the BNG and ODN and the full definition of the BNG (OSGB36 National Grid) is available on the Ordnance Survey OS Net GPS site: https://www.ordnancesurvey.co.uk/oswebsite/gps/

A general introductory guide to BNG is provided at: https://getoutside.ordnancesurvey.co.uk/guides/beginners-guide-to-grid-references/

Irish Grid Reference System – for data in Northern Ireland
All geometric data supplied to NUAR within Northern Ireland must use the Irish Grid Reference System. This uses the OSNI 1952 (EPSG 29901) geodetic datum and a single Transverse Mercator projection for the onshore Northern Ireland area. Positions on this projection are described using easting and northing coordinates in units of metres.

Where geometries need to be transformed to the Irish Grid Reference System, the OSi/OSNI polynomial transformation should be used.

Height above datum
Suppliers may provide z-ordinates in their geometries that represent height above datum. In which case the verticalCRS attribute value will become mandatory.

7.2 Feature Types and Attributes

The Data Model documentation provides the full details of Feature Types, Attributes and Codelist values that must be adhered to. The packages represented in the data model are as follows:

- Drainage
- Electricity Network
- Gas Network
- Highways Assets
- Sewer Network
- Telecommunications Network
- Transport
- Water Network

Additionally, there is a data model package for Reference Data that describes features and attributes for:

- Protected Sites
- Historical Land Use
- Planning Data

7.3 File Format

Limiting the range of supported formats means that the platform can be built and maintained more efficiently, as it makes data loading pipelines fewer, shorter and less complex, and makes schema updates following data model changes simpler and less prone to error. Therefore, data can be supplied in the NUAR Harmonised Data Model schema in two different formats: GeoJSON and Geography Mark-up Language (GML).

These two formats have been chosen as they are both open standards and have broad tooling support, and offering both widens the options for supply and helps to lower the barrier to providing data to the NUAR platform.

7.3.1 GeoJSON

GeoJSON is an open standard format for the interchange of geospatial data. The format is based on JavaScript Object Notation (JSON). It defines several types of JSON objects and the manner in which they are combined to represent data about geographic features, their properties, and their spatial extents.

The GeoJSON standard specifies that the Coordinate Reference System to be used is World Geodetic System 1984 (WGS84/EPSG:4326), with the ordinates represented as units of decimal degrees. However, by community agreement it is acceptable to use a different CRS. Therefore, data supplied in GeoJSON must be supplied with OSGB36/EPSG:27700 coordinates.

7.3.2 GML

Geography Mark-up Language (GML) v3.2. GML is an open standard that is widely supported by off-the-shelf, and often open source, GIS tooling. As GML is based on XML, validation of GML files can be performed with standard XML parsing libraries and applications.

This section describes how to structure the data to be supplied in GML format. Recommended documents to support understanding and the creation of GML documents are:

- The Open Geospatial Consortium (OGC) document, Geography Markup Language v3.2.1 (see **Related Documents**)
- The eXtensible Mark-up Language (XML) specifications, that GML is a profile of, are available from the World Wide Web Consortium (W3C) website: http://www.w3.org. An understanding of XML and XML Schemas is required

GML documents are a subset of XML documents and can be validated against a schema. The NUAR schema defines the construct of NUAR application-specific features, attributes, data types, and codelists.

- NUAR_HarmonisedModel_v1-2.xsd (see **Related Documents**)

XML Declaration
The document must begin with the XML declaration:

```
<?xml version="1.0" encoding="UTF-8"?>
```

Feature Collection
All information shall be provided in a GML feature collection. This element must include the following XML namespaces declarations:

xsi http://www.w3.org/2001/XMLSchema-instance
gml http://www.opengis.net/gml/3.2
nuar URL tbc

Example:

```
<gml:FeatureCollection
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:gml="http://www.opengis.net/gml/3.2"
xmlns:nuar="<<NUAR schema URI goes here>>"
gml:id="<<identifier goes here>>"
xsi:schemaLocation="http://www.os.uk/schemas/NUAR/1.0 nuar.xsd" >
...
 </gml:FeatureCollection>
```

8. Cross-section diagrams

Cross-section diagrams are files that contain information about the disposition and depth of multiple assets at a given point on the network, with their location on the network being comprised of an X/Y position and a direction of view. These files are a valuable contextual source of information for excavation activities.

These files co-exist with a vector feature, and that vector feature is used to locate the cross-section diagram such that it can be displayed in the NUAR user interface in the correct position.

Requirement A06

Cross-section Markers

Cross-section diagrams shall be supplied with a corresponding CrossSectionMarker feature. The CrossSectionMarker feature must include the file name of the relevant cross-section diagram file. For details on the attributes of the CrossSectionMarker refer to the Data Model documentation.

Cross section diagrams may be supplied when data is being supplied according to the Local Schema Data Supply section. The exact format of supply should be agreed with the NUAR platform team.

Requirement A07

Cross-section File Naming

Cross-section diagrams shall be supplied with a file name that is unique to the supplying organisation, in order that file naming conflicts can be avoided. Other requirements for file naming can be found in the **File Naming** section.

8.1.1 Anticipated Formats

The format of cross-section drawings is likely to include formats primarily aimed at web graphics such as PNG, or formats that work at multiple resolutions that also support hard copy print outs such as PDF.

8.1.2 Data Submission

Refer to the **Data Submission** section.

8.1.3 Supply of Change

Refer to the **Supply of Change** section.

9. File Naming

.zip Archive Naming Convention

Requirement A08

Parent .zip archive

To ensure consistency of supply of data all individual data files must be provided within a “parent” .zip archive.

The following is a proposal for a naming convention that enables data to be processed reliably and efficiently:

{Data Asset Owner name}_{Date of data extraction}_{fs/cou}.zip

{Data Asset Owner name}	Unique name or code of the Data Asset Owner – as specified in the **Organisation Data Supply** section. With this in the file name it will be possible to perform initial required checks on data processing, such as “Does the Data Asset Owner have a valid Data Distribution Agreement?” without even opening the file.
{Date of data extraction}	Date the data was extracted from the source system – in format YYYYMMDDHHMM. The date can be compared against the dates of previously processed files from a Data Asset Owner to ensure that old files do not overwrite new data.
{fs/cou}	Indication of whether the data is a “full supply” or a “change only update”. This will determine the process flow for the data files.

Example:

NGW_202003161534_FS.zip

9.1 Individual File Naming Convention

9.1.1 Vector Data

{Data Asset Owner name}_{Date of data extraction}_{fs/cou}_{file identifier}.{file extension}

{file extension}	Character string that identifies the format of the file.
{file identifier}	The primary purpose of this value is to ensure that files can be named uniquely within the parent directory.

Example:

GDHS_202003161534_FS_001.gml

9.1.2 Geo-referenced Raster and Supplementary Files

{Data Asset Owner name}_{unique file reference}.{file extension}

{unique file reference}	This must be a reference that is unique to the data asset owner. This will ensure that all such files can be stored on the NUAR platform without causing file-naming conflicts.

Example:

ABC_34ryt.pdf

10. Supply of Change

10.1 Vector Data

Typical update schemes for geospatial dataset into a repository distinguish between two very different approaches. It is possible to:

1. supply change by providing either a complete re-supply of data (full supply), or
2. provide only the features that have changed, i.e. features that are new, updated, or no longer exist (change only update or COU).

It needs to be decided which of these update approaches should be supported for different data providers.

10.1.1 Full Supply

Full supply could be for the entire data holdings of a Data Asset Owner, or it could be a full supply of a particular feature type. When a “full supply” is received from a Data Asset Owner, any vector data that currently exists for the supplied feature types for that Data Asset Owner will be removed from the system, to be replaced with the data in the incoming file(s), with an audit trail stored reflecting the data load.

10.1.2 Change Only Update

Change Only Update (COU) is where, in order to maintain a current view of a Data Asset Owners data on the NUAR platform, only the features that are new, have been updated or deleted are supplied to the platform. This enables efficiencies to be gained in preparing data to be sent, the transmission of significantly reduced volume of data, and in the loading of data into the NUAR platform.

However, in order to provide COU data to the platform, there needs to be an absolute requirement to allocate unique persistent asset identifiers (assetID) to the Data Asset Owner’s data. These identifiers will be used to update records for features that already exist on the platform, or remove features that should no longer be stored on the platform.

For data providers and data sets for which COU is viable, a COU schema will need to be defined that supports the specification of deleted records, as well as new and updated records.

10.2 Geo-referenced Raster and Supplementary Files

It will be possible to update copies of both Geo-referenced raster files and supplementary files, either by full supply or COU. Similarly to vector data, if a full supply of either of these types of data is submitted, it will result in a full replacement of the data that exists on the NUAR platform for that data provider. Alternatively, if the Geo-referenced Raster or Supplementary files can be provided with a persistent

name, it will be possible to supply and therefore update only the files that need to be updated.

When replacing raster data with vectorised data, a full re-supply of the remaining raster data for a Data Asset Owner will ensure that redundant raster data is removed from the NUAR platform.

11. Data Submission

All data to be hosted on the NUAR platform will be uploaded in the target NHDM schema via the data ingestion subsystem in the NUAR platform. Data Asset Owners who supply their data in their “local schema” will need to supply data via the NUAR Transformation Platform which will perform the transformation to the target schema, based on data being supplied in a consistent repeatable form as documented and agreed in the data ingestion specification for that asset owner.

12. Federated Data Supply – Features

It will be possible to provide data to the NUAR platform via a compliant web service. This web service could be hosted by the Data Asset Owner or by a service provider engaged by them. Details of the requirements for ingestion of external inbound APIs are defined in the supplied documentation.

The web service must follow a recognised OGC open standard that defines the API and supports discovery and query operations. These are described in the OGC Features API documentation as:

- *Discovery operations enable clients to interrogate the API to determine its capabilities and retrieve information about this distribution of the dataset, including the API definition and metadata about the feature collections provided by the API.*
- *Query operations enable clients to retrieve features from the underlying data store based upon simple selection criteria, defined by the client.*

12.1 Payload

Requirement D02

API Payload Format

The payload of an API query response must be in GML or GeoJSON format.

12.2 Conformance

Data that is supplied via API will be assessed against the **Data Asset Owner Conformance Matrix**.

13. Related Documentation

Document	Version	Date	Description
Linked identifier schemes: Best Practice Guide		October 2019	Guidance, published by the Geospatial Commission, on the design, creation and management of identifiers. https://www.gov.uk/government/publications/linked-identifier-schemes-best-practice-guide
Description of the NUAR Harmonised Data Model	1.2	October 2020	Defines and documents the harmonised data model on which the National Underground Asset Register Pilot is based.
OGC MUDDI Engineering Report		November 2019	Describes the design and development of conceptual UGII integration model MUDDI (Model for Underground Data Definition and Interchange). http://docs.opengeospatial.org/per/17-090r1.html