National Land Data Programme (NLDP): Land Data Specification

Version 3.5

OFFICIAL

Synopsis

This document is a data product specification for land data. The specification was developed as part of the Geospatial Commission's National Land Data Programme to investigate optimal approaches for how information about land could be represented to support land use decision making, in particular national scale scenario modelling of land use options.

At the topic level the specification considers three core facets of land:

- The 'as is' nature of land and its use patterns (land use, land cover). 'As is' also includes planned use that is intended, but has not been realised.
- Constraints that prevent or modify how land can be used (hazards, land use restrictions, and protections)
- Potential opportunities for land use based on agreed typologies (e.g. 'housing development potential', 'solar energy potential')

Across all facets the model supports representation of land as simple features (points, lines, polygons) and simple coverages (regular grids).

The purpose of this specification is to:

- provide a guide to the creation and use of land data utilising the NLDP model
- provide a framework for creating a data specification for a system implementation of the NLDP data model

The specification contains a conceptual model for land, building on the concepts for modelling land developed as part of The INSPIRE Regulations 2009, aggregating them to give a packaged, holistic view of land that could support a wide range of decision making.

A database implementation of the model was created as part of the development of the specification and used for testing the model across a range of different land data sets.

Contents

Glossary of Terms

Term	Definition
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
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 a platform built on the NLDP Model
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. It is part of the Cabinet Office.
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
INSPIRE	Infrastructure for Spatial Information in the European Community. Conceptual and logical models developed by the EU for representation of spatial datasets covering a wide range of thematic groups.
NLDP	National Land Data Program
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

Normative References

The NLDP model has been built around a number of pre-existing standards to ensure interoperability and consistency. These are detailed below:

ISO 19103:2015 – A profile of the Unified Modelling Language (UML) for use with spatial data. Establishes a set of data type definitions that have been used for feature class attribution in the NLDP model. Link: https://www.iso.org/standard/56734.html

ISO 19107:2019 – Conceptual schemas for describing, representing, and manipulating the spatial characteristics of geographic entities. Defines a set of vector and raster geometry types for representing spatial data. Link: https://www.iso.org/standard/66175.html

ISO 19144:2009 Part 1- establishes the structure of a land cover classification system, together with the mechanism for defining and registering the classifiers for such a system. ISO - ISO 19144-1:2009

ISO 19144:2012 Part 2 - specifies a Land Cover Meta Language (LCML) expressed as a UML metamodel that allows different land cover classification systems to be described. ISO - ISO 19144-2:2012

ISO 19157:2013 Part 1 - establishes the principles for describing the quality of geographic data. ISO - ISO 19157:2013

INSPIRE Land Themes – The INSPIRE themes for land data were chosen to be the basis for the NLDP model, although they have been adapted to meet the requirements of the NLDP project. The INSPIRE Directive led to the creation of a set of logical data models representing 34 categories of spatial data of interest to policy makers within the European Union. Of interest to the NLDP project were the following four INSPIRE themes:

- Land Cover: https://inspire.ec.europa.eu/Themes/123/2892
- Land Use (both current and planned): https://inspire.ec.europa.eu/Themes/129/2892
- Protected Sites: https://inspire.ec.europa.eu/Themes/117/2892
- Natural Risk Zones: https://inspire.ec.europa.eu/Themes/140/2892

1. Introduction

1.1. Context of the NLDP Model

The NLDP Data Product Specification is a project commissioned by the Geospatial Commission (GC). The aim is to investigate how a dataset about land can be designed to support strategic decision making to optimise land use.

Our land is a finite resource. There are competing demands for housing, infrastructure, farming, mineral resources, flood protection, carbon sequestration and natural habitat protection. Competition for land is also set to increase in the future with the need to reach net zero greenhouse gas emissions.

This requires a joined-up approach to decisions about land use at all levels of government. Data about land use is fundamentally important to ensure that decision making is not only evidence based but will also stand up to public scrutiny. The access to an authoritative land dataset underpinned by a common specification will allow a better understanding of our land and increase collaboration between groups of land professionals and landowners. This will contribute to better policy making in the future.

With this background there is an imperative for up-to-date land use and land cover mapping to be available to any land professional with an interest in how our land is managed:

- Land use is defined as territory characterised according to its current and future planned functional dimension or socio-economic purpose (e.g. residential, industrial, commercial, agricultural, forestry, recreational).
- Land cover is inextricably linked to land use and is defined as an abstraction of the physical and biophysical cover on the earth’s surface. Land cover data provides a description of the surface of the earth by its physical characteristics.

Taken together, land use and land cover describe what is currently present on the land surface and how it is used. For example, an area could be described as grassland and its use being for agriculture. Alternatively, an urban area could be described as residential.

In addition, this project considered other information about land that contributed to making strategic land use decisions. This included the capability to capture natural risks, such as areas of flooding and landslides as well as the ability to record protected sites and land use potential for new activities. Three distinct viewpoints on land were identified:

- As Is: How is the land utilised currently? This typically means the land use and land cover. Land use provides the current land function and land cover its form. Very often one defines the other.
- Constraints: What can impact our choices? This can include our preferences to how we protect our land, e.g. historic buildings or national parks, and also hazards such as flooding, subsidence or erosion
- Opportunity: What are the potential options for land? Land can be typified according to combinations of its topology, climate, soil and infrastructure amongst others to support particular potential uses, e.g. carbon sequestration, recreation, renewable energy.

If the ‘opportunity’ is combined with the ‘constraints’ and the ‘as is’ we found it provides a good basis for strategic decision making; striking a balance between precision and ease of use. This is summarised in the figure below.

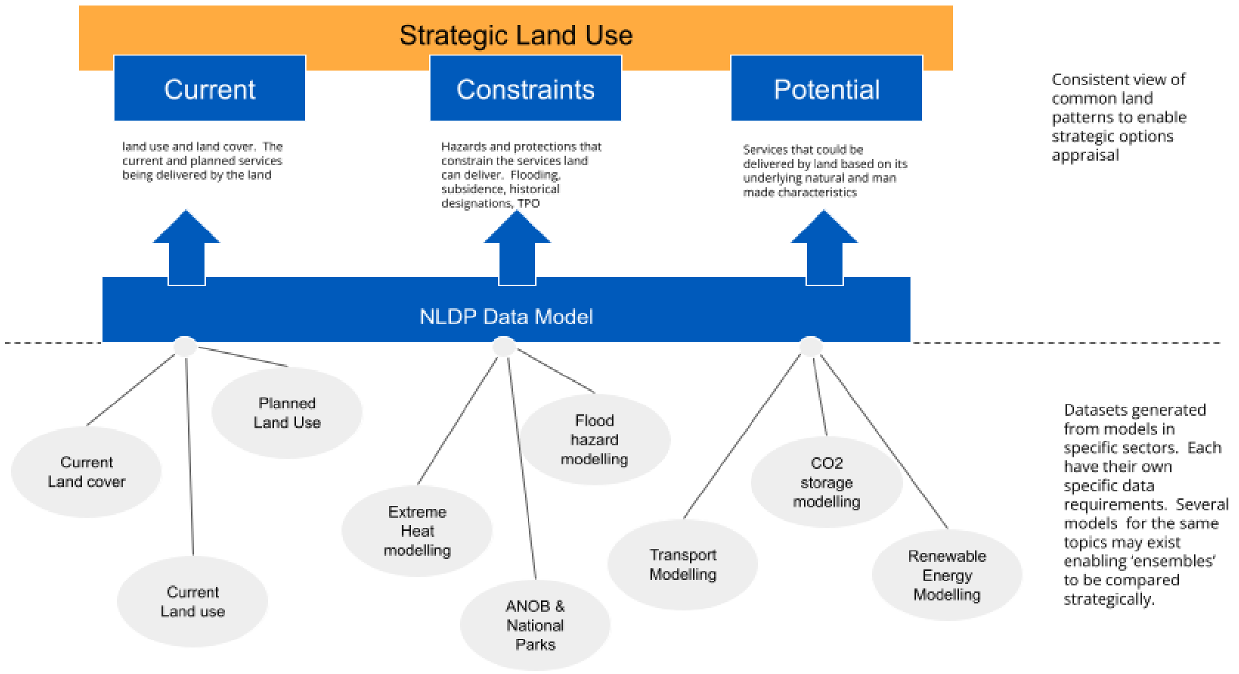

Figure 1 Context of the NLDP Land Data Model

It is recognised that for strategic decision making the emphasis is on ‘ready to use’ information. For example, a map of flood hazards is typical of the type of information needed for appraising land use options. This data set is created from the analysis of many other data such as rainfall, river flows, flood defences and land cover amongst others. The same is true for land use ‘opportunity’; for example, land that has the potential for onshore wind will have a number of characteristics that would need to be combined from many different datasets.

The NLDP data specification recognises that there will also be different ways to derive ‘flood hazard’ or ‘wind farm potential’ using different analysis techniques and different data. The NLDP data specification does not seek to support this myriad of use cases. Instead it provides a standardised way to compare the outputs of such analysis and where several analytical approaches exist, land use options appraisal would benefit from inter comparing these different derivations.

1.2. Approaches to develop the NLDP Model

This logical data model has been designed to facilitate the harmonisation of land data across the UK to support strategic land use options appraisal. The challenge is that land data is not a single topic and often many viewpoints of land need to be considered when making land use decisions

To model land data, NLDP uses the published INSPIRE themes as a foundation but builds upon these and adapts the model for the identified use cases. INSPIRE was chosen for the flexibility that the specification offers, particularly in incorporating the land use management layers defining risks, protected sites and planned land use.

Conceptual and logical models are developed based on the following four INSPIRE themes:

- Land Use (covering both current and planned land use)
- Land Cover
- Natural Risk Zones
- Protected Sites

The conceptual data model was created in two stages, and was designed 'top down' around the requirements of the NLDP programme, rather than 'bottom up' determined by the land datasets available. The first stage considered the data layers required to describe the physical features on the ground and how they interact. Land Use and Land Cover were considered essential to meet this requirement.

The second stage added the data layers required to manage the land. Data layers describing, planned land use, natural risk zones and areas of protected land were considered important in providing the necessary information to facilitate land management decision making.

Examples of planned land use could be for housing, or national infrastructure. Natural risk zones are areas where there is a potential risk to use of the land such as flooding or land slips. Finally protected land could be conservation areas, green belt, and areas of outstanding natural beauty (AONB).

The INSPIRE Area Management and Administrative Units were considered, but not utilised as there was adequate provision within the selected themes to capture the information required for strategic planning of land, avoiding repetition and keeping the data model as light as possible.

Similarly, foundational INSPIRE themes such as Geology, Soils, Hydrography and Transport Network layers were also considered. These layers require further specialist analysis to extract the relevant information for land use planning and were discounted in favour of a model containing 'decision ready' data derived from these foundational layers and suitable for interpretation for a wider range of professionals.

1.3. Content of the NLDP Land Data Specification

This document describes the NLDP Logical Model, provides insight into the rationale behind the model's elements and gives some high-level guidance on what to consider when implementing the NLDP model, either as a creator, user or publisher of land data. It contextualises and complements the Feature Catalogue (Appendix A) and UML Logical Model artefacts.

This data specification will:

- Assist data owners who wish to create datasets that conform to part or all of the NLDP model.
- Assist data users who wish to analyse data already conforming to the NLDP model.
- Assist data users who wish to undertake national scale spatial modelling of land data, for whom a common target model would enable easier comparison of datasets from different jurisdictions.
- Assist developers who wish to create datastores conforming to the NLDP model.

The precise requirements of the above will of course vary depending on their intended use cases, but this document aims to signpost the key considerations when utilising the NLDP model, and explain the rationale behind decisions taken during the development of the model.

2. NLDP Logical Model Description

The following section provides an informal description of the different sections of the logical model.

Section 2.1 provides general comments about the approach to geometry types across the model. Section 2.2 describes the approach to coordinate reference systems. The remainder of the sections in this chapter describe the different sections (packages) of the NLDP model. Apart from NLDP Core (which defines a common set of attribution used throughout the model), these packages directly map to part or all of an INSPIRE theme:

NLDP Package	INSPIRE Theme
Existing Land Use	Land Use
Planned & Potential Land Use	
Land Cover	Land Cover
Protected Sites	Protected Sites
Risks	Natural Risk Zones

A full UML Feature Catalogue detailing every element of the NLDP logical model has been provided as Appendix A.

2.1. Treatment of geometry within the NLDP Model

The key rationale behind developing the NLDP model was to support decision making by facilitating the comparison and analysis of different land datasets. This data could be in many different forms and degrees of complexity. Therefore, the NLDP model does not specify a particular geometry type (point, line, polygon) for many features, instead using the generic GM_Object datatype. This allows, say, Protected Sites to be represented by polygons (e.g. a National Park extent) or points (e.g. a Tree Preservation Order).

Other features do have a geometry type specified, such as Dataset features (it would not make sense for the extent of a dataset to be anything other than a polygon).

This is a departure from the INSPIRE models, but it was felt that supporting the easy aggregation of land datasets was more important than interoperability with INSPIRE.

One downside of this approach is that it could cause complications with some technologies that do not support features having different geometry types within the same feature class. It would be relatively simple, however, to amend the model and define geometry types at the feature class level if needed.

2.2. Coordinate Reference Systems

The choice of a coordinate reference system (CRS) would depend on the use-case(s) for a specific implementation of the NLDP model. Many datasets focussing on Great Britain will use the British National Grid, but if a wider focus is needed (such as UK-wide) then it would make sense to choose a different CRS that covers the entire area of interest. The model allows CRS information to be stored at both the dataset and feature level, so it would be possible, assuming the implementation allows, to use different systems for different datasets. This could cause issues when trying to compare datasets in different CRSs, so in many cases it could be more appropriate to transform all datasets to one CRS.

2.3. NLDP Core

This section of the model serves to harmonise the approach to data provenance, authority, and temporality across the NLDP model. It does this by establishing common sets of attribution that are inherited by theme-specific feature classes throughout the NLDP model. It also establishes INSPIRE's common structure for feature-level identifiers, as well as the common approach to citation of legislation, external documents, and journals.

There are two basic "building blocks" of the NLDP logical model that form the foundation for all other feature types in the model:

- NLDPDataUnit represents individual spatial features. These could be, for example, land parcels, planning areas or protected sites.
- NLDPDataset represents a collection of spatial features. This collection will likely represent an individual source dataset or perhaps an individual release of a regularly updated dataset.

All other feature classes within the NLDP model inherit a common set of attribution from one of these classes, as outlined below.

NLDPDataUnit represents individual spatial features. These could be, for example, land parcels, planning areas or protected sites.

NLDPDataset represents a collection of spatial features. This collection will likely represent an individual source dataset or perhaps an individual release of a regularly updated dataset.

All other feature classes within the NLDP model inherit a common set of attribution from one of these classes, as outlined below.

inspireId represents the unique identifier of the feature or dataset. This attribute follows the INSPIRE identifier standard and consists of a **namespace** (a character string uniquely identifying the data source), a **localid** (a character string or numeric that uniquely identifies a feature within a dataset or a dataset within a series) and a **versionId** (uniquely identifying a particular version of the spatial object or dataset).

validFrom and **validTo** represent the validity of the real-world phenomenon represented by a spatial data feature or dataset. For example, it might represent the date at which a land parcel gained its current land use, or when a site's protected status expires.

beginLifespanVersion and **endLifespanVersion** represent the temporal validity of that specific version of the spatial data feature or dataset representing real-world phenomena. It is not necessarily linked to any real-world change but could reflect the update frequency of a dataset.

For example, a land cover dataset that is updated at the start of every new year would have a LifespanVersion covering, say, 2022-01-01 to 2022-12-31.

sourceDataset a character string uniquely identifying the source dataset.

name (NLDPDataset only) a character string that contains the name of the dataset.

sourceResourceLocator (NLDPDataset only) a URI that can be used to provide a link to the original location of the dataset.

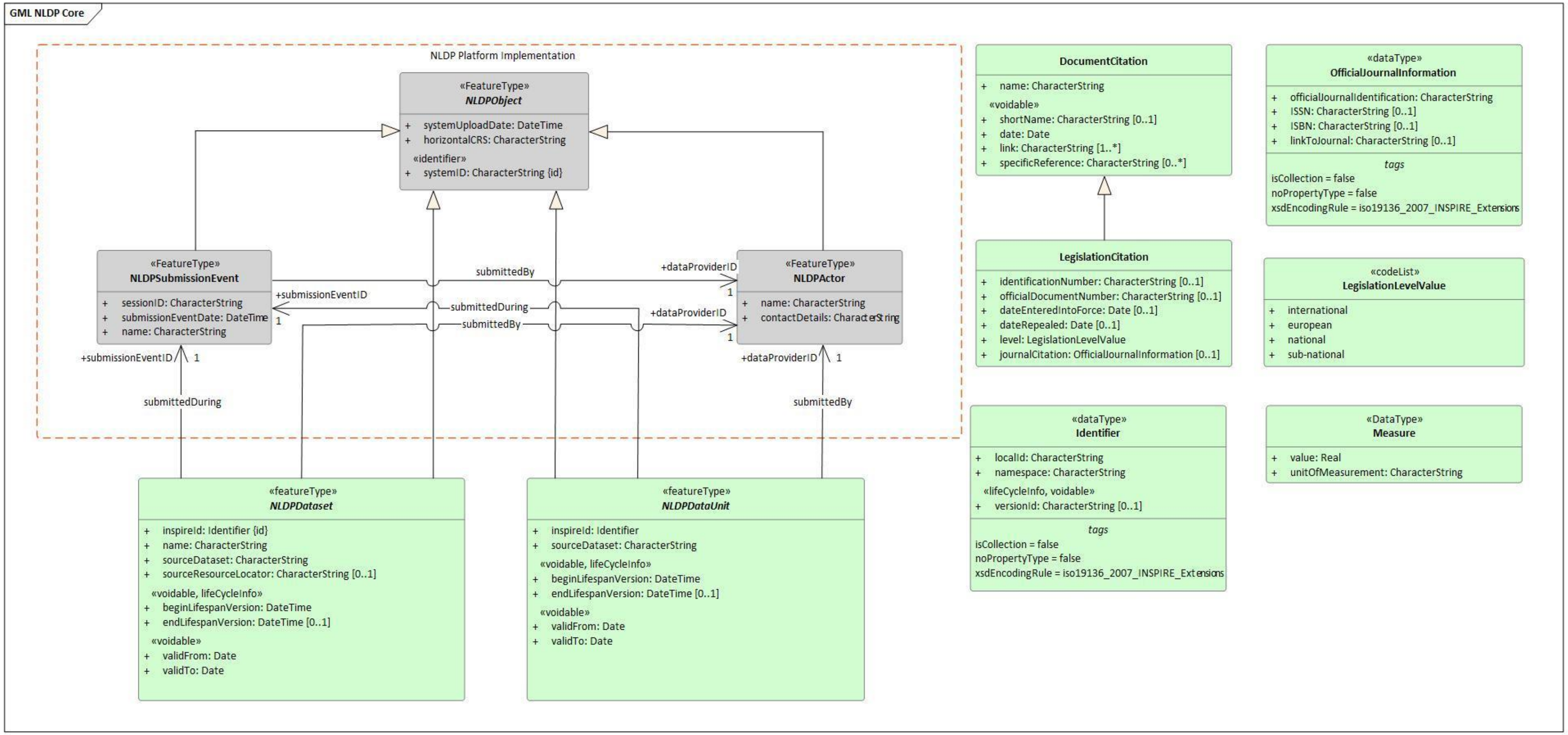

Figure 2 UML diagram of the NLDP Core package

2.4. Land Use

The National Land Data Programme defines Land Use in the same way as the INSPIRE Land Use theme:

“Land Use is defined as the use and functions of a territory. It is the description of land in terms of its socio-economic and ecological purpose. Land Use is different from Land Cover (Annex III, theme number 3), dedicated to the description of the surface of the earth by its (bio-)physical characteristics. Land Cover and Land Use are, however, related and often combined in practical applications. Data sets combining Land Use and Land Cover often emphasise land use aspects in intensively used areas and land cover aspects in extensively used areas. An example of the difference in description of the same piece of land would be an agricultural area with grass. In Land Cover it would be described as a pasture (the cover of the land is grass), while in Land Use it would be classified as agriculture (its use would be for grazing cattle).” (INSPIRE LU, 2013)

This section consists of three separate but related concepts, each of which is explored further in the following subsections. The three concepts are:

- Existing Land Use refers to either the current or past land use and may be represented as vector data (generally polygons but could also be point samples) or raster (gridded) data.
- Potential Land Use represents the land usage that could potentially occur on a parcel. For example, a dataset might represent areas suitable for solar farms, or former industrial areas that could be used for residential development after clean-up. This is not to say that a particular land use is mandated or planned, but just that the parcel has the characteristics that could support that use at some point in the future.
- Planned Land Use represents the various components of a Spatial Plan, which together would dictate how land within its extent is developed.

At the root of this package, three feature classes are defined:

CommonLandUseDataset represents discrete collections of land use features. It inherits the attribution from NLDPDataset while adding:

Extent	Polygon geometry representing the spatial extent of the dataset
nomenclatureDocumentation	A way of referencing an external code list or classification schema. This means each separate dataset can use its own classification scheme

CommonLandUseObject represents vector land use features. It inherits from NLDPDataUnit and adds the following attributes:

geometry	The spatial representation of the land use feature (generally polygons but this is not enforced in the model to allow for other types)
specificLandUseObservation	The land use classification assigned to the feature by the data creator, with optional observation date. This could be a single classification value (class) or multiple values (mosaic). Where the mosaic concept is used, it is possible to assign coverage percentages to the different classifications, and flag the main or

	primary land use classification. A land use observation can also be assigned a level (Above Surface, Surface, Below Surface)
hilucsPresence	The Hierarchical INSPIRE Land Use Classification System (HILUCS) classification that best corresponds to the specificLandUseObservation. Allows a common classification scheme to be applied to different datasets without overwriting the original classifications. Also supports single or multiple classification values as above. If HILUCS does not fit the requirements for an implementation, a different common classification scheme could be substituted here.

CommonLandUseImagery represents raster land use grids. It inherits from NLDPDataUnit and adds the following attributes:

coverage	The raster image holding the gridded land use data

As there are several ways of classifying land use, it was decided to keep these feature classes as flexible as possible. Therefore, the NLDP model supports different approaches to classifying land use. Depending on the use case, it may be desirable to map the classification schemes of different datasets to a common scheme (in this case HILUCS) to allow easy cross-comparison. The model does not enforce any particular method, but rather gives the user the tools to represent land use as they see fit and link supporting documentation directly to the land use data.

This package remains close to the original INSPIRE Land Use model, but there are two key enhancements:

- The concept of "level" is added to allow for land uses above and below, as well as upon, the surface. This is particularly important in urban areas but has utility elsewhere. For example, agricultural land could have solar panels built on it, without hindering its use as pasture for livestock at ground level. Level can hold one of four values:
  - Subsurface
  - Surface
  - Above Surface
  - Unknown

  This approach was chosen over a more relative concept of depth (where features are, say, given a numerical indication of their level with respect to other features in the same dataset) as it was felt to have the most utility when considering land usage. It is generally more useful to be able to say definitively that a land use occurs at, say, ground level, as this will be easier to understand without having to look at other features. Consider a building with a bar in the cellar, a shop on the ground floor and a residential flat on the first floor:

Land Use	Floor	NLDP Level	Relative Level (e.g.)
Bar	Cellar	Subsurface	0
Shop	Ground Floor	Surface	1

Land Use	Floor	NLDP Level	Relative Level (e.g.)
Residential	First Floor	Above Surface	2

Other features in the same dataset might have their ground floor land use set to a relative level of 0 if there is no land use recorded below ground, making a quick determination of the land use at ground level potentially difficult. The NLDP approach avoids this complexity.

- The "mosaic" structure was adopted from the INSPIRE Land Cover theme. This method of articulating multiple land use classifications on one feature allows for a clearer identification of the primary land use, and for classifications to be given different level values (in the example above, where a raised solar farm is built over land used for grazing, the classification "Grazing" could be given a "Surface" level value and "Energy Production" could be given the level value "Above Surface"). This also allowed for a greater harmony between the representations of Land Use and Land Cover, which are frequently seen represented on the same land parcel features in existing datasets.

2.4.1. Existing Land Use

This package adds three features representing past or present land usage (i.e. land uses that have actually occurred or are occurring on a land parcel):

ExisitingLandUseDataset represents discrete collections of past or present land use features. It inherits the attribution from CommonLandUseDataset.

ExistingLandUseObject represents vector existing land use features. It inherits from CommonLandUseObject and features must relate to an ExistingLandUseDataset feature.

ExisitingLandUseImagery represents raster land use grids. It inherits from CommonLandUseImagery and features must relate to an ExistingLandUseDataset feature.

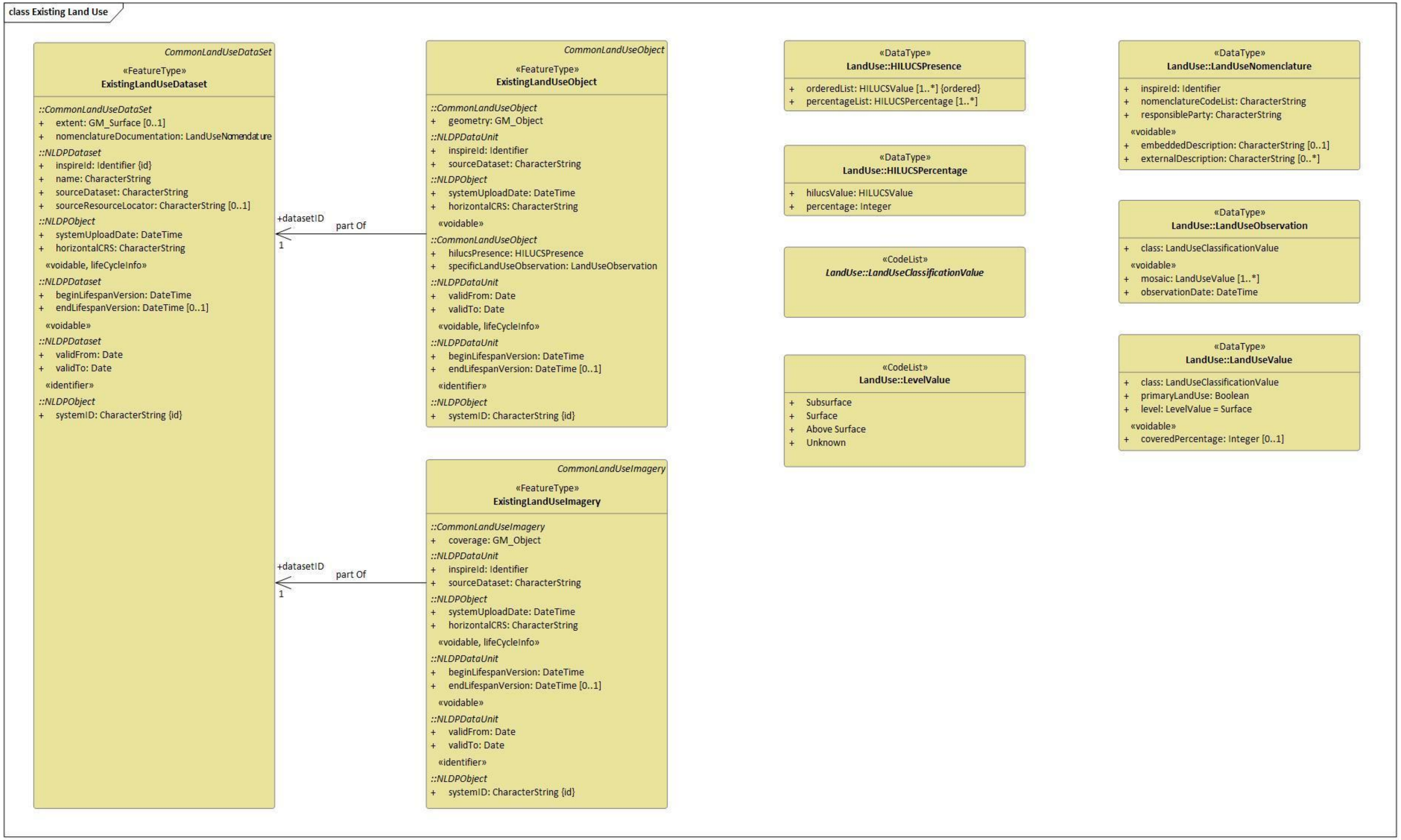

Figure 3 UML diagram of the Existing Land Use package

2.4.2. Potential Land Use

This package adds three features representing potential land usage (i.e. land uses that could occur on a land parcel, but are not currently occurring):

PotentialLandUseDataset represents discrete collections of potential land use features. It inherits the attribution from CommonLandUseDataset.

PotentialLandUseObject represents vector potential land use features. It inherits from CommonLandUseObject and features must relate to a PotentialLandUseDataset feature.

PotentialLandUseImagery represents raster land use grids. It inherits from CommonLandUseImagery and features must relate to a PotentialLandUseDataset feature.

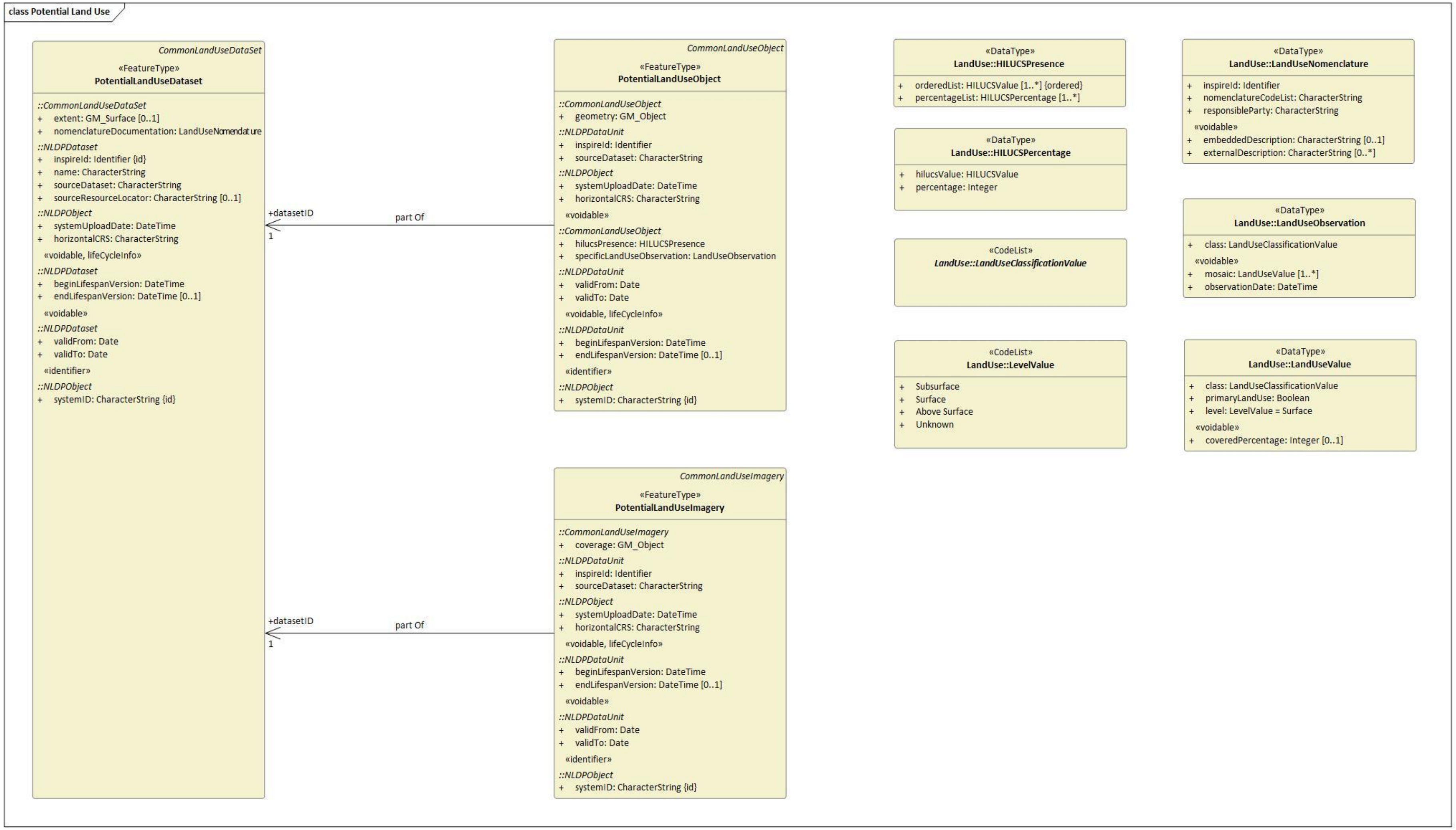

Figure 4 UML diagram of the Potential Land Use package

2.4.3. Planned Land Use

This package remains closely aligned to the original INSPIRE Planned Land Use theme model (INSPIRE LU, 2013). The key concept in this section is that of a Spatial Plan. These may be made by authorities at different sub-national and even inter-national levels, and while there is no standard approach to these plans, they are generally composed of the following elements:

- The overall strategic framework that describes the development aims of the competent authority. This is generally a text document.
- Regulations that may apply across the area affected by the Spatial Plan or to specific zones within it. These dictate what is possible and impossible with respect to land use.
- Cartographic representations of these various elements. These features may have defined boundaries with binding conditions within them or may just be illustrative.

The aim of this package is to provide a framework for representing these elements spatially and in a consistent manner. The feature classes provided for this are as follows:

PlannedLandUseDataset represents a collection of spatial plan features (for example, if a local authority had a number of spatial plans covering different parts of their jurisdiction, with different overall planning objectives) and inherits from NLDPDataset. It has one additional attribute:

extent	Polygon geometry representing the area covered by the collection of spatial plans

SpatialPlan represents a spatial planning document that sets out the overall aims of the competent authority in regards to land use planning within a specified area. It inherits from NLDPDataUnit and adds the following attributes:

extent	Polygon geometry representing the area effected by the Spatial Plan
officialTitle	The official name of the Spatial Plan document
levelOfSpatialPlan	The administrative level at which this plan has been adopted (e.g. Local Authority, National Government, Devolved Government)
planTypeName	The type of spatial plan
alternativeTitle	Any alternative or secondary names given to the plan
processStepGeneral	A general indication of the planning stage at which this plan sits (e.g. scoping, consultation, in force, obsolete etc.)
backgroundMap	Provides a link to the background map used when constructing the plan (e.g. a specific version of OS MasterMap, a builder's plan for a new development etc.)
ordinance	A reference to any relevant administrative ordinance that underpins the plan

ZoningElement represents areas where a particular land use (or several) has been designated through the Spatial Plan. It inherits of NLDPDataUnit and adds the following attributes:

geometry	The polygon geometry representing the area of designated future land use
regulationNature	An indication of the of the legal nature of the regulation underpinning the Spatial Plan
hilucsPresence	The Hierarchical INSPIRE Land Use Classification System (HILUCS) classification that best corresponds to the specificLandUseObservation. Also supports single or multiple classification values as above
specificLandUseObservation	The land use classification assigned to the feature, with optional observation date. This could be a single classification value (class) or multiple values (mosaic). Where the mosaic concept is used, it is possible to assign coverage percentages to the different classifications, and flag the main or primary land use classification. A land use observation can also be assigned a level (Above Surface, Surface, Below Surface)
processStepGeneral	A general indication of the planning stage at which this plan sits (e.g. scoping, consultation, in force, obsolete etc.)
backgroundMap	Provides a link to the background map used when constructing the plan (e.g. a specific version of OS MasterMap, a builder's plan for a new development etc.)
dimensioningIndication	A way to represent the dimensions (generally the area) of the element. Dimensioning can be accomplished with a character string, numerics or using the Measure data type (value and measurementUnit)

SupplementaryRegulation represents locations where a specific regulation limiting or otherwise affecting future land use applies and is supplementary to the zoning elements. They could be area polygons, or could form buffer zones around real-world features (represented by points or lines). It inherits from NLDPDataUnit and adds the following:

geometry	The spatial representation of the location at which a regulation applies. Can be a point, line or polygon
regulationNature	An indication of the of the nature of the regulation underpinning the Spatial Plan
supplementaryRegulation	An indication of the type of supplementary regulation[1]

[1] In INSPIRE, this is an EU-wide managed codelist. For the purposes of NLDP, the nationally-focussed specificSupplementaryRegulation may be sufficient. Alternatively, this could be used to distinguish between UK, devolved and local regulation.

specificSupplementaryRegulation	A more in-depth indication of the type of supplementary regulation (for example, a sub-type of the supplementaryRegulation type above)
processStepGeneral	A general indication of the planning stage at which this plan sits (e.g. scoping, consultation, in force, obsolete etc.)
backgroundMap	Provides a link to the background map used when constructing the plan (e.g. a specific version of OS MasterMap, a builder's plan for a new development etc.)
dimensioningIndication	A way to represent the dimensions (generally the area) of the element. Dimensioning can be accomplished with a character string, numerics or using the Measure data type (value and measurementUnit)
inheritedFromOtherPlans	Boolean. Indicates whether this regulation is derived from a different spatial plan, for example a plan at a higher administrative level
specificRegulationNature	An indication of the of the legal nature of the specific regulation underpinning the Spatial Plan
name	The official name of the supplementary regulation

OfficialDocumentation represents the documentation that underpins the spatial plan. This could be any or all of the legislation, regulations, cartographic representations and descriptive text that together form the spatial plan. It inherits from NLDPDataUnit and adds the following attributes:

legislationCitation	A reference to any relevant legislation, following the structure established in NLDP Core.
regulationText	The full text of the regulation.
planDocument	A reference to any relevant external documentation, for example, scanned text, a raster image or vector geometry.

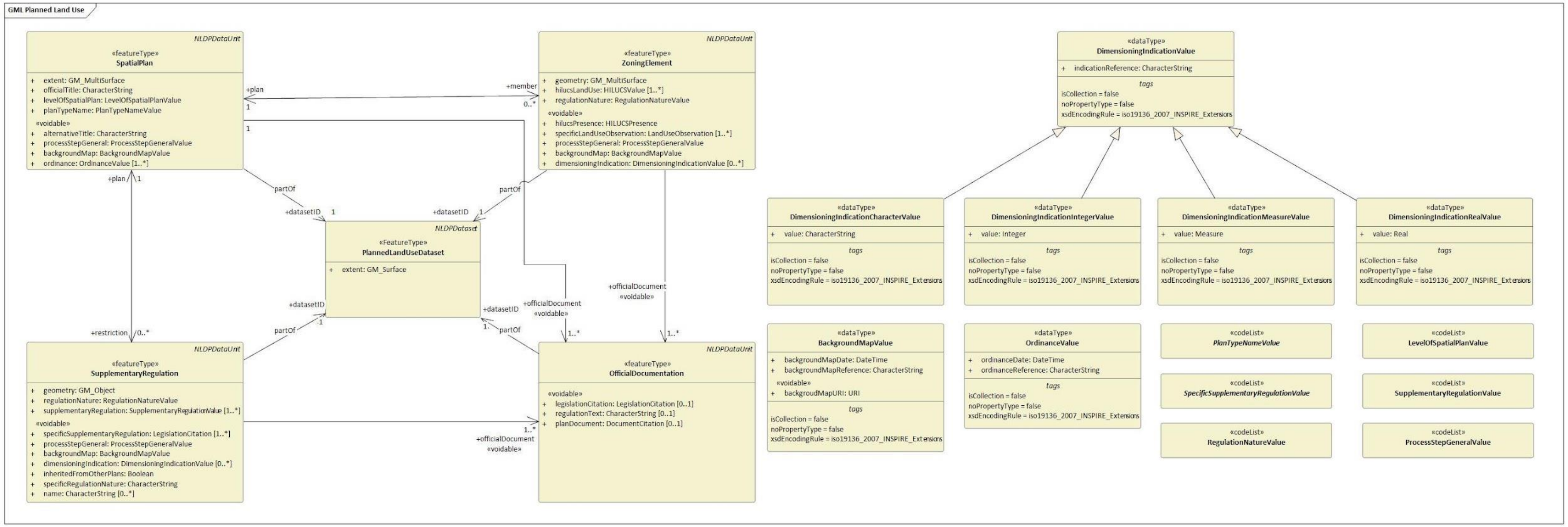

Figure 5 UML diagram of the Planned Land Use package

2.5. Land Cover

Land cover data represents a physical or biological description of the Earth's surface. It differs from land use data which describes how that surface area is utilised.

INSPIRE's Land Cover theme (INSPIRE LC, 2013) is well suited to the identified use cases of the NLDP, so few changes were made. The instantiable feature classes were reorganised to inherit from the NLDP Core feature types and raster imagery has been added to the same package as the vector data representations to harmonise the approaches to the two different types of spatial representation.

The package consists of the three following feature types:

LandCoverDataset represents a collection of land cover units. These units may be points, polygons or raster images. It inherits from NLDPDataset and introduces the following attributes:

extent	Polygon geometry representing the spatial extent of the dataset
nomenclatureDocumentation	A way of referencing an external code list or classification schema. This means each separate dataset can use its own classification scheme

LandCoverUnit represents discrete locations with one or more land cover observations with vector (generally polygon) geometries. It inherits from NLDPDataUnit and adds the following attribution:

geometry	The spatial representation of the land cover feature (generally polygons but this is not enforced in the model to allow for other types)
landCoverObservation	The land cover classification assigned to the feature, with optional observation date. This could be a single classification value (class) or multiple values (mosaic). Where the mosaic concept is used, it is possible to assign coverage percentages to the different classifications. Alongside the class value from the dataset's classification scheme, it is also possible to populate the equivalent value from the EU's Co-Ordination of Information on the Environment (CORINE) land cover classification scheme (if any), to allow for easier cross-comparison between datasets with different classification schemes. CORINE does not have to be used here, another classification scheme could be adopted for this purpose with minimal effort

LandCoverImagery represents raster land cover grids. It inherits from NLDPDataUnit and adds the following attributes:

coverage	The raster image holding the gridded land cover data

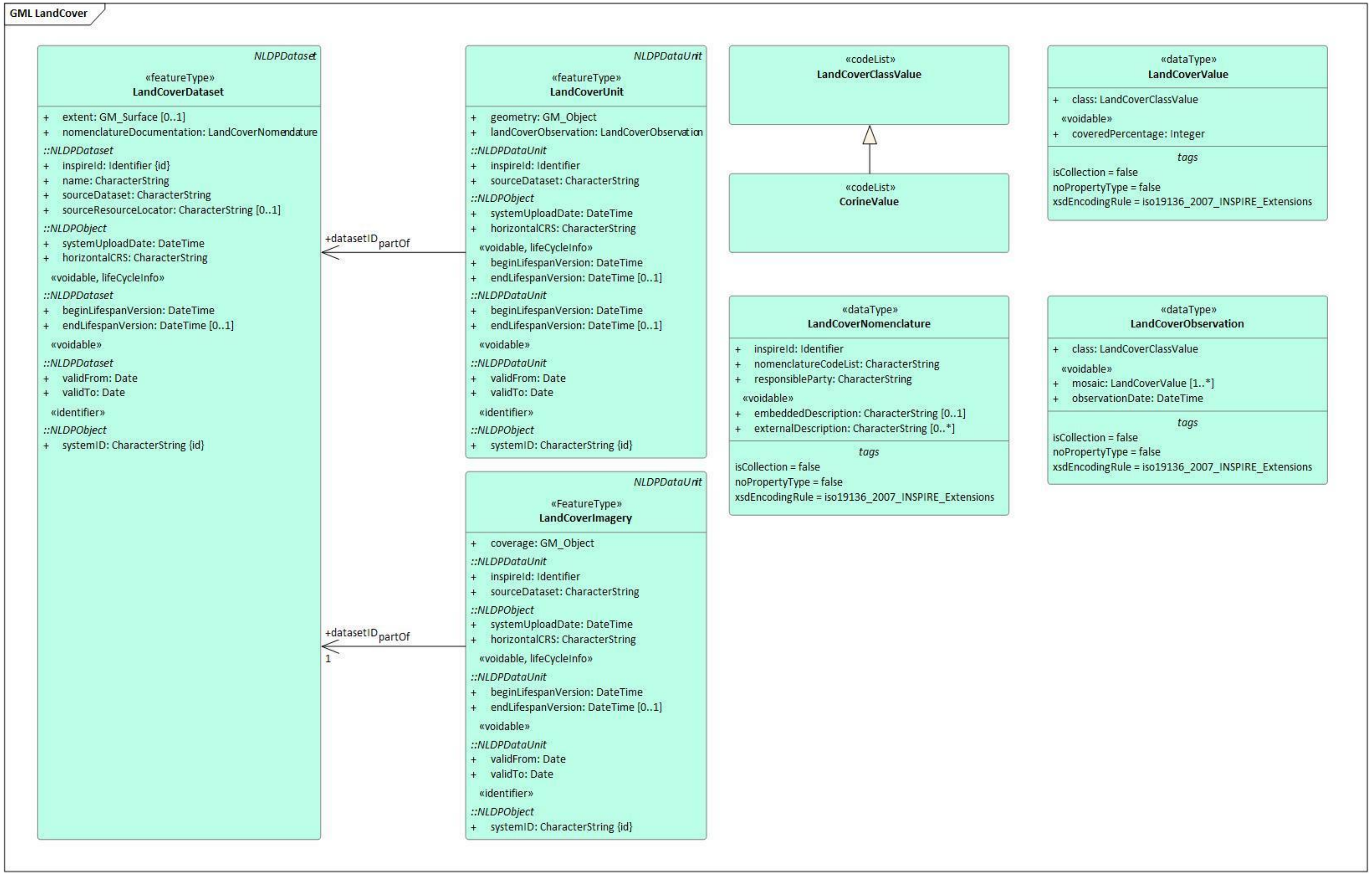

Figure 5

UML diagram of the Land Cover package

2.6. Protected Sites

A Protected Site within the context of the NLDP is an area of land designated for the protection and maintenance of biodiversity, natural resources, culturally important assets, or any other resource deemed to warrant protection. It may be managed through legal or other effective means.

The types of protected site and the reasons for their designation vary greatly. Therefore, this package aims to harmonise the representation of these sites and provide a way to link any external documentation or legislation that is of relevance. It aligns very closely to the existing INSPIRE Protected Sites theme (INSPIRE PS, 2014). The package consists of two feature classes:

ProtectedSiteDataset represents a collection of Protected Sites (such as the complete list of Areas of Outstanding Natural Beauty in England, for example). It inherits from NLDPDataset and adds the following:

extent	Polygon geometry representing the spatial extent of the dataset

ProtectedSite represents the individual sites where protection has been established. It inherits from NLDPDataUnit and adds the following:

geometry	The spatial extent of the site, generally a polygon but this is not enforced to allow for other geographical representations.
legalFoundationDate	The real-world date on which the site was created (i.e. not the date on which the data feature was created)
legalFoundationDocument	A reference to the legislation establishing the site (if any), following the structure established in NLDP Core
siteDesignation	The type of protected site. This attribute can have one or more values and is a complex data type that contains the following child attributes: ● designationScheme – the scheme under which the site is designated. The codelist is populated by a number of common international schemes and can be extended easily ● designation – the type of site as defined by the designation scheme ● percentageUnderDesignation – an integer value representing the proportion of the extent that is covered by the designation value
siteName	The name of the protected site
siteProtectionClassification	The general purpose of the protected site (e.g. Nature Conservation, Cultural etc.)

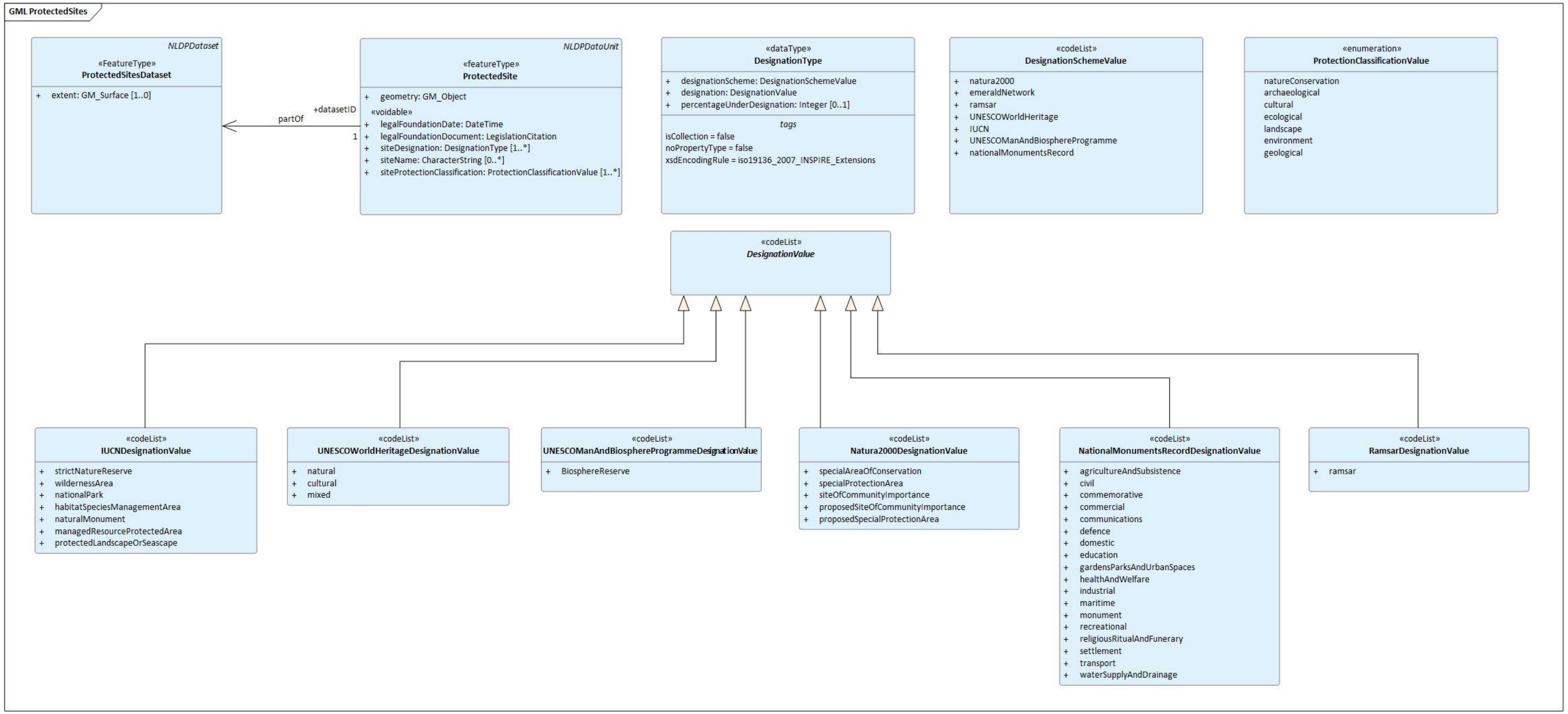

Figure 7 UML diagram of the Protected Sites package

2.7. Risks

Risk Zones are zones where hazardous areas are coincident with highly populated areas and/or areas of environmental, cultural, or economic value. Risk in this context is defined as:

Risk = Hazard x Exposure x Vulnerability

The NLDP has adopted the same pre-existing definitions for each of these terms as the INSPIRE Natural Risk Zone theme (INSPIRE NZ, 2013):

- **Risk (R):** Risk is the combination of the consequences of an event (hazard) and the associated likelihood/probability of its occurrence, (ISO 31010).
- **Hazard (H):** A dangerous phenomenon, substance, human activity or condition that may cause loss of life, injury or other health impacts, property damage, loss of livelihoods and services, social and economic disruption, or environmental damage, (UNISDR 2009).
- **Exposure (E):** People, property, systems, or other elements present in hazard zones that are thereby subject to potential losses, (UNISDR 2009).
- **Vulnerability (V):** The characteristics and circumstances of a community, system or asset, that makes it susceptible to the damaging effects of a hazard, (UNISDR 2009).

While this package has been adapted from INSPIRE's Natural Risk Zone theme, the scope has been broadened to include man-made hazards such as pollution. It contains the following five feature types:

RiskDataset is a collection of risk elements. For example, a dataset on flood risk may include historical flooding extents (ObservedEvent), HazardAreas (areas at risk of flooding), ExposedElements (houses, fields, infrastructure in the vicinity) and RiskZones (surrounding ExposedElements that fall within the HazardAreas). Inherits from NLDPDataset and adds the following:

extent	Spatial extent of all the risk elements that make up the dataset (polygon)

ObservedEvent represents past or ongoing events that have been observed, relevant to the risk being examined. For example, a previously flooded area or the extent of an ongoing oil spill. It inherits from NLDPDataUnit and adds the following:

typeOfHazard	The type of hazard being observed. Consists of hazardCategory (e.g. Flood) a specificHazardType (e.g. Flash Flood)
geometry	The spatial location of the event. Can be a polygon, line, or point
magnitudeOrIntensity	A quantitative and/or qualitative assessment of the hazard. Qualitative assessments are expressed in a character string. Quantitative assessments are expressed as a Measure (value and unit of measurement). Also includes assessmentMethod, which allows an external document to be properly cited as evidence for the assessment.

HazardArea represents the area within which the hazard posed by a phenomenon is non-zero. An example could be the area predicted to flood in a once-in-a-century storm. Inherits from NLDPDataUnit and adds the following:

determinationMethod	Codelisted attribute – either "modelled" (i.e. directly derived from a model) or "indirectly determined" (i.e. inferred indirectly through interpretation of other data)
typeOfHazard	The type of hazard present in the area. Consists of hazardCategory (e.g. Flood) a specificHazardType (e.g. Flash Flood)
geometry	A polygon representing the spatial extent of the hazard area
likelihoodOfOccurrence	An expression of the likelihood of the hazard event occurring. Consists of: ● qualitativeAssessment – a character string ● quantitativeAssessment – the probability of occurrence expressed numerically ● returnPeriod – integer number of years on average between each occurrence
magnitudeOrIntensity	A quantitative and/or qualitative assessment of the hazard. Qualitative assessments are expressed in a character string. Quantitative assessments are expressed as a Measure (value and unit of measurement). Also includes assessmentMethod, which allows an external document to be properly cited as evidence for the assessment.

RiskZone is a spatial representation of the areas where a HazardArea with non-zero magnitude and likelihood intersects an ExposedElement with a non-zero Vulnerability. Inherits from NLDPDataUnit and adds the following attributes:

sourceOfRisk	The type of hazard causing the risk. Consists of hazardCategory (e.g. Flood) a specificHazardType (e.g. Flash Flood)
geometry	The polygon extent of the zone at risk
levelOfRisk	A quantitative and/or qualitative assessment of the risk. Qualitative assessments are expressed in a character string. Quantitative assessments are expressed as a Measure (value and unit of measurement) and can be calculated from the formula Risk = Hazard x Exposure x Vulnerability. Also includes assessmentMethod, which allows an external document to be properly cited as evidence for the assessment

ExposedElement represents people, property, systems, or other elements present in hazard zones that are thereby subject to potential losses. ExposedElements that have a vulnerability to a hazard that's greater than zero will seed RiskZones where they intersect HazardAreas. Inherits from NLDPDataUnit and adds the following attributes:

geometry	The point, line or polygon spatial extent of the exposed element
assessmentOfVulnerability	A complex data type expressing an assessment of the vulnerability of the element to the hazard. Consists of: ● sourceOfVulnerability – as per HazardArea:typeOfHazard ● levelOfVulnerability – a qualitative and/or quantitative assessment of the vulnerability of the element to the hazard ● magnitudeOrIntensityOfHazard – HazardArea:magnitudeOrIntensity ● typeOfElement – composed of a category (e.g. Residential Property) and type (e.g. Bungalow)

These concepts are illustrated well by the following diagram, reproduced from the INSPIRE Natural Risk Zone Data Specification:

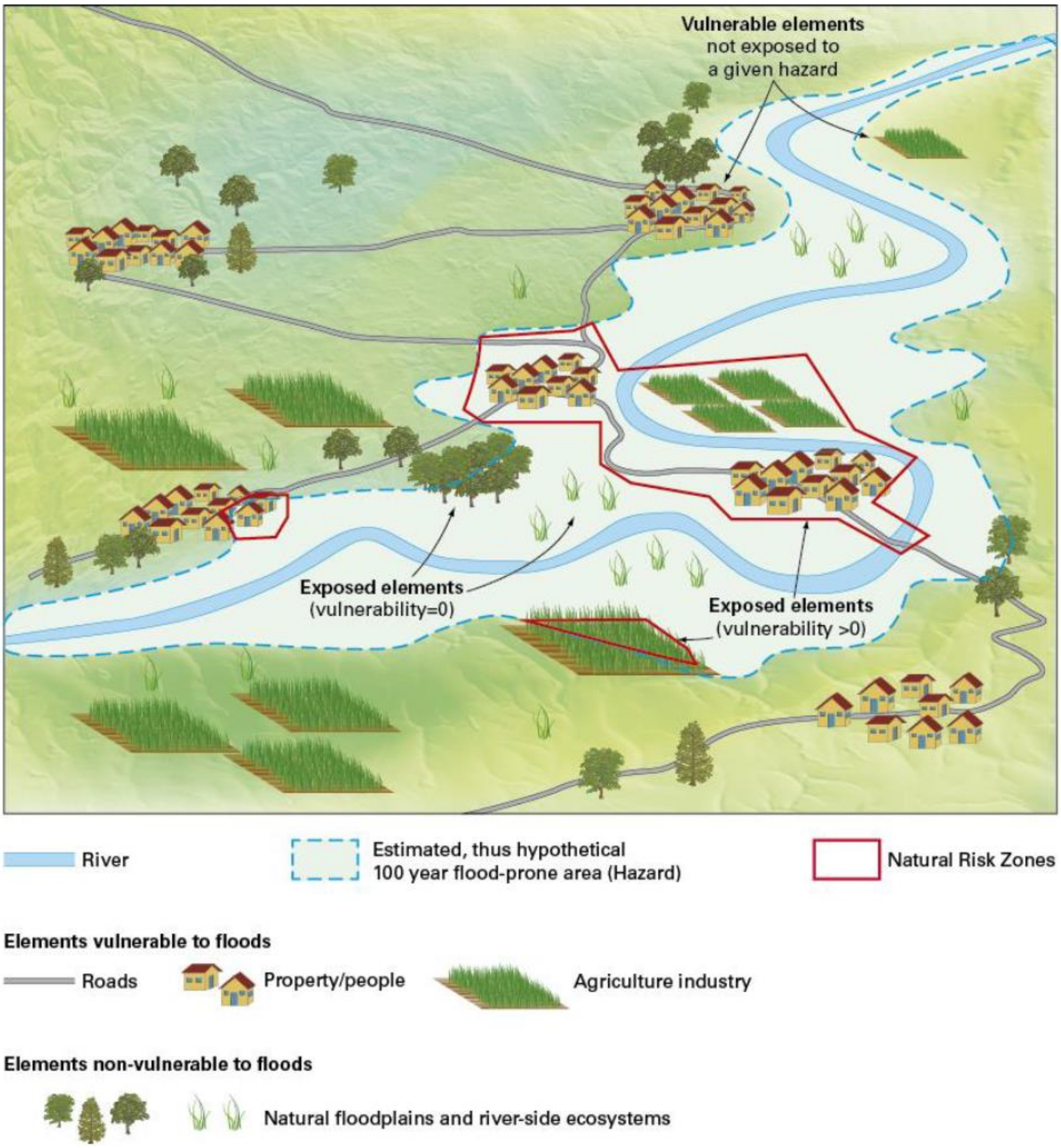

Figure 8 Illustration of a real-world example of the risk elements detailed above. In this case, the predicted extent of the 100-year flood would be represented by a HazardArea, the areas around the houses and agriculture would be RiskZones and the houses and fields themselves would be ExposedElements. The trees are also ExposedElements, but with a Vulnerability (and therefore Risk) of 0. If previously flooded areas were also shown on the illustration, they would be examples of ObservedEvents. (INSPIRE NZ, 2013 p2)

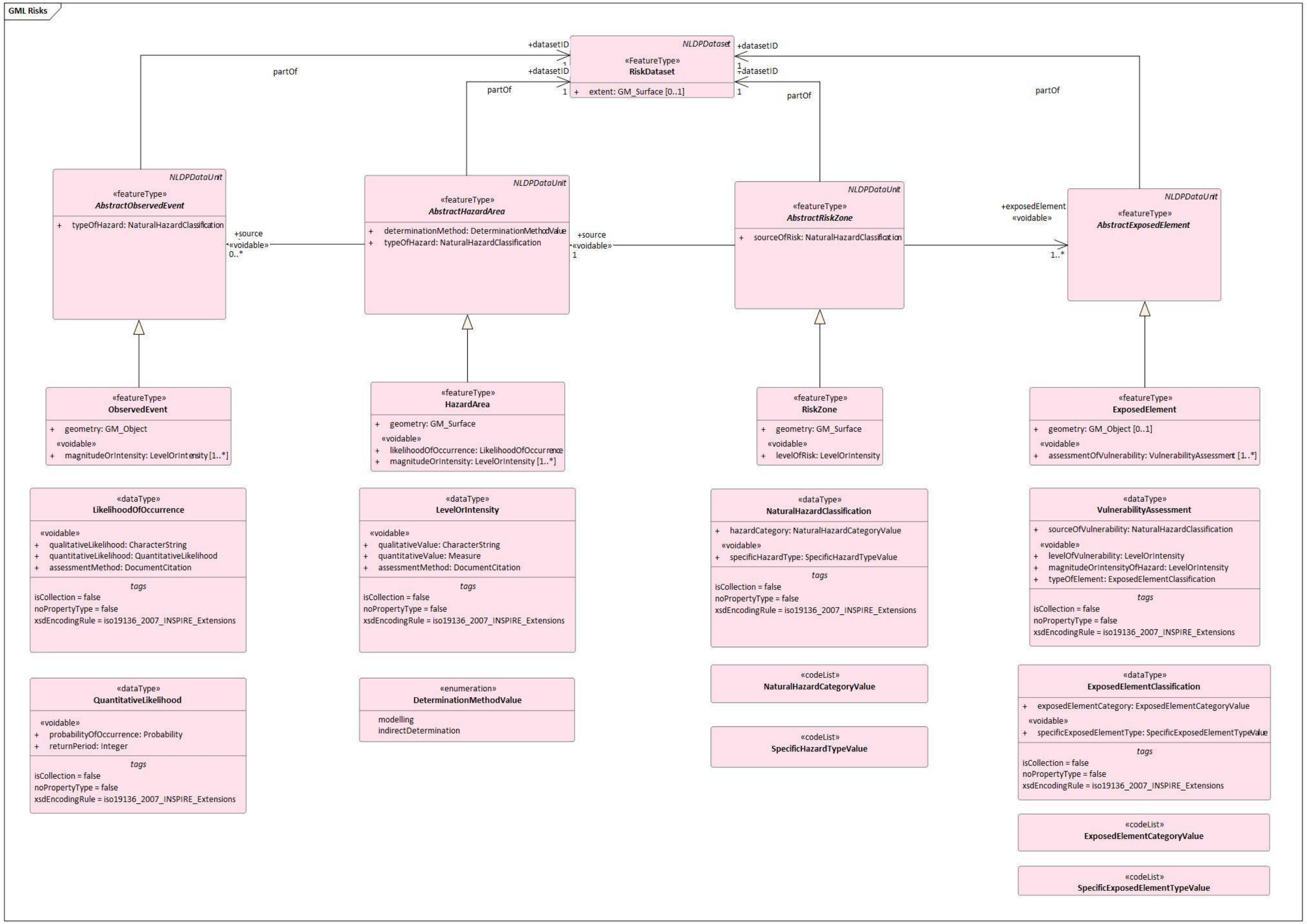

Figure 9 UML diagram of the Risk package

3. Data Quality

The quality of the data presented to end users is critical to the success of any future use of the NLDP model. This section explains a possible approach to data quality, which includes definitions of the terms and how they apply to features, datasets and data providers.

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

- Positional Accuracy
- Completeness
- Domain Consistency
- Currency

Where appropriate ISO 19157 terms are used, otherwise the dimensions are created specifically for the NLDP context.

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

- Identifiers
- Geometry

3.1. Positional Accuracy

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

Requirements on positional accuracy would be implementation-specific, however it is envisaged that an accuracy of a few metres would be sufficient for most land-orientated datasets.

The model does not currently support vertical coordinates, as the Z-axis is only represented in the "level" attribute in the Land Use feature classes and is restricted to simply above surface, surface level and below surface. However, it would be relatively straightforward to introduce a vertical CRS attribute if required by an implementation.

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.

In this document, Completeness considers "omission" in the following contexts:

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

This would require the identification of a minimum set of attributes that are required to meet the use-cases of an implementation. These attributes would likely need to be made mandatory in the implementation's Data Product Specification, and datasets that could not populate these attributes would be rejected.

Completeness of Features: The proportion of features from a Data Owner that are successfully loaded into an implementation of the NLDP model

For a given implementation, it would likely always be advisable to set up some basic data validation on ingestion. This could trap invalid or incorrect geometries, attributes with invalid values, duplicate features etc. In some cases, Completeness might be more important than other aspects of data quality – in these cases the validation rules could be relaxed.

Completeness of Data Supply: For a given area in which NLDP-transformed data is held, the proportion of data that has been received from known Data Owners

It is generally useful to understand where gaps are caused by a lack of data, rather than any real-world consideration. For example, if a local authority has a brownfield dataset that has only been partially digitised, it is important to flag areas where there may be brownfield sites present in the real world, but the corresponding features simply don't exist yet in the dataset. This is difficult to address in a conceptual or even a logical model, but it should be a consideration in any implementation and form a part of the Data Product Specification.

3.3. Domain Consistency

ISO19157: Adherence of values to the value domains.

The NLDP model has adopted the Nomenclature concept from INSPIRE, which allows individual datasets to reference specific domain values. This allows disparate datasets to be collated even if it is not possible or desirable to map their domain values to a unified target set. However, the model does allow for every domain value to be mapped to a common target set if required. This allows for maximum utility as the strictness of domain consistency can be tailored to meet the specific use cases of an implementation.

3.4. Currency

The currency features of INSPIRE have been integrated into the NLDP model. This is managed through two sets of attributes:

- beginLifespanVersion and endLifespanVersion - these attributes refer to the lifecycle of a specific version of a data feature
- validFrom and validTo – these attributes refer to the validity of the real-world feature or observation

This would allow an implementation to store historical and future (planned) versions of datasets, allowing users to step backwards or forwards through time. For example, a user could analyse how land use in a given area has changed over time, or how a given land cover classification has grown or shrunk.

3.5. Identifiers

The INSPIRE ID standard (namespace, local id and version id) was retained throughout the logical model. This allows source feature identifiers to be preserved between different versions of the source datasets.

This also allows for the preservation of any links between features that already exist in the source data. For example, Existing Land Use Objects might be explicitly linked to, say, intersecting or adjacent Protected Sites via their source IDs – by preserving these the model allows for these links to be maintained in an implementation if desirable.

3.6. Geometry

The model generally uses generic (GM_Object) geometry type. This was so that the model can cater for land data in a myriad of formats, and a recognition that existing land use data comes in all shapes and sizes. However, the model structure allows geometry types to be specified at a feature-class level if needed. The model uses ISO 19107 compliant geometry types to ensure consistent implementation and handling of geometries.

Specifying a geometry type at the feature-class level could be a good way of filtering out invalid or incorrect geometries but would likely lead to greater feature rejection rate on ingestion. Again, it is an implementation-specific decision on whether completeness is more important than other aspects of data quality.

4. Conformance

Conformance describes the level to which a feature, dataset, or Data Provider meets the data quality criteria of an NLDP model implementation.

4.1. Mandatory Attributes

The enforcement of mandatory attributes is very much dependent on the use-cases identified for a given implementation. However, a subset of attributes has been identified that would need to be mandatory for all use-cases:

Table 1: Mandatory Attributes

Feature Class	Mandatory attributes
LandCoverUnit	geometry, landCoverObservation (either single value or mosaic values)
LandCoverImagery	geometry
ExistingLandUseObject	geometry, specificLandUseObservation (either single value or mosaic values)
SpatialPlan	geometry, officialTitle
ZoningElement	geometry, specificLandUseObservation (either single value or mosaic values)
SupplementaryRegulation	geometry, name, specificSupplementaryRegulation
OfficialDocumentation	geometry, planDocument
ProtectedSite	geometry, siteName, siteProtectionClassification
ObservedEvent	geometry, typeOfHazard
HazardArea	geometry, typeOfHazard
RiskZone	geometry, sourceOfRisk
ExposedElement	geometry, assessmentOfVulnerability::sourceOfVulnerability

4.2. Completeness

There is generally a balance to be struck between Completeness and other elements of data quality. The stricter the validation requirements, such as around mandatory values, the more difficult it can become to populate an implementation with valid data.

One workaround to this could be to establish conformance levels with increasing levels of strictness. Below is presented an example of a simple conformance level regime:

Table 2: Example of Conformance Level criteria for a Land Cover dataset

Conformance Level	Minimum Data Requirements
Bronze	Valid geometry, land cover class value
Silver	As above, plus survey dates and coverage percentages for multi-class features
Gold	As above, plus a unique source ID

Depending on the implementation, it might be appropriate to add a conformanceLevel attribute at the NLDPDataset level of the model that would allow a platform to indicate the conformance level to the user.

4.3. Frequency of Updates

This is essentially a use-case and dataset specific consideration. In general, it is not envisaged that any dataset within the scope of the NLDP model would be so fluid as to require anything more frequent than monthly updates. Existing and Planned Land Use datasets would likely benefit from 6 monthly updates, particularly in urban areas, while Land Cover datasets might only require updating annually. Historical Land Use datasets would be completely static. Protected Sites might only change very infrequently, such as Areas of Outstanding Natural Beauty. Natural Risk Zones might be static and predictable, or might evolve in response to external factors (e.g. flood risks increasing as a side-effect of development, or subsiding due to mitigating infrastructure).

Any implementation-specific Data Product Specification should set out clearly the update frequency required to support the use-cases, likely at a target feature-class level.

5. Further Considerations on Implementation

The following section aims to collate some further key considerations that must be explored in order to fully utilise the NLDP model.

5.1. Symbology

The requirements for visualising land data will vary greatly by use case. However, one of the key use cases identified at the start of the National Land Data Programme was for national-scale spatial planning, with a particular focus on cross-border (inter and sub-national) analysis. To this end, choosing the right symbology allows users to quickly glean details and highlight features of interest.

Broadly, symbology can be defined at three levels: feature class, dataset and feature. This section will briefly explore scenarios in which each might be appropriate.

- **Feature class-level symbology**: Giving all features within a feature class a common symbology allows users to signpost the different types of data available at a given location, but gives no further detail. It might be used, say, for an overview of the types of datasets used to produce a piece of analysis, or as part of a catalogue view in a data platform.
- **Dataset-level symbology:** Symbolising data at the dataset level could allow users to signpost different sub types of data under analysis (e.g. a user looking at Protected Sites might be interested in Tree Preservation Orders, National Parks and Conservation Areas, all of which could be symbolised differently). As with feature class level symbology, this is likely best suited to providing a high-level overview of datasets in focus rather than to glean any particular insight.
- **Feature-level symbology**: This is likely the most powerful and insightful use of symbology. Features could be symbolised by practically any of their attributes, from land use classification to validity dates, from the administrative level of a spatial plan to the intensity of a hazard. The NLDP model provides a plethora of attribution on which to symbolise and style features, and the choice of which to use would entirely depend on the use case.

Consideration should be given to using stacked symbologies where there are overlapping features or features with multiple (mosaic) classifications. This can allow users to highlight areas where, for example, multiple types of land cover exist within a feature or different land uses occur on different floors of the same building.

An example of this can be seen in figure 8, where land parcels have been classified by their primary (solid line) and secondary land use classifications.

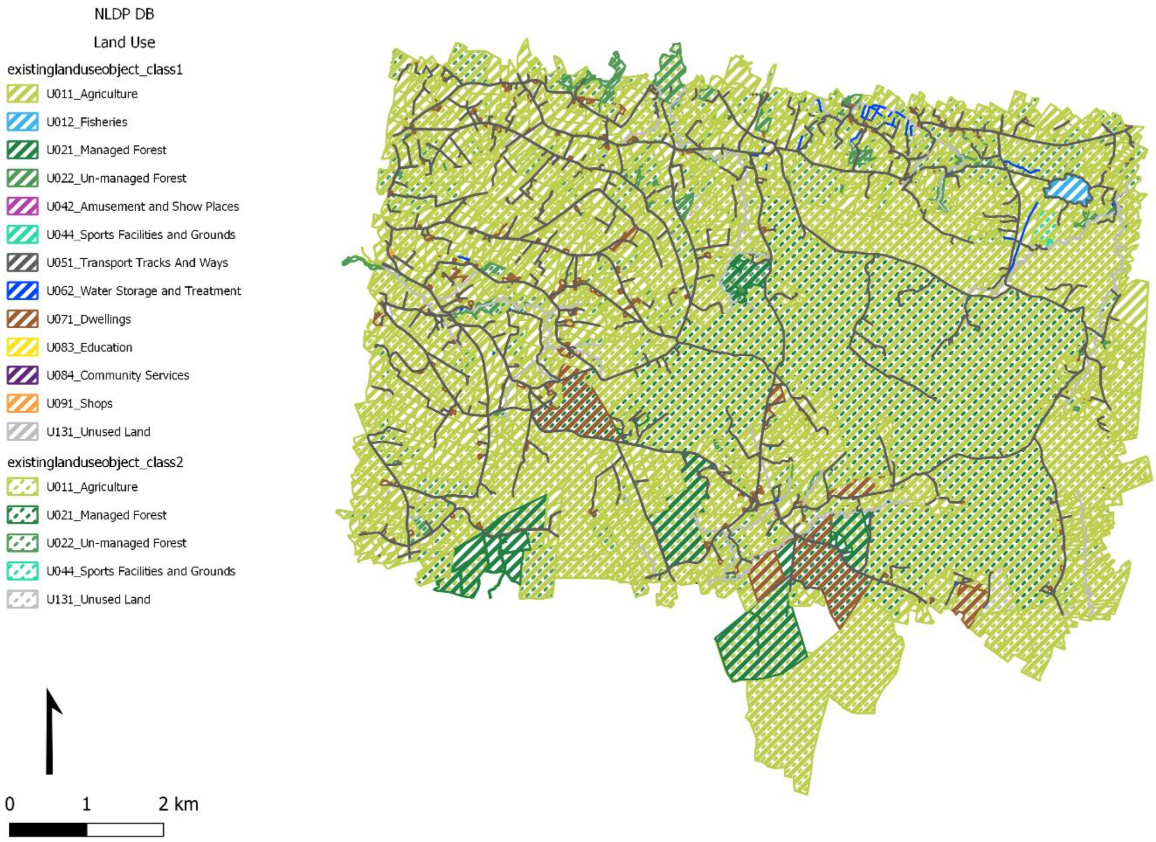

Figure 10 Primary and Secondary Land Use, courtesy of OSNI's FUSION dataset (OSNI FUSION, 2022)

The use of a common classification scheme can be a powerful way of gaining insights from different datasets and/or between different regions. By mapping source classifications to a common scheme (for example HILUCS for Land Use), users can also use different datasets to fill in gaps in coverage. One example of this is shown in figure 9, where a sample from Ordnance Survey Northern Ireland's FUSION land cover dataset has been mapped to the European Environment Agency's CORINE classification scheme and then overlaid on the CORINE 2018 data. This could allow, for example, analysis of several scattered, smaller scale datasets (such as the FUSION sample), with the gaps filled in by larger-scale datasets (such as CORINE).

There are a number of factors to be aware of with this approach. Firstly, even where it is possible to map a dataset's classification scheme to a common one, there may be significant differences in the way the data is captured. The same features could be given different classifications in different datasets. An urban park with trees and a pond might be classified as one patch of urban greenspace, or as individual features for the grass, trees and water. Land cover features may be digitised from aerial imagery and given one classification, or might be clipped to administrative or parcel boundaries (and therefore potentially have multiple classifications per feature). The scale might be different – this is evident in figure 9, where the subnational FUSION dataset has a much higher resolution than the continental CORINE dataset.

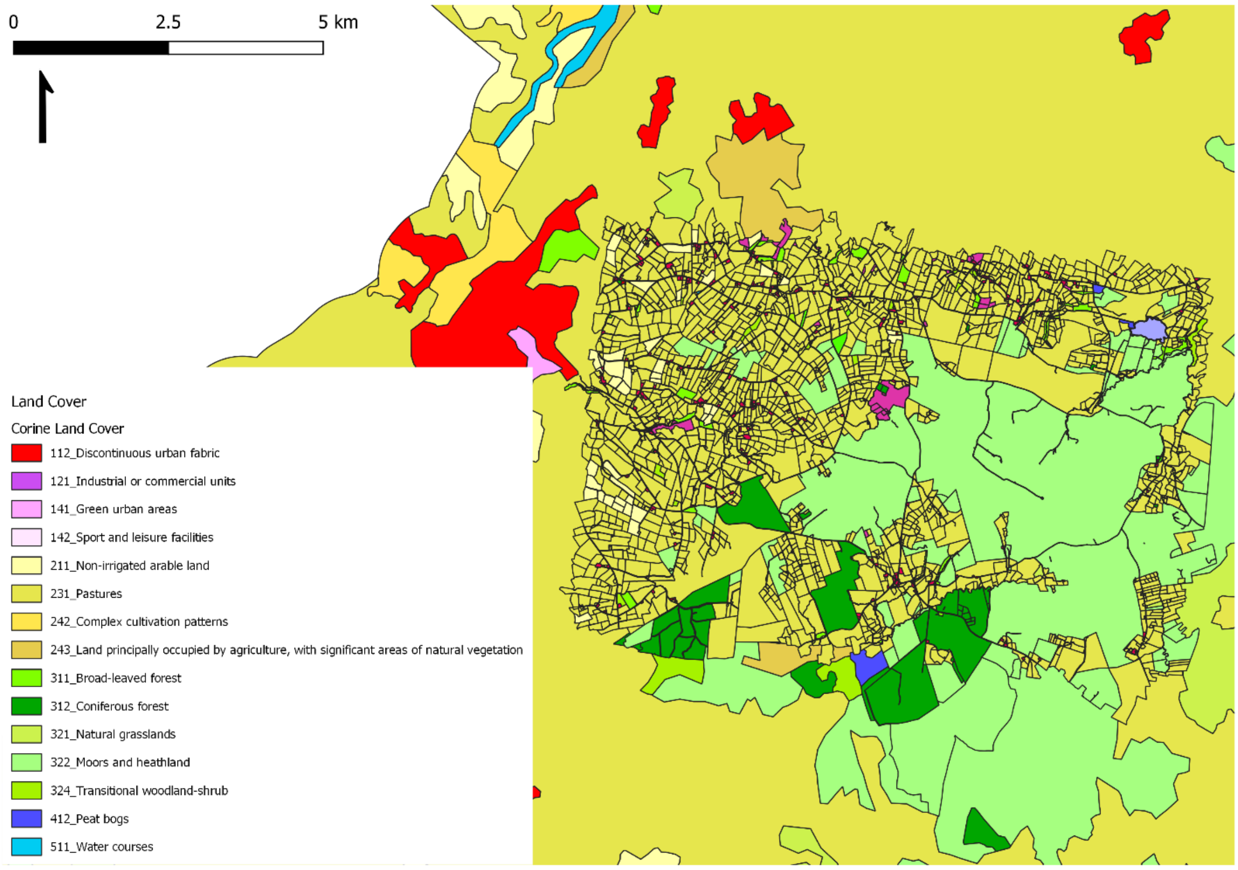

Figure 11 OSNI Fusion Land Cover and EEA's CORINE Land Cover data 2018 near Strabane, Co. Tyrone, Northern Ireland. Both datasets are symbolised using the CORINE land cover classification scheme. The FUSION dataset corresponds to the area of many small polygon features.

None of this should deter users from utilising the NLDP model to compare and aggregate land data, however. As long as the differences between datasets are well understood they can be accounted for, and even used to highlight land management challenges.

5.2. Data Supply

A Data Specification should detail how it is envisaged that data will be ingested into a system. There are broadly three categories of data supply method:

Local Schema Data Supply
The data is supplied to a platform in the form in which it is stored at source. This would require that a process is available to transform the source data to the NLDP target model as part of ingestion. Once transformed, data is stored locally to the platform.

NLDP Schema Data Supply
Data is supplied to a platform already in an NLDP-conformant structure. The data is either transformed from a different source schema by the Data Provider or is created to be NLDP-conformant to begin with. The data is stored locally to the platform.

Federated Data Supply
Data is held by Data Providers and can be accessed through an OGC-compliant web service API. The platform simply serves as a portal through which to access federated data.

6. Initial Data Testing

As part of the development of this model, a number of datasets were identified to be in the scope of the NLDP. These datasets were transformed to the target model using FME (https://www.safe.com/), and the conformance was measured by recording for each dataset which of the NLDP attributes could be populated and which source attributes did not fit into the model. The datasets in question are detailed below:

Table 3: Datasets used in the Phase 1 testing of the NLDP model. Unsuccessful tests shown in italics.

Dataset	Data Owner	NLDP Theme
FUSION[2]	Ordnance Survey NI	Land Use & Land Cover
Land Cover Map 2020	UK Centre for Ecology & Hydrography	Land Cover
CORINE Land Cover 2018	European Environment Agency	Land Cover
Certificates of Immunity	English Heritage Trust	Planned Land Use
Areas of Outstanding Natural Beauty England	Natural England	Protected Sites
Building Preservation Notices	English Heritage Trust	Protected Sites
English Local Authority Green belt dataset	Department for Levelling Up, Housing & Communities	Protected Sites
Listed Buildings England	Historic England	Protected Sites
Local Nature Reserves - England	Natural England	Protected Sites
National Nature Reserves (England)	Natural England	Protected Sites
Historic Flood Map	Environment Agency	Risks – Risk Zone
Historic Landfill	Environment Agency	Risks – Hazard Area
Coalfields	Coal Authority	Risks – Hazard Area
Peaty Soil Location (Zarr)	*DEFRA*	*Land Cover*
Sites of Special Scientific Interest England (Zarr)	*DEFRA*	*Protected Sites*

[2] https://www.nidirect.gov.uk/articles/osni-fusion

The only datasets that could not be transformed were two Zarr-format multi-dimensional arrays, however this was down to the choice of testing platform and the current availability of tools for handling Zarr archives, rather than any structural incompatibility with the NLDP model.

Appendix A: Feature Catalogue

The full Feature Catalogue is available via this link.