SC353DI07171

D04.02.1: NUTS RDF Model revised

Date: 07/03/2018

Document Metadata

Property	Value
Date	2018-03-07
Status	Accepted
Version	1.00
Authors	Makx Dekkers – AMI Consult Ioana Novacean – PwC EU Services
Reviewed by	Nikolaos Loutas – PwC EU Services Willem Van Gemert – Publications Office of the EU Eugeniu Costetchi – Publications Office of the EU Christine Laboudi – Publications Office of the EU Hannes Reuter – Eurostat
Approved by	Susanne Wigard - European Commission, ISA² Programme

This study was prepared for the ISA Programme by:

PwC EU Services

Disclaimer:

Table of Contents

1 INTRODUCTION

The current document describes a revised version of the RDF model[1] for the Nomenclature of Territorial Units for Statistics (NUTS)[2], and the approach to provide persistent URIs for NUTS codes. Besides the details of the NUTS RDF model, the document proposes a set of URI patterns, and a way to preserve links between different NUTS versions.

The publication of NUTS codes as linked data could help users/Eurostat to exploit information which originally is encoded only by postcodes or local administrative units. Other users could track changes over time using links between NUTS versions, and users of e-procurement or e-tendering systems could link their data to persistent NUTS codes to enhance data discovery.

[1] Initially documented under Semic 8, "D04.01_Persistent URIs for NUTS"
[2] http://ec.europa.eu/eurostat/web/nuts/overview

2 NUTS LEGAL BASIS

The Nomenclature of Territorial Units for Statistics (NUTS) acquired a legal status with the European Union Regulation No 1059/2003[3]. This regulation sets out the rules for the NUTS and for future amendments of the classification. The rules for future amendments ensure the consistency of the NUTS through time. The NUTS classification divides the economic territory of the Member States into territorial units and assign to each territorial unit a specific code and name.

As anticipated by the regulation, several amendments were integrated into the initial regulation[4], for example when new countries joined the European Union in 2005 and 2008.

The latest version of the NUTS levels 1, 2 and 3 breakdown will be effective as from 1 January 2018, following the Regulation (EC) No 2066/2016 amending the initial regulation of 2003.

The latest effective version of the NUTS applicable in the European Union was released before 9 December 2013 and the Regulation No 1319/2013 which provided the NUTS levels 1, 2 and 3 breakdown effective as of 1 January 2015[5]. The different NUTS versions can be found on the Eurostat website: http://ec.europa.eu/eurostat/web/nuts/history.

Eurostat is responsible for the development and maintenance of the different versions of the NUTS classification. It approves and processes the changes proposed by the Member States as amendments to the Regulation No 1059/2003.

[3] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32003R1059
[4] http://ec.europa.eu/eurostat/web/nuts/legislation
[5] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32013R1319

3 MODELLING NUTS IN RDF

3.1 Modelling considerations

The following sections introduce the entities that are modelled in this specification. The formal definitions of the entities as classes with their properties are detailed in sections 3.2 through 3.7.

3.1.1 NUTS code

The entity *NUTS Code* is a description of the characteristics of a code, including its notation, its label, its level, its status, additional information and relationships with other codes and with *NUTS Versions*.

It is modelled as a *Concept* as defined in SKOS (class *skos:Concept*): "an idea, a notion or unit of thought".

NUTS Codes are not versioned because their meaning is stable – a code is withdrawn if fundamental aspects change, e.g. when the region that it represents changes.

As *NUTS Codes* are not versioned and *NUTS Versions* (see below) are versioned, a *NUTS Code* may be a member of more than one *NUTS Version*.

Changes in label are modelled using two labels: the 'preferred label' (property *skos:prefLabel*) and the 'alternative label' (property *skos:altLabel*). The preferred label is the label that is/was associated with the *NUTS Code* in the most recent *NUTS Version* of which it is/was a member. All other labels that were associated with the *NUTS Code* in the past are included as *skos:altLabel*.

At this time (February 2018), neither the latest nor earlier versions of NUTS have parallel labels in multiple languages. For multilingual countries, a single label may contain names in more than one language, e.g. "Région de Bruxelles-Capitale/Brussels Hoofdstedelijk Gewest". However, in the current specification, the preferred label is repeatable to allow for future use of parallel, multilingual labels.

Post Codes and Local Administrative Units are made available by Eurostat for download from the NUTS website[6] as Excel files linked from the NUTS-3 codes. These files contain only the Post Codes or LAU codes that correspond to that particular NUTS-3 code.

Relationships between *NUTS Codes* are modelled as 'replacement', 'merging' or 'splitting':

- Replacement is the situation when the *NUTS Code* is changed but the geospatial characteristics remain the same.
- Merging is the situation when two *NUTS Codes* that represent smaller regions A and B are replaced by one *NUTS Code* that represents the region C that combines the regions A and B: C=A+B.
- Splitting is the situation that a *NUTS Code* that represents a region G is replaced by two or more *NUTS Codes*, e.g. D and E, that represent regions that together are equal to region G: D=G-E and E=G-D.

NUTS Codes on a NUTS level with a lower level number are 'broader' than *NUTS Codes* with a higher level number: a *NUTS Code* in NUTS-2 is 'broader' than a *NUTS*

[6] http://ec.europa.eu/eurostat/web/nuts. Post Codes: http://ec.europa.eu/eurostat/web/nuts/correspondence-tables/postcodes-and-nuts; Local Administrative Units: http://ec.europa.eu/eurostat/web/nuts/local-administrative-units

Code in NUTS-3, because the *NUTS Code* in NUTS-2 represents a larger area that includes the area represented by the *NUTS Code* in NUTS-3. Conversely, the *NUTS Code* in NUTS-3 is 'narrower' than the *NUTS Code* in NUTS-2. Following this approach, NUTS-2 Code FRC1 (France—Bourgogne) is 'broader' than NUTS-3 Code FRC11 (France—Côte-d'Or) as the area represented by FRC11 is part of the area represented by FRC1.

3.1.2 Geometry

The entity *Geometry* is a description of the geospatial characteristics of the region represented by the *NUTS Code*, including its type (Regions, Boundaries or Labels), the scale of the data, and the projection as expressed with a code from the EPSG Geodetic Parameter Registry[7].

This entity is modelled as a *Geometry* as defined in the Core Location Vocabulary (class *locn:Geometry*): "a location identified as a point, line, polygon, etc. expressed using coordinates in some coordinate reference system". It is also modelled as a *Dataset* as defined in DCAT (class *dcat:Dataset*): "a collection of data, published or curated by a single agent, and available for access or download in one or more formats".

The modelling as DCAT *Dataset* allows to link to the data file modelled as a DCAT *Distribution*.

3.1.3 Distribution

The entity *Distribution* is a description of the data that is associated with a *Geometry*, including the URL from where the file containing the data can be downloaded, the format of the file and the licence under which the data is available.

It is modelled as a *Distribution* as defined in DCAT (class *dcat:Distribution*): "a specific available form of a dataset".

3.1.4 NUTS version

The entity *NUTS Version* is a description of a published version of NUTS, including the date of its publication, its name and description, and a link to the Regulation that governs its publication.

It is modelled as a *Concept Scheme* as defined in SKOS (class *skos:ConceptScheme*): "an aggregation of one or more SKOS concepts".

*NUTS Version*s are versioned with the year of publication as the distinguishing characteristic. Links to previous and next versions are provided, if such versions exist, using properties *dct:replaces* and *dct:isReplacedBy*.

The relationship between a *NUTS Code* and a *NUTS Version* is included in the description of the *NUTS Code*, using the property *skos:inScheme*. The full set of *NUTS Code*s that are members of a particular *NUTS Version* can be generated by selecting all *NUTS Code*s that link to that *NUTS Version*.

3.1.5 Regulation

The entity *Regulation* is a description of the legal act that governs the publication of a particular *NUTS Version*.

[7] http://www.epsg-registry.org/

It is modelled as a *Legal Resource* as defined in the ontology of the European Legislation Identifier. The link between the NUTS Version and the Regulation is expressed using the property *dct:isRequiredBy*.

This specification does not specify the elements of the description of the *Regulation*. Such a description can be based on the elements in the ELI ontology and other vocabularies.

3.2 Namespaces

The following namespaces are used in the remainder of this section.

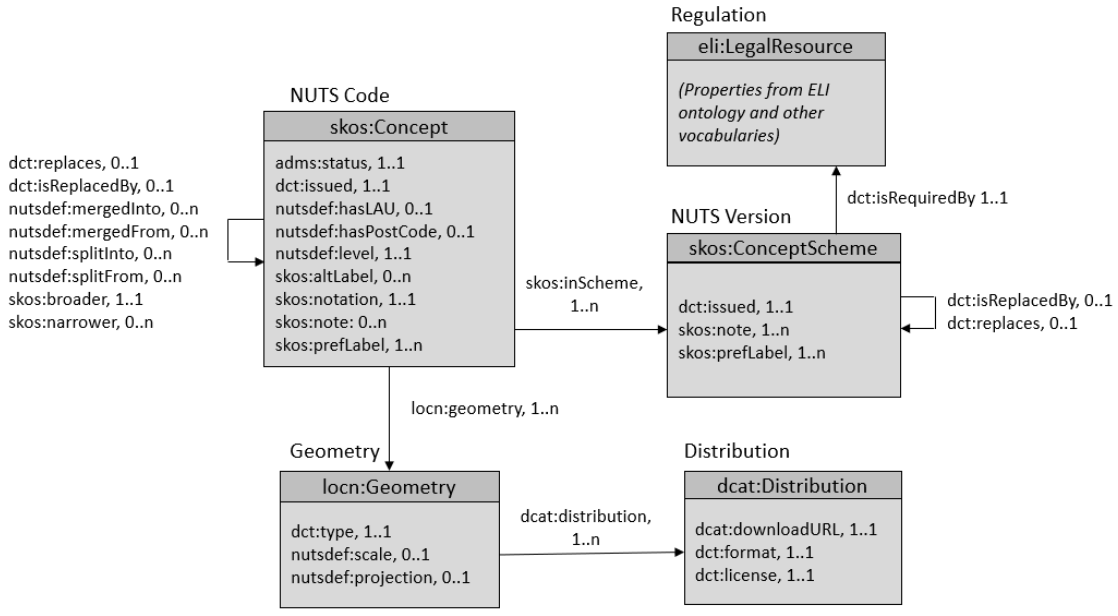
Prefix	Reference	Namespace URI
adms	Asset Description Metadata Schema. https://www.w3.org/TR/vocab-adms/	http://www.w3.org/ns/adms#
dcat	Data Catalog Vocabulary. https://www.w3.org/TR/vocab-dcat/	http://www.w3.org/ns/dcat#
dct	DCMI Metadata Terms. http://dublincore.org/documents/dcmi-terms/	http://purl.org/dc/terms/
eli	European Legislation Identifier. http://publications.europa.eu/mdr/eli/	http://data.europa.eu/eli/ontology#
locn	ISA Programme Location Core Vocabulary. https://www.w3.org/ns/locn	http://www.w3.org/ns/locn#
nuts	Namespace for NUTS instance data	http://data.europa.eu/nuts/
nutsdef	Model elements for the NUTS Linked Open Data model defined in this specification	http://data.europa.eu/<xyz>/; <xyz> to be assigned by the URI Committee
skos	SKOS Simple Knowledge Organization System. https://www.w3.org/2004/02/skos/	http://www.w3.org/2004/02/skos/core#
xsd	XML Schema	http://www.w3.org/2001/XMLSchema#

3.3 Classes

Label	Definition	Class URI
NUTS code	An idea, a notion or unit of thought. https://www.w3.org/TR/skos-reference/#concepts	skos:Concept

Label	Definition	Class URI
NUTS Version	An aggregation of one or more SKOS concepts https://www.w3.org/TR/skos-reference/#schemes	skos:ConceptScheme
Distribution	A specific available form of a dataset. https://www.w3.org/TR/vocab-dcat/#class-distribution	dcat:Distribution
Geometry	A location identified as a point, line, polygon, etc. expressed using coordinates in some coordinate reference system. https://www.w3.org/ns/locn#locn:Geometry This class is at the same time modelled as a Dataset to enable the data to be linked as a Distribution: A collection of data, published or curated by a single agent, and available for access or download in one or more formats. https://www.w3.org/TR/vocab-dcat/#class-dataset	locn:Geometry, dcat:Dataset
Regulation	A legal act or any component of a legal act, like an article. In: http://publications.europa.eu/mdr/eli/documentation/ELI_Ontology-v1.1.pdf	eli:LegalResource

3.4 Conceptual model diagram



3.5 Data types

Table for types of Literals:

Label	Definition	URI
NUTS code notation	A string of characters conforming to the pattern: Two uppercase alphabetic characters from the set of country codes defined at http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Country_codes, followed by zero, one, two or three alphanumeric characters from the set A-Z, 0-9.	nutsdef:notation
Date	An interval of exactly one day in length	xsd:date
String	A sequence of characters	xsd:string

3.6 Properties per class

3.6.1 NUTS Code

Label	Definition	Range	Card.
adms:status	Status of the code in the context of the latest NUTS Version, i.e. all codes that are part of the latest NUTS Version have status "CURRENT", and all codes that are in previous versions but not in the latest version have status "DEPRECATED".	skos:Concept	1..1
dct:issued	Date of formal issuance (e.g., publication) of the resource. http://dublincore.org/documents/dcmi-terms/#terms-issued. This is the issue date of the first NUTS Version in which the described NUTS Code appeared.	rdfs:Literal, datatype xsd:date	1..1
dct:isReplacedBy	A related resource that supplants, displaces, or supersedes the described resource. http://dublincore.org/documents/dcmi-terms/#terms-isReplacedBy. This refers to a new NUTS Code that replaces the described NUTS Code.	rdfs:Resource	0..1

Label	Definition	Range	Card.
dct:replaces	A related resource that is supplanted, displaced, or superseded by the described resource. http://dublincore.org/documents/dcmi-terms/#terms-replaces. This refers to a NUTS Code that is replaced by the described NUTS Code.	rdfs:Resource	0..1
locn:geometry	Associates any resource with the corresponding geometry. https://www.w3.org/ns/locn#locn:geometry. This refers to the description of the geometry associated with the described NUTS Code.	locn:Geometry	1..n
nutsdef:hasLAU	A document that contains a list of the Local Administrative Units corresponding to the described NUTS code.	foaf:Document	0..1
nutsdef:hasPostCode	A document that contains a list of the Post Codes corresponding to the described NUTS code.	foaf:Document	0..1
nutsdef:level	A value that indicates the level of the Code; permissible values are 0, 1, 2 or 3.	rdfs:Literal, datatype xsd:integer (0|1|2|3)	1..1
nutsdef:mergedFrom	This refers to a NUTS Code in a preceding NUTS Version that represented a smaller area that is part of the area represented by the described NUTS Code. Subproperty of dct:replaces.	skos:Concept	0..n
nutsdef:mergedInto	This refers to a NUTS Code in a subsequent NUTS Version that represents an area that includes the area represented by the described NUTS Code. Subproperty of dct:isReplacedBy.	skos:Concept	0..1
nutsdef:splitFrom	This refers to a NUTS Code in a preceding NUTS Version that represented a larger area that includes the area represented by the described NUTS Code. Subproperty of dct:replaces.	skos:Concept	0..1

Label	Definition	Range	Card.
nutsdef:splitInto	This refers to a NUTS Code in a subsequent NUTS Version that represents an area that is part of the area represented by the described NUTS Code. Subproperty of dct:isReplacedBy.	skos:Concept	0..n
skos:broader	A broader concept. https://www.w3.org/TR/skos-reference/#semantic-relations This refers to a NUTS Code on a higher NUTS level than the described NUTS Code.	skos:Concept	0..1
skos:inScheme	A Concept Scheme of which the described Concept is a member. https://www.w3.org/TR/skos-reference/#schemes This is the NUTS Version of which the described NUTS Code is a member.	skos:ConceptScheme	1..n
skos:narrower	A narrower concept. https://www.w3.org/TR/skos-reference/#semantic-relations This is a NUTS Code on a lower NUTS level than the described NUTS Code.	skos:Concept	0..n
skos:altLabel	An alternative lexical label for the described Resource. https://www.w3.org/TR/skos-reference/#labels This is an alternative label for the described NUTS Code. It is used to record previous labels in cases where a code was relabelled.	rdfs:Literal, datatype xsd:string plus language tag	0..n
skos:notation	A string of characters used to uniquely identify a concept within the scope of a given concept scheme, and not normally recognizable as a word or sequence of words in any natural language. https://www.w3.org/TR/skos-reference/#notations This is the character string for the described NUTS Code, e.g. "UKC11".	rdfs:Literal, datatype nutsdef:notation	1..1

Label	Definition	Range	Card.
skos:note	Information relating to a SKOS concepts. https://www.w3.org/TR/skos-reference/#notes This gives additional information about the NUTS Code, for example in case of complex merging or splitting operations. It is repeatable to allow for information in multiple languages.	rdfs:Literal, datatype xsd:string plus language tag	0..n
skos:prefLabel	A preferred lexical label for the described Resource. https://www.w3.org/TR/skos-reference/#labels This is the label for the described NUTS Code. It is repeatable to allow for labels in multiple languages.	rdfs:Literal, datatype xsd:string plus language tag	1..n

3.6.2 Geometry

Label	Definition	Range	Card.
dcat:distribution	Connects a dataset to its available distributions. https://www.w3.org/TR/vocab-dcat/#Property:dataset_distribution This refers to a description of a file that contains the data for the Geometry.	dcat:Distribution	1..n
dct:type	The nature or genre of the resource. http://dublincore.org/documents/dcmi-terms/#terms-type This is the kind of the Geometry. The values must be taken from a controlled vocabulary, see section 3.7.1.	skos:Concept	1..1
nutsdef:scale	The scale or generalisation of the Geometry.	rdfs:Literal, datatype integer	0..1
nutsdef:projection	The projection of the Geometry expressed as a code from the EPSG Geodetic Parameter Registry, http://www.epsg-registry.org/	rdfs:Literal, datatype xsd:string	0..1

3.6.3 Distribution

Label	Definition	Range	Card.
dcat:downloadURL	A file that contains the distribution of the dataset in a given format. https://www.w3.org/TR/vocab-dcat/#Property:distribution_downloadurl This refers to the location where the data can be downloaded.	rdfs:Resource	1..1
dct:format	The file format, physical medium, or dimensions of the resource. http://dublincore.org/documents/dcmi-terms/#terms-format The values must be taken from a controlled vocabulary, see section 3.7.2.	dct:MediaTypeOrExtent	1..1
dct:license	A legal document giving official permission to do something with the resource. http://dublincore.org/documents/dcmi-terms/#terms-license The values must be taken from a controlled vocabulary, see section 3.7.3.	dct:LicenseDocument	0..1

3.6.4 NUTS Version

Label	Definition	Range	Card.
dct:isReplacedBy	A related resource that supplants, displaces, or supersedes the described resource. http://dublincore.org/documents/dcmi-terms/#terms-isReplacedBy This refers to a newer NUTS version.	rdf:Resource	0..1
dct:isRequiredBy	A related resource that requires the described resource to support its function, delivery, or coherence. http://dublincore.org/documents/dcmi-terms/#terms-isRequiredBy This refers to the Regulation that forms the legal basis for the described NUTS version.	rdf:Resource	1..1

Label	Definition	Range	Card.
dct:issued	Date of formal issuance (e.g., publication) of the resource. http://dublincore.org/documents/dcmi-terms/#terms-issued This is the issue date of the described NUTS Version.	rdfs:Literal, datatype xsd:date	1..1
dct:replaces	A related resource that is supplanted, displaced, or superseded by the described resource. http://dublincore.org/documents/dcmi-terms/#terms-replaces This refers to the previous NUTS version.	rdf:Resource	0..1
skos:note	General documentation. https://www.w3.org/TR/2009/NOTE-skos-primer-20090818/#secdocumentation This gives a description of the NUTS version. It is repeatable to allow for descriptions in multiple languages.	rdfs:Literal, datatype xsd:string with language tag	1..n
skos:prefLabel	A preferred lexical label for the described Resource. https://www.w3.org/TR/skos-reference/#labels This is the name of the NUTS Version, e.g. "NUTS 2016". It is repeatable to allow for labels in multiple languages.	rdfs:Literal, datatype xsd:string with language tag	1..n

3.6.5 Regulation

The Regulation under which the NUTS Version is issued is identified with a European Legislation Identifier (ELI) and can be described by properties defined in the ELI ontology[8] and properties from other vocabularies.

3.7 Controlled vocabularies

3.7.1 Type of Geometry

For the type of Geometry, three values are relevant: Regions, Boundaries and Labels. URIs for these three types must be minted as part of a controlled vocabulary.

3.7.2 Format of Distribution

For the format of Distributions, the MDR File type NAL[9] must be used. The formats GeoJSON and TopoJSON should be added to that NAL.

[8] http://publications.europa.eu/mdr/eli/Fregions
[9] http://publications.europa.eu/mdr/authority/file-type/

3.7.3 Licence of Distribution

For the licence of Distributions, the MDR Licence NAL[10] must be used.

3.8 Versioning

The proposed approach is to version the NUTS as a whole, not to version the individual codes.

NUTS Codes are not versioned because semantic changes, e.g. major changes in the territory that the code refers to, lead to creation of a new code and retirement of the old code (e.g. EL2 was replaced by EL6 between NUTS2010 and NUTS2013). Codes that remain the same across NUTS versions can simply have two occurrences of *skos:inScheme* pointing to two *Concept Schemes*.

The only change that occurs, albeit sporadically, is a change in the label associated with a code while the code remains unchanged. An example is the change for code PL2 that had label MAKROREGION POŁUDNIOWY in NUTS2016 and REGION POŁUDNIOWY in NUTS2013.

As it is not necessary to track which label was associated with the code at a specific time, the current label can be expressed with *skos:prefLabel* and previous labels with *skos:altLabel*. This enables finding the correct code when searching on any of the labels that were associated with the code at any time.

3.9 URI patterns

Entity	URI pattern	Example URIs
NUTS Version	http://data.europa.eu/nuts/scheme/<year>	http://data.europa.eu/nuts/scheme/2016
NUTS Code	http://data.europa.eu/nuts/code/<XYnnn>	http://data.europa.eu/nuts/code/ES521
Geometry	http://data.europa.eu/nuts/geometry/<XYnnn>-<type>-<scale>-<projection> with <type> one of (regions, boundaries, labels) and <scale> expressed as 1m, 3m etc.	http://data.europa.eu/nuts/geometry/ES521-regions-1m-EPSG:4258 http://data.europa.eu/nuts/geometry/ES521-labels
Distribution	http://data.europa.eu/nuts/distribution/<XYnnn>-<type>-<scale>-<projection>-<year>.<format> with <format> one of (shape, geojson, topojson, pbf, wkt)	http://data.europa.eu/nuts/distribution/ES521-regions-1m-EPSG4258-2013.geojson

[10] http://publications.europa.eu/mdr/authority/licence/