

Questionnaire on unique identifiers in INSPIRE

Purpose of this questionnaire

Your input to this questionnaire is valuable in that it will help determine and define the Implementing Rules that will complement the implementation of the INSPIRE Directive. This will ensure that your views will be considered when the Directive becomes law and the Implementing Rules Regulations to be applied to your information and data.

INSPIRE and the Implementing Rules of INSPIRE will only apply to data that you publish externally and not to internal datasets and internal data processes.

The purpose of this questionnaire is to provide the INSPIRE Data Specifications Drafting Team the current status of using unique identifiers in your organisation, and to clarify if the proposed structure and restrictions in the lexical rules for unique identifiers are feasible.

The proposed measures are considered to be feasible if the national unique identifiers can be mapped in a unique identifier in INSPIRE.

Timeframe and how to respond

Deadline for submission: 6 July 2007 17:00 CET

It may happen that your organisation uses different systems for unique identifiers for each different product. In this case, please fill as many questionnaires as you think it is useful to give us a clear idea of your practises about and explain us at the end of the questionnaire what is the context of your answer.

Please note that even incomplete answers are useful, so please do complete the questionnaire as far as you can.

Follow up

The results of this survey will enable the drafting Team to decide if the proposed lexical rules:

- are feasible, as they are
- are feasible with some amendments
- are not feasible at all

In the two first cases, a recommendation about lexical rules for "unique identifiers" will be added to document D2.5 "Generic Conceptual model". In the last case, no recommendation on this topic will be added.

Furthermore, an analysis of the results of this survey will be published and made available for all SDIC/LMO.

Background

Article 8 (2) of the INSPIRE Directive requires that for datasets corresponding to the themes listed in annexes I and II, the implementing rules shall address

[...]

(a) a common framework for the unique identification of spatial objects, to which identifiers under national systems can be mapped in order to ensure interoperability between them;"

This means all spatial objects of Annexes I and II shall carry a unique identifier property.

Spatial objects are abstract representations of a real-world phenomenon related to a specific location or geographic area. Spatial objects include vector data (e.g. building, forest, railway) and coverages i.e. grids (e.g. orthophoto, DTM). "Spatial objects" in INSPIRE do not refer the entities of the real world, but their abstractions, sometimes called features. The same entity of the real world may be represented by several features (e.g. at different levels of details), each of them having their own identifier.

- For vector data, "spatial objects" to be identified are the individual features in a database, e.g. each segment of road in a transport dataset.
- For coverages, as there is no need to identify each cell, the "spatial object" to be identified is the file (the dataset).

In the INSPIRE context, "unique identifiers" mean the external identifiers, used to identify spatial objects in a specific database and supplied to the users.

- Internal identifiers (used to manage a database and not supplied to users) are out of the scope of this questionnaire.
- Thematic identifiers (which are the same for a given real-world entities in different databases, as e.g. administrative codes) are also out of the scope of this questionnaire.

Recommended structure of Unique Identifiers

To address the requirement of the Directive, the Drafting Team recommends the following system for "unique identifiers" in particular for spatial objects that belong to a theme in Annex I or II.

Structure of "unique identifiers":

"Unique identifiers" of spatial objects shall consist of two parts:

1. a namespace to identify the data source
2. a local identifier, assigned by the data provider (must be unique within the namespace)

All namespaces shall start with a two letter code.

In case of a data provider associated with a Member State this shall be the ISO 3166 code.

- Example: FR for France, DE for Germany, NL for the Netherlands, etc.
- In case of multinational data providers, a pseudo code that does not conflict with ISO 3166 shall be registered in an INSPIRE register.

All remaining characters of the namespace shall uniquely identify the data source within the Member State (or multinational organisation). Often, it will include one or several of the following information :

- the data provider name (or acronym)
- the product name (or acronym)
- It may also include the spatial object type (e.g. railway, building, parcel, road).

EXAMPLES :

- "FR.IGNF.BDCARTO" may be the namespace used by IGN France for spatial object in their BD CARTO product.
- "NL.TOP10NL" may be the namespace for spatial objects in the TOP10 NL product in the Netherlands.
- "EG.ERM" may be the namespace for spatial objects in the EuroRegionalMap product from EuroGeographics (assuming that "EG" would be the registered abbreviation of EuroGeographics)
- "UK.OSGB" and "UK.OSNI" may be the namespaces used by the Ordnance Survey of Great Britain and Northern Ireland respectively within the UK.
- A road segment may have as unique identifier "FR.IGNF.BDCARTO.1789752321", in the IGNF BD CARTO database.
- A lake may have as unique identifier "UK.OSGB1234567890123456", in the OSGB OS MasterMap data base.

Coverages (grids):

The rules for "unique identifiers" of spatial objects apply also for coverages.

As coverage is typically represented by a file, the current identifier conventions generally apply at file-level. Typically these identifiers are not opaque, but they incorporate 'metadata' (e.g. data product

INSPIRE Data Specifications Drafting Team	Reference: Questionnaire_ui_v2_0.doc	
	2006-09-07	Page 3 of 7

types, timestamps, data origin, sensor information, processing information). However, there is no uniform approach across agencies and data providers. In terms of the structure specified above, the file name is the local identifier.

EXAMPLES :

- A Landsat scene archived in the UK NRC Earth Observation Data Centre may have as unique identifier: UK.NRC.NEODC.LANDSAT_MSS
- A tile of BD ORTHO in IGNF may have as unique identifier: FR.IGNF.BDORTHO.56.2002.0256.2463.IA1

Lexical rules

Recognising the central role of “unique identifiers” and the need to allow for an easy processing of “unique identifiers” in software applications, it is under consideration to restrict the lexical rules for both the namespace and the local identifier according to the following rules:

Rule 1 : both for features (vector data) and for coverages:

- **use only the following set of characters { “A”..”Z”, “a”..”z”, “0”..”9”, “_”, “.”, “,”, “-“ }**

(i.e only capital or small letters from Latin alphabet, digits from 0 to 9, underscore, point, comma, and dash.)

Rule 2 : both for features (vector data) and for coverages:

- **the length of the namespace is limited to 24 characters**

Rule 3 : for features (vector data)

- **the length of the local identifier is limited to 24 characters**

Rule 4 : for coverages

- **the length of the local identifier is limited to 128 characters**

(as the local identifier is the file name, it may require more characters than a vector feature)

Questionnaire

Vector data

1. Does your organisation already use “unique identifiers”?
- Yes No Unknown

Comments or remarks:

2. “Unique identifiers” are strongly recommended for objects belonging to themes in annexes I and II (list below). Please, tell us what is the situation in your organisation for each theme :

- Geographical names
 Data with “unique identifiers” Data without “unique identifier” no data
- Administrative units
 Data with “unique identifiers” Data without “unique identifier” no data
- Transport
 Data with “unique identifiers” Data without “unique identifier” no data
- Hydrography
 Data with “unique identifiers” Data without “unique identifier” no data
- Protected sites
 Data with “unique identifiers” Data without “unique identifier” no data
- Elevation
 Data with “unique identifiers” Data without “unique identifier” no data
- Addresses
 Data with “unique identifiers” Data without “unique identifier” no data
- Cadastral parcels
 Data with “unique identifiers” Data without “unique identifier” no data
- Land cover
 Data with “unique identifiers” Data without “unique identifier” no data
- Geology
 Data with “unique identifiers” Data without “unique identifier” no data

Comments or details:

3. Are your “unique identifiers” binary, strings, numbers?

- Numbers only Strings Binary Other

If “other“, please, explain:

4. If Strings: Which alphabet?

- Latin Cyrillic Greek Other

If your answer is "Other", please tell us which alphabet you use.

5. If Strings: Which separators or other characters are allowed in addition of the proposed set of characters (e.g. ",", ".", "/" etc.)?

6. Which character set (e.g. ISO 8859-1 (Latin-1), {"a" "z", "0 "9"}) do you use for "unique identifiers" ?

How many characters are there in this character set ?

7. What is the maximum length of the "unique identifier", if any?

- Unlimited /

8. Would the set of characters { "A".."Z", "a".."z", "0".."9", "_", ".", ",", "-" } proposed by INSPIRE be acceptable for your local "unique identifiers" ?

- Yes No

Please consider in the case that the proposed set of characters is different than what is used in the SDIC/LMO, would it be possible to map between that character set (e.g. if Cyrillic characters are used) and the proposed character set and vice versa?

- Yes No

If No, please explain:

9. Do you distinguish between different versions of the same entity and if yes, how do you identify them?

Coverages (grids) :

10. Do you use the same character set for identification of coverages as for vector data identification?

Yes No

If no, answer the following questions.

If yes, go directly to question 11.

a) Are your "unique identifiers" binary, strings, numbers?

Numbers only Strings Binary Other

If "other", please explain :

b) If Strings: Which alphabet?

Latin Cyrillic Greek Other

If your answer is "Other", please tell us which alphabet you use.

c) If Strings: Which separators or other characters are allowed in addition of the proposed set of characters

(e.g. ",", ":", "/" etc.)?

d) Which character set do you use for "unique identifiers" ?

How many characters are there in this character set ?

e) Would the set of characters { "A".."Z", "a".."z", "0".."9", "_", ".", ",", "-" } proposed by INSPIRE be acceptable for your local "unique identifiers" ?

Yes No

Please consider in the case that the proposed set of characters is different than what is used in the SDIC/LMO, would it be possible to map between that character set (e.g. if Cyrillic characters are used) and the proposed character set and vice versa?

Yes No

If No, please explain:

11. Do you use more than 128 characters for identification of coverages ?

Yes No

12. Do you distinguish between different versions of the same entity and if yes, how do you identify them?

Namespace :

Within your organisation, which of the following information would need to be part of the namespace to ensure uniqueness of the "unique identifier":

13. Data provider code?

Yes No Unknown

14. Product code of the product the spatial object belongs to?

Yes No Unknown

15. Name of the spatial object type / feature type?

Yes No Unknown

16. Anything else?

17. For the namespace, will you use a different character set than for your local "unique identifiers"?

Yes No

If yes, explain.

18. Do you think 24 characters will be enough for the global namespace in your case? (the global namespace includes the country code and may include the data provider identification)

Yes No

If no, explain why :

Other comments or remarks :