Discussion paper: IR on data interoperability – Change proposal on voidability and multiplicities

Detailed issue addressed
Attributes with multiplicity \([0..1]\) not marked as voidable (section 1.2.1 in the change proposal summary document)

Summary of the issue(s)
In some data models (e.g. for Statistical Units and Population Distribution & demography), there are attributes, which in the model have multiplicity \([0..1]\) and were not marked as voidable. Multiplicity is not translated into Implementing Rules, therefore these attributes are mandatory according to the Implementing Rules\(^1\).

This raises a larger issue about the way the notions of voidability and multiplicities are currently translated into the Implementing Rules.

The concept of voidability (see detailed explanation in the Annex) refers to the principle that all properties\(^2\) defined in the INSPIRE data models that are directly available in the source data or can be mapped to the INSPIRE data models at reasonable costs need to be made available in compliance with the Implementing Rules. Put the other way around, all existing data needs to be shared according to the IRs (as long as the costs for doing so are not unreasonable), but no new data needs to be collected. Therefore, only properties without which a spatial object would not make any sense (e.g. it geometry, type or identifier) should be declared as non-voidable in the data models.

The concept of multiplicity describes how many values a certain property can take in the real world. If a property can have no value (e.g. the \(\text{hasChild}\) property for a person for people without children), the minimum multiplicity should be set to 0. If a property can have more than one value (again, the \(\text{hasChild}\) property for a person with several children), the maximum multiplicity should be set to "any" (*) or the maximum possible value (e.g. an edge will have exactly 2 nodes). However, the appropriate multiplicity is not always that clear, in particular for properties that are "information artefacts", e.g. a description or comments property, which exists only in the source database and not in the "real world". Therefore, multiplicities have not been included in the Implementing Rules.

Using these two concepts, it is possible to distinguish between situations where a value is known to be non-existent (e.g. we know that a person has no children) and a situation where a value is unknown (we don't know whether a person has children or not), unpopulated in the data set or confidential. In the first case, the property \(\text{hasChild}\) would receive no value (be omitted), in the second case it would be declared as "void".

Proposed change(s)
The proposed change in the change proposal is to mark all attributes with multiplicity \([0..1]\) as voidable both in the Implementing Rules and the Technical Guidelines.

Since the concepts of voidability and multiplicities have raised a lot of questions over the years, an alternative could be to fundamentally change the way they are included in the Implementing Rules,

---

\(^1\) Note that this is not an issue in the XML schema encodings, since there the multiplicities are taken into account, and therefore the attributes can be omitted.

\(^2\) attributes or association roles
e.g. by completely removing the concept of voidability and replacing it by the principle that all properties that are directly available in the source data or can be mapped from the source data at reasonable costs need to be made available in compliance with the Implementing Rules.

Possible options
Preferred option (for discussion at MIG-T meeting #48): 2 or 3.

Option 2 would be addressing the change proposal, but option 3 would be removing a lot of ambiguity that still exists around the notion of voidability.

Option 1: No change in the IR
- **Description:** No change will be made to the IR, but it will be clarified in the Technical Guidelines or a legal interpretation note that also non-voidable properties can be omitted if they do not exist in the real world / source database.
- **Impact:** There will be no impact on the IR or on existing implementations, but the situation will still be not very clear (implementers will have to be aware of the clarification in the TG/legal note).

Option 2: Limited IR amendment addressing only the identified issue
- **Description:** The IR (and TGs) will only be amended to mark all attributes with multiplicity [0..1] as voidable.
- **Impact:**
  - The impact on the IR will be relatively small, but the proposed change would also impact on the TGs, data model repository and XML schemas.
  - The impact on existing implementations should be low, since the change to the schemas should be backwards-compatible.

Option 3: IR amendment removing the notion of “voidability”
- **Description:** The amendment will remove the notion of “voidability” from the IR (in Art. 2(20) and all tables in the Annexes). Instead, Art. 4 (Types for the Exchange and Classification of Spatial Objects) will be amended along the following lines (proposed change in red):

  1. For the exchange and classification of spatial objects from data sets meeting the conditions laid down in Article 4 of Directive 2007/2/EC, Member States shall use the spatial object types and associated data types, enumerations and code lists that are defined in Annexes II, III and IV for the themes the data sets relate to.

  2. Spatial object types and data types shall comply with the definitions and constraints and. They shall further include the attributes and association roles set out in the Annexes, where corresponding values are contained in the spatial data sets maintained by the Member States or corresponding values can be derived from existing values at reasonable costs.

This will effectively make all properties conditional, i.e. all properties would have to be provided, but only if they are contained in the spatial data sets maintained by the Member States or corresponding values can be derived from existing values at reasonable costs.

- **Impact:**

The TGs and the XML schemas will be updated accordingly.
o The one-time impact on the IR will be considerable (removing many pages of text) and might not be easy to implement as a "remove each occurrence of x" type of amendment (we may need to replace the complete annexes). But it would greatly improve clarity and make the IR easier to maintain in the future.

o Also the effort to once update all TGs, data models and schemas will be high, but these updates could be dealt with through corrigenda and/or automatic bulk updates.

o There will be no impact on existing implementations serving INSPIRE data, since the schema changes will be backwards-compatible (in XML terms, all elements in the schema will become optional). Contrary there will be impacts on client implementations, as long as they expect the currently non-voidable elements to be present.
Annex – Voidable characteristics (from the DT template)

The «voidable» stereotype is used to characterise those properties of a spatial object that may not be present in some spatial data sets, even though they may be present or applicable in the real world. This does not mean that it is optional to provide a value for those properties.

For all properties defined for a spatial object, a value has to be provided – either the corresponding value (if available in the data set maintained by the data provider) or the value of void. A void value shall imply that no corresponding value is contained in the source spatial data set maintained by the data provider or no corresponding value can be derived from existing values at reasonable costs.

Recommendation 2 The reason for a void value should be provided where possible using a listed value from the VoidReasonValue code list to indicate the reason for the missing value.

The VoidReasonValue type is a code list, which includes the following pre-defined values:

- **Unpopulated**: The property is not part of the dataset maintained by the data provider. However, the characteristic may exist in the real world. For example when the —elevation of the water body above the sea level— has not been included in a dataset containing lake spatial objects, then the reason for a void value of this property would be „Unpopulated“. The property receives this value for all spatial objects in the spatial data set.

- **Unknown**: The correct value for the specific spatial object is not known to, and not computable by the data provider. However, a correct value may exist. For example when the —elevation of the water body above the sea level— of a certain lake has not been measured, then the reason for a void value of this property would be „Unknown“. This value is applied only to those spatial objects where the property in question is not known.

- **Withheld**: The characteristic may exist, but is confidential and not divulged by the data provider.

NOTE It is possible that additional reasons will be identified in the future, in particular to support reasons / special values in coverage ranges.