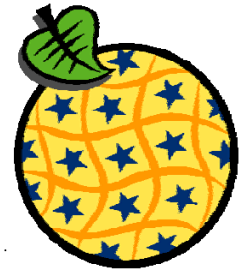
IGN
INSTITUT NATIONAL
DE L'INFORMATION
GÉOGRAPHIQUE
ET FORESTIÈRE



DESIGN AND USE OF MATCHING TABLES TO PREPARE SCHEMA TRANSFORMATION

dominique.laurent@ign.fr
marie.lambois@ign.fr
18/06/2014

Context

- **IGN is a National Mapping Agency**
  - Many data (several themes and LoD) to be made compliant with INSPIRE
  - 2 steps approach:
    - Short-term: transform existing data off-line
      - First, sample data on limited geographic extend
      - Then, whole France
    - Long-term approach: make our internal model close to the INSPIRE one => enable to transform data on-the-fly

- **Themes considered until now (short-term step):**
  - AU - AD - GN
  - TN (road and railway) - CP - HY

IGN
INSTITUT NATIONAL
DE L'INFORMATION
GÉOGRAPHIQUE
ET FORESTIÈRE

DESIGN OF MATCHING TABLES

Design of matching tables: principles

- **Matching tables (Excel) include:**
  - Left side: the INSPIRE model
  - Right side: source model
  - Middle: transformation column

To be filled during matching table meetings

- **The purpose is to derive the left side from INSPIRE data models**

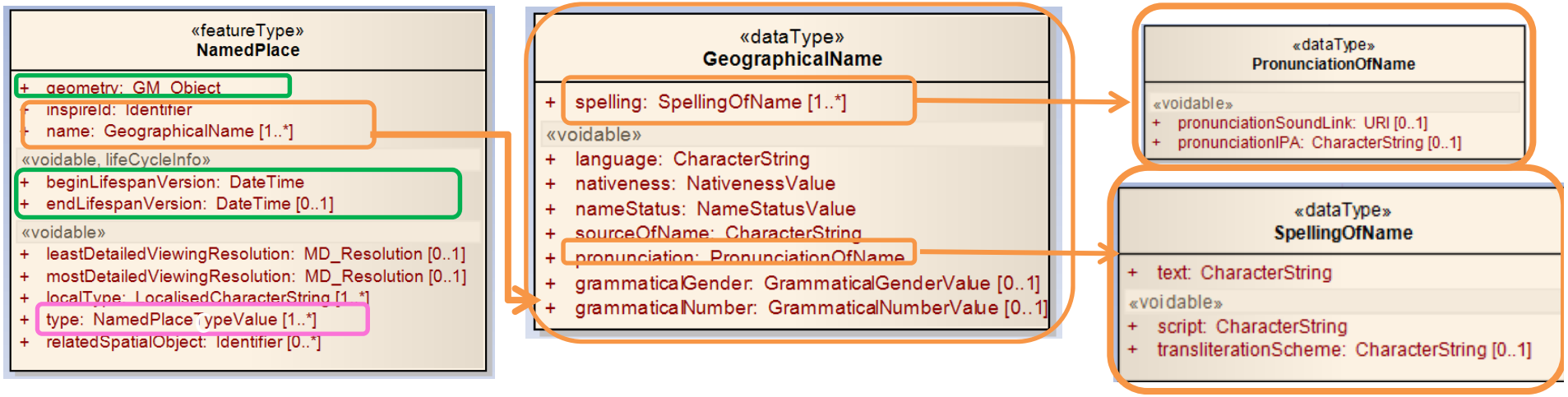
GeographicalNames										Transformation	Source :				
Classe	Element	Type complexe	Documentation	Attribut Lien Contrainte	Attribut Lien Contrainte documentation	Type	Nombre minimal d'occurrences	Nombre maximal d'occurrences	Voidable?		Classe	Attribut Lien Contrainte	Attribut Lien Contrainte documentation	Type	Remarques
NamedPlace	name	Geographical Name	Proper noun applied to a real world entity.	language	Language of the name, given as a three letters code, in accordance with either ISO 639-3 or ISO 639-5. NOTE 1More precisely, this definition refers to the language used by the community that uses the name. NOTE 2 The code "mul" for "multilingual" should not be used in general. However it can be used in rare cases like official names.	string	1	1	VRAI						
				nativeness	Information enabling to acknowledge if the name is the one that is/was used in the area where the feature is situated at the instant when the name is/was in use.	gn:NativenessValue	1	1	VRAI						

INSPIRE model

Transformation operations

Source model

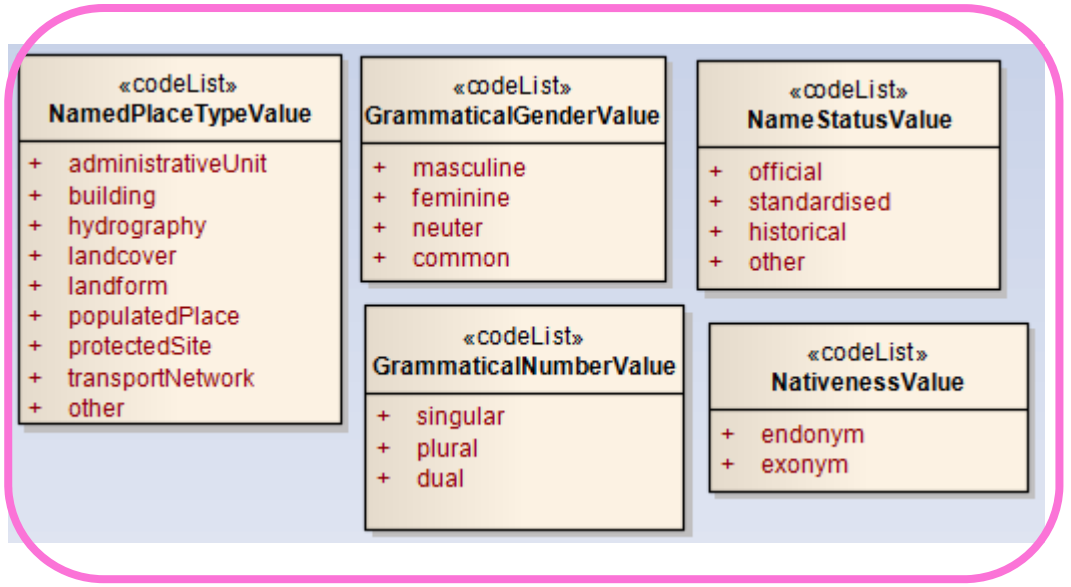
Design of matching tables: principles



Simple attributes (character strings, numbers, date, boolean, geometry)

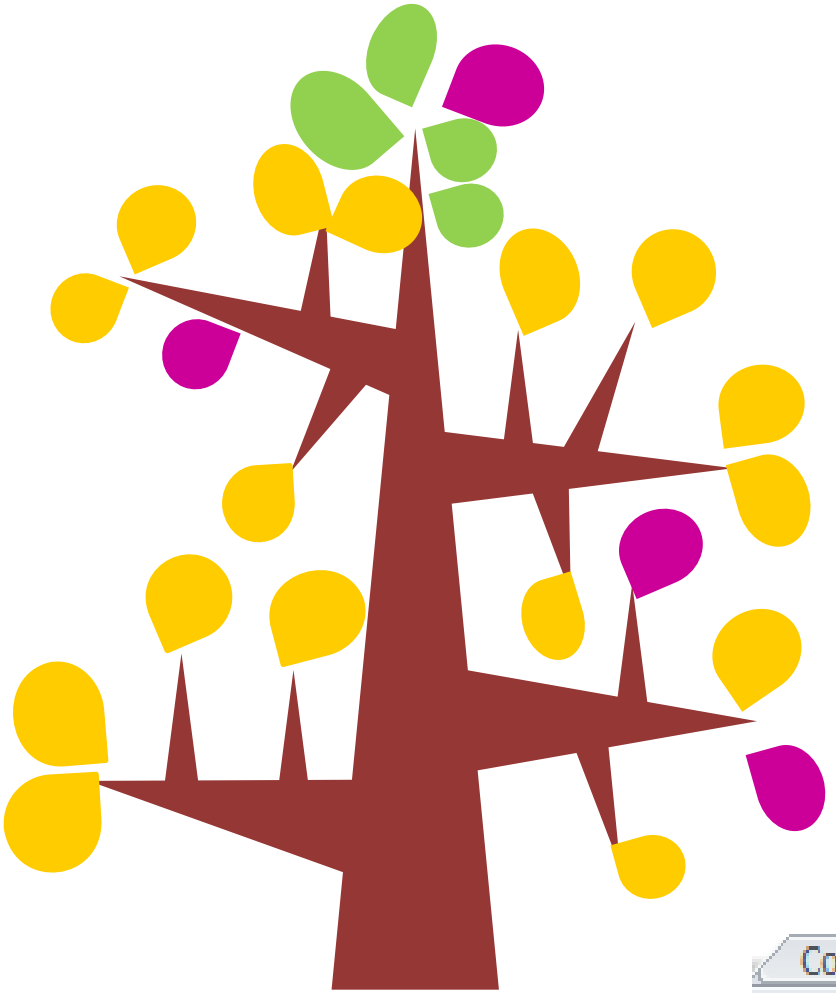
Attributes as data types

Transformation may be done only at « atomic » level, on simple attributes or on values of code lists



Attributes as code lists: transformation generally to be done for each value

Design of matching tables: principles

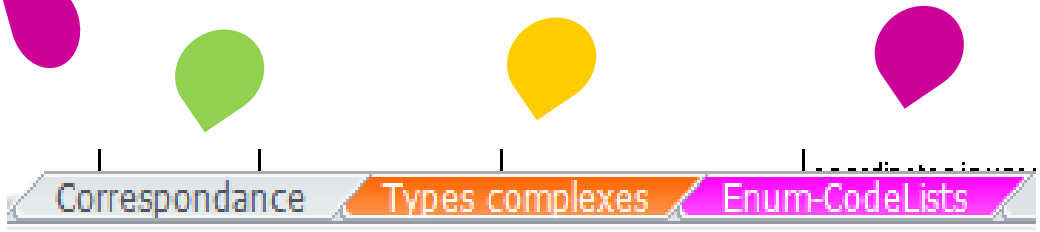


Feature types

Simple attributes (in feature types)

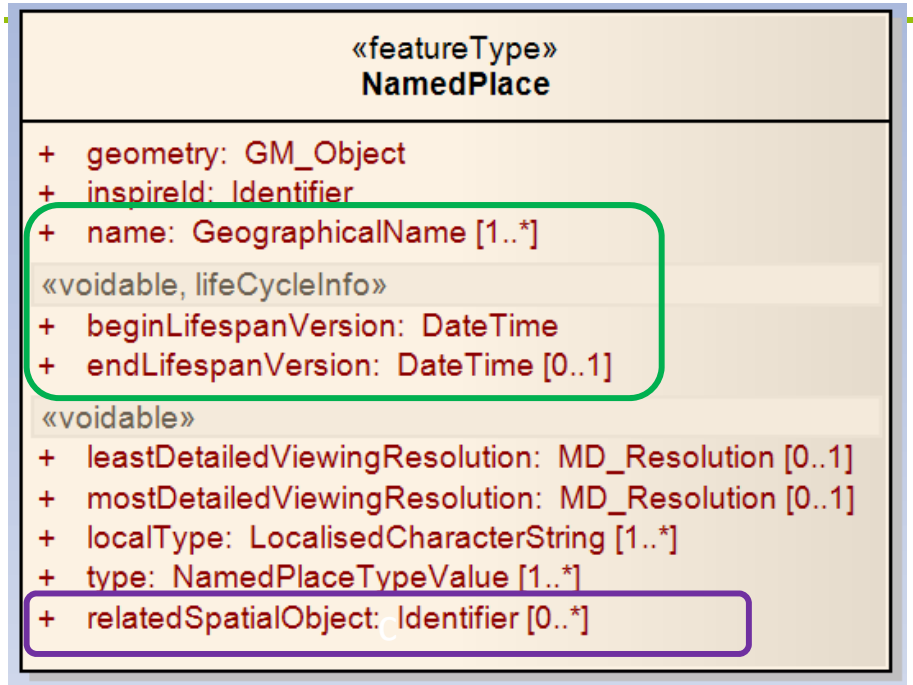
Simple attributes in data types)

Attributes as code lists



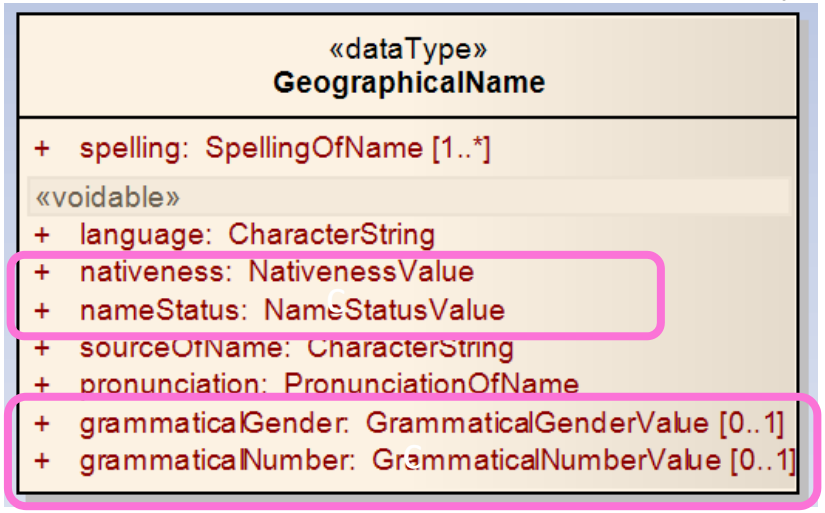
Matching table includes 3 sheets

Design of matching tables: GN example



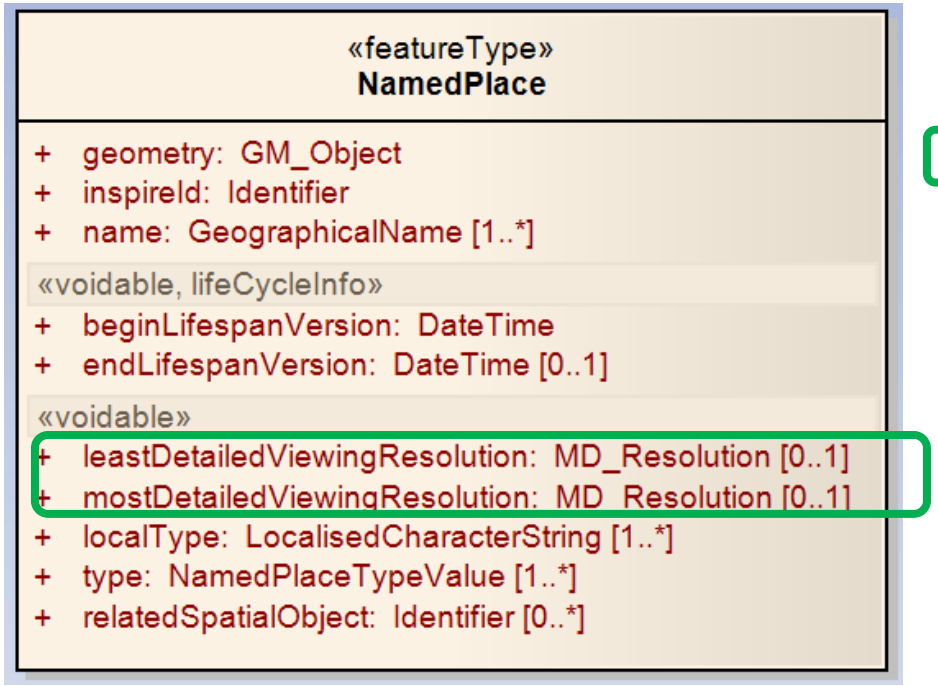
Attributes and data types : basic or defined in GN data model

Data type defined in Generic Conceptual Model (Base types)

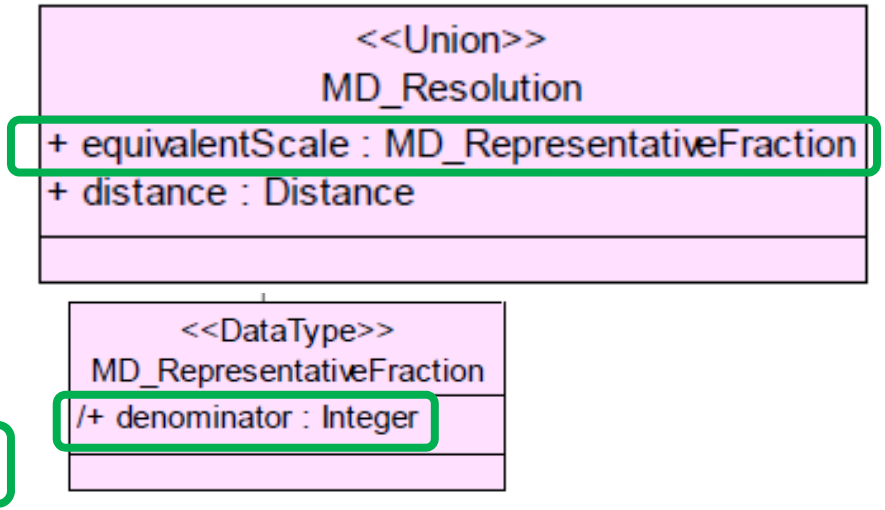


Code lists defined in theme GN (but not in the XSD file)

Design of matching tables: GN example



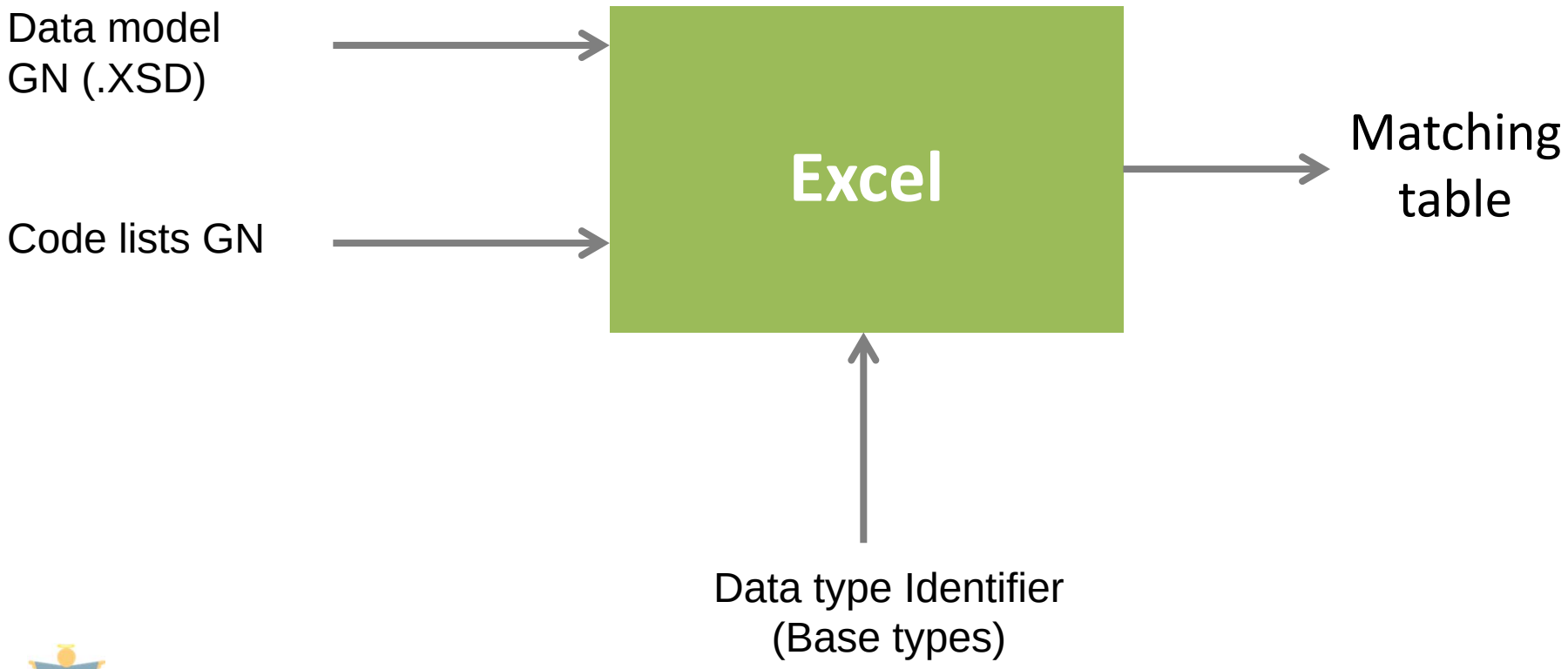
A data type coming from ISO 19115



The data type MD_Resolution is quite simple => no importation in Excel , filled directly in the transformation column.

GeographicalNames				
Classe	**Attribut** **Lien** **Contrainte**	**Type** **types complexes** **énumérations**	**Transformation**	**Classe**
NamedPlace	leastDetailedViewingResolution	gmd:MD_Resolution	equivalentScale=50000	Zone d'activité ou d'intérêt
	mostDetailedViewingResolution	gmd:MD_Resolution	equivalentScale=2000	

Design of matching tables: GN example



Design of matching tables: GN example

GeographicalNames							Source :		
Classe	Attribut Lien Contrainte	Type types complexes énumérations	Nombre	Nombre max	Voidable ?	Transformation	Classe	Attribut Lien Contrainte	Remarques
NamedPlace	beginLifespanVersion	dateTime	1	1	VRAI				
	endLifespanVersion	dateTime	0	1	VRAI				
	geometry	gml:GeometryPropertyType	1	1	FAUX				
	inspireId	base:IdentifierPropertyType	1	1	FAUX				
	leastDetailedViewingResolution	gmd:MD_Resolution	0	1	VRAI				
	localType	gmd:LocalisedCharacterString_PropertyType	1	inf	FAUX				
	mostDetailedViewingResolution	gmd:MD_Resolution	0	1	VRAI				
	name	gn:GeographicalNamePropertyType	1	inf	FAUX				
	relatedSpatialObject	base:Identifier	0	inf	VRAI				
	type	gn:NamedPlaceTypeValue	1	inf	VRAI				

« Correspondance » sheet for feature types and their direct attributes

IGN

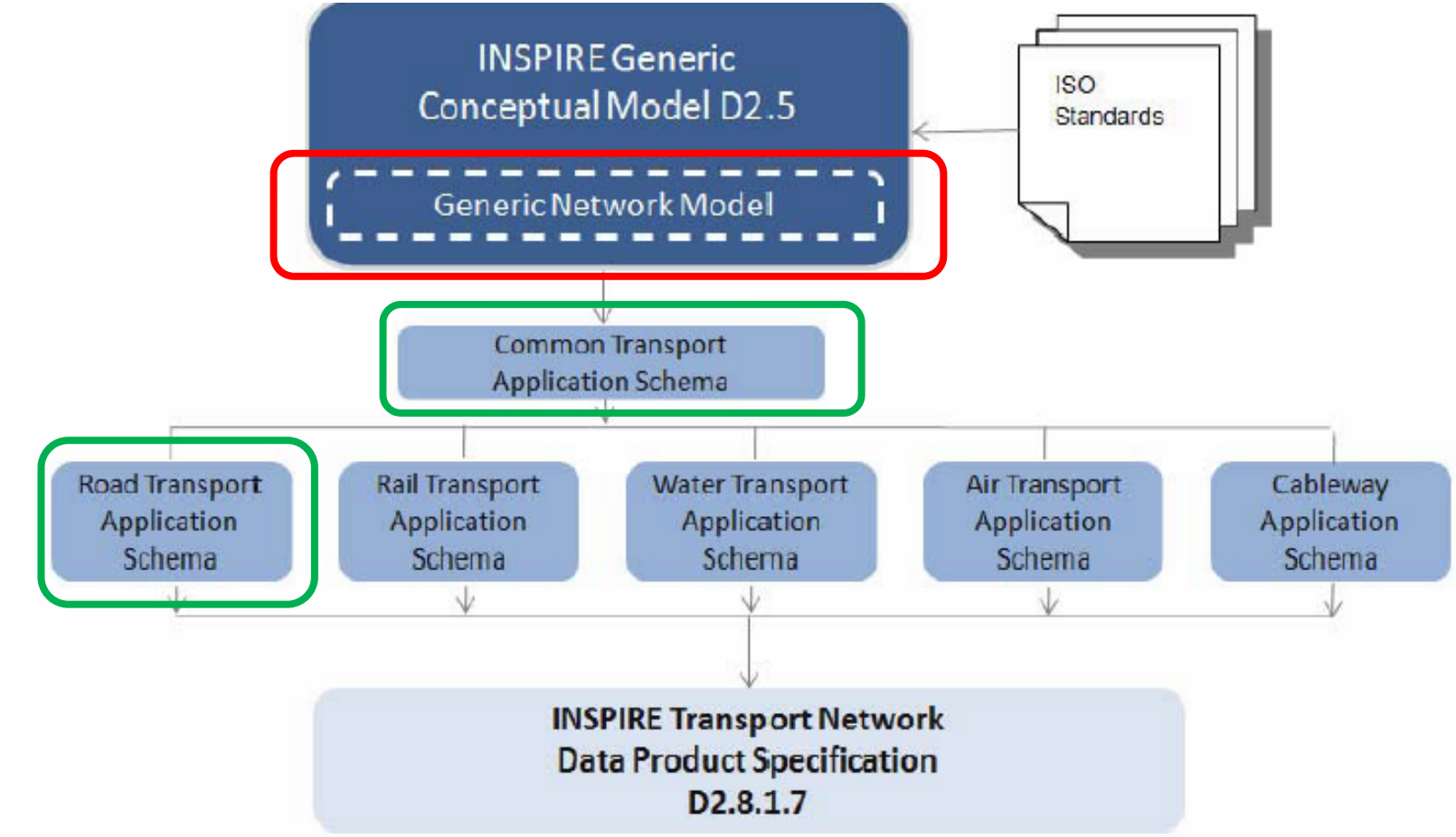
Design of matching tables: GN example

GeographicalNames									Source :		
Classe	Element	Type complexe	Attribut Lien Contrainte	Type	Nombre mini	Nombre maxim	Voidable ?	Transformation	Classe	Attribut Lien Contraint	Remarques
NamedPlace	name	GeographicalName	language	string	1	1	VRAI				
			nativeness	gn:NativenessValue	1	1	VRAI				
			nameStatus	gn:NameStatusValue	1	1	VRAI				
			sourceOfName	string	1	1	VRAI				
			pronunciation	gn:PronunciationOfName	1	1	VRAI				
			spelling	gn:SpellingOfNamePropertyType		unbounded					
			grammaticalGender	gn:GrammaticalGenderValue	0	1	VRAI				
			grammaticalNumber	gn:GrammaticalNumberValue	0	1	VRAI				
	name/GeographicalName/spelling	SpellingOfName	text	string							
			script	string	1	1	VRAI				
			transliterationScheme	string	0	1	VRAI				
	name/GeographicalName/pronounciation	PronunciationOfName	pronunciationSoundLink	URI	0	1	VRAI				
			pronunciationIPA	string	0	1	VRAI				
NamedPlace	InspireId	Identifier	localId	string	1	1					
			namespace	string	1	1					
			versionId		0	1	VRAI				
NamedPlace	relatedSpatialObject/InspireId	Identifier	localId	string	1	1					
			namespace	string	1	1					
			versionId		0	1	VRAI				

« Types complexes» sheet for data types and their attributes

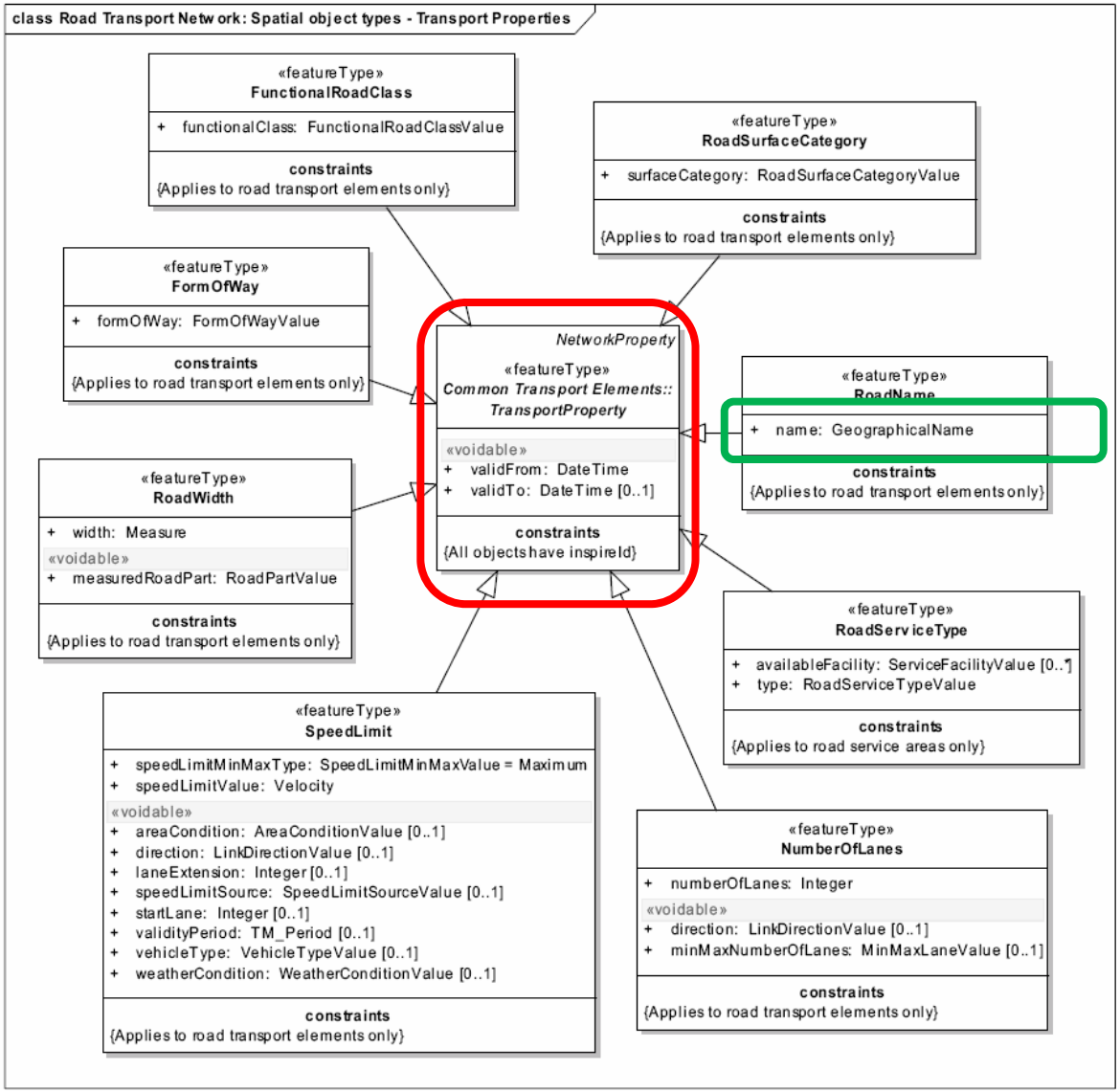
IGN

Design of matching tables: TN road example



The Road Transport inherits from Common Transport and from the Generic Network Model (GCM)

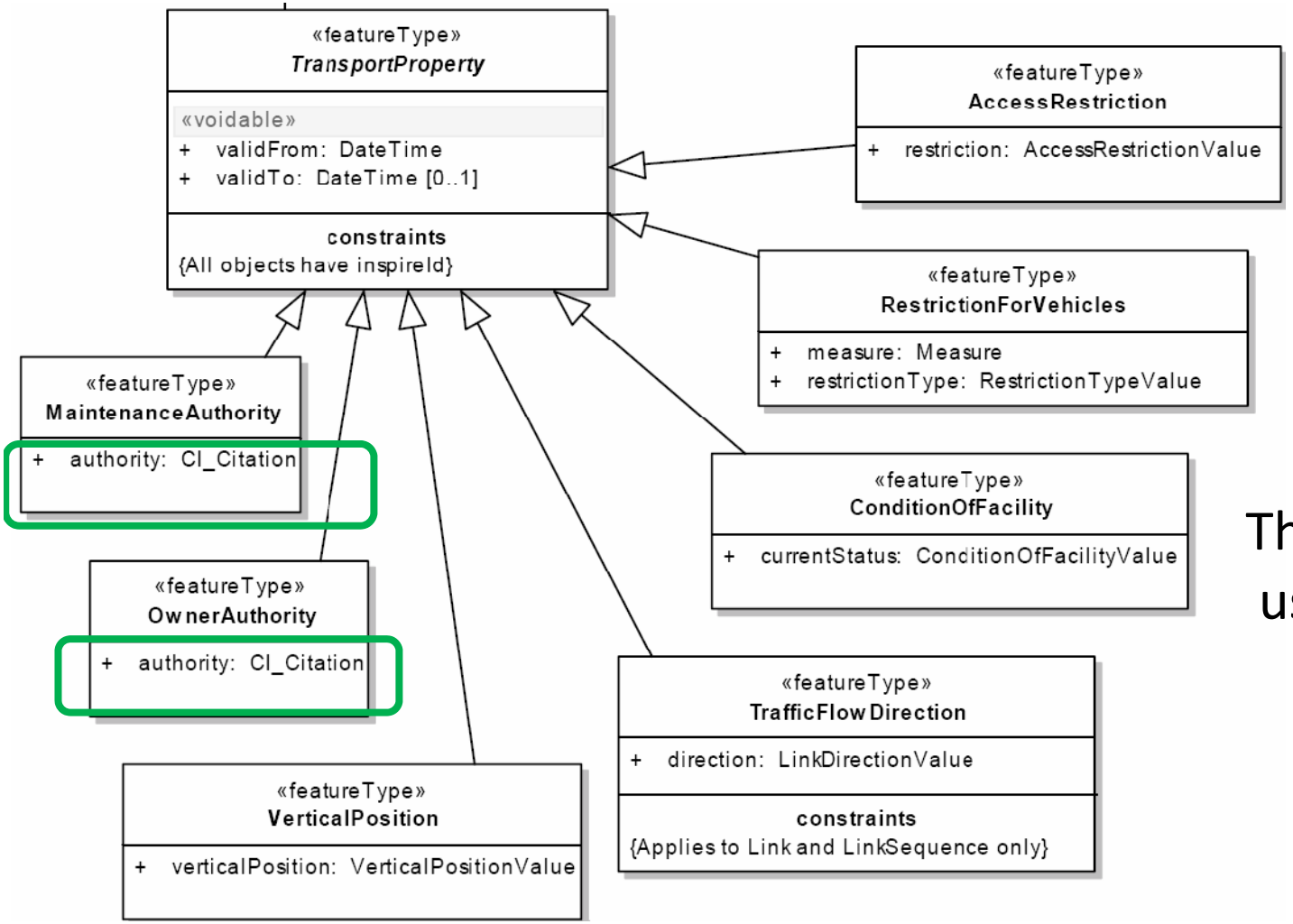
Design of matching tables: TN road example



Transport properties are considered as feature types

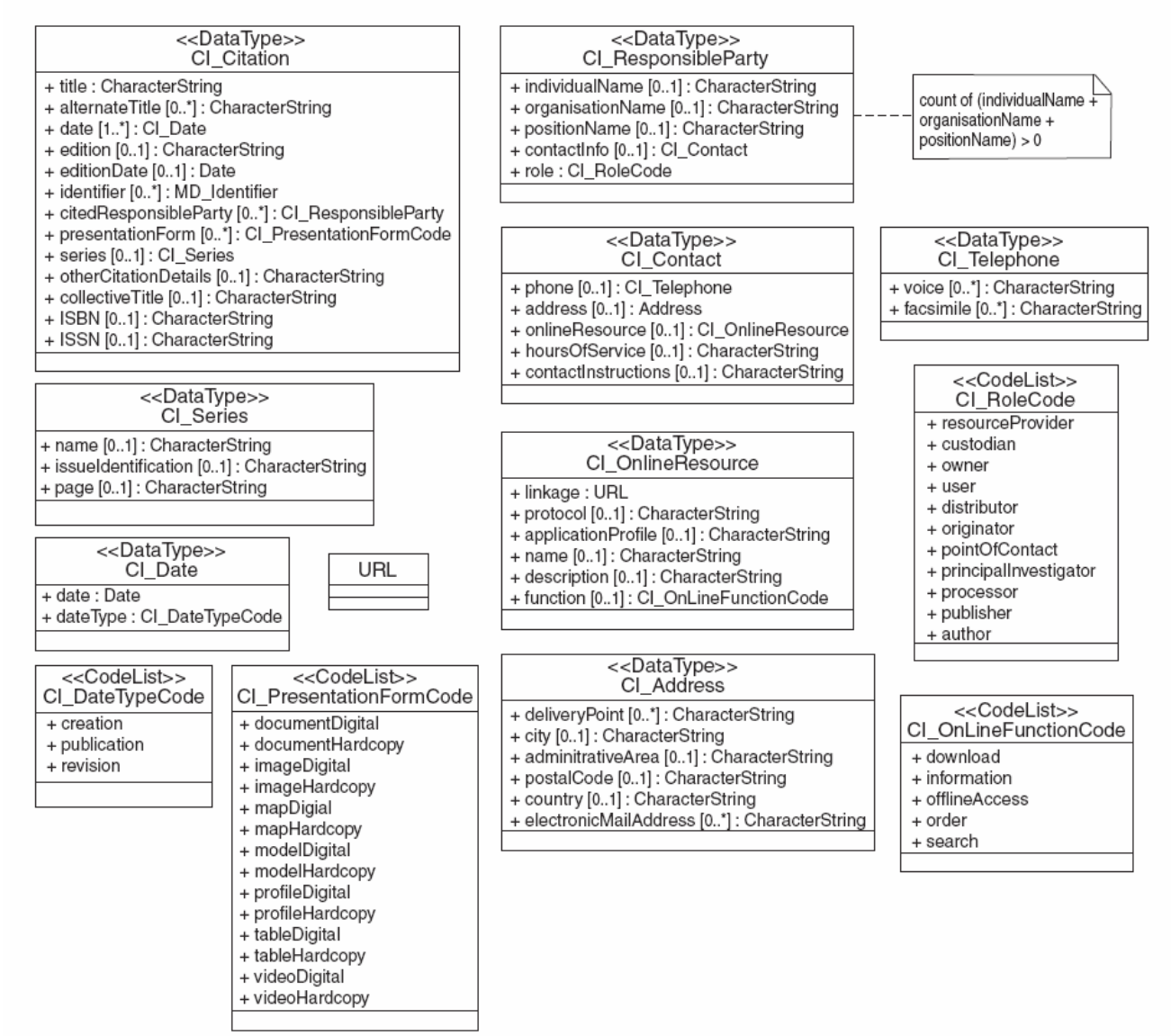
The data type GeographicalName is defined in theme GN

Design of matching tables: TN road example



The Road Transport uses the data type CI_Citation

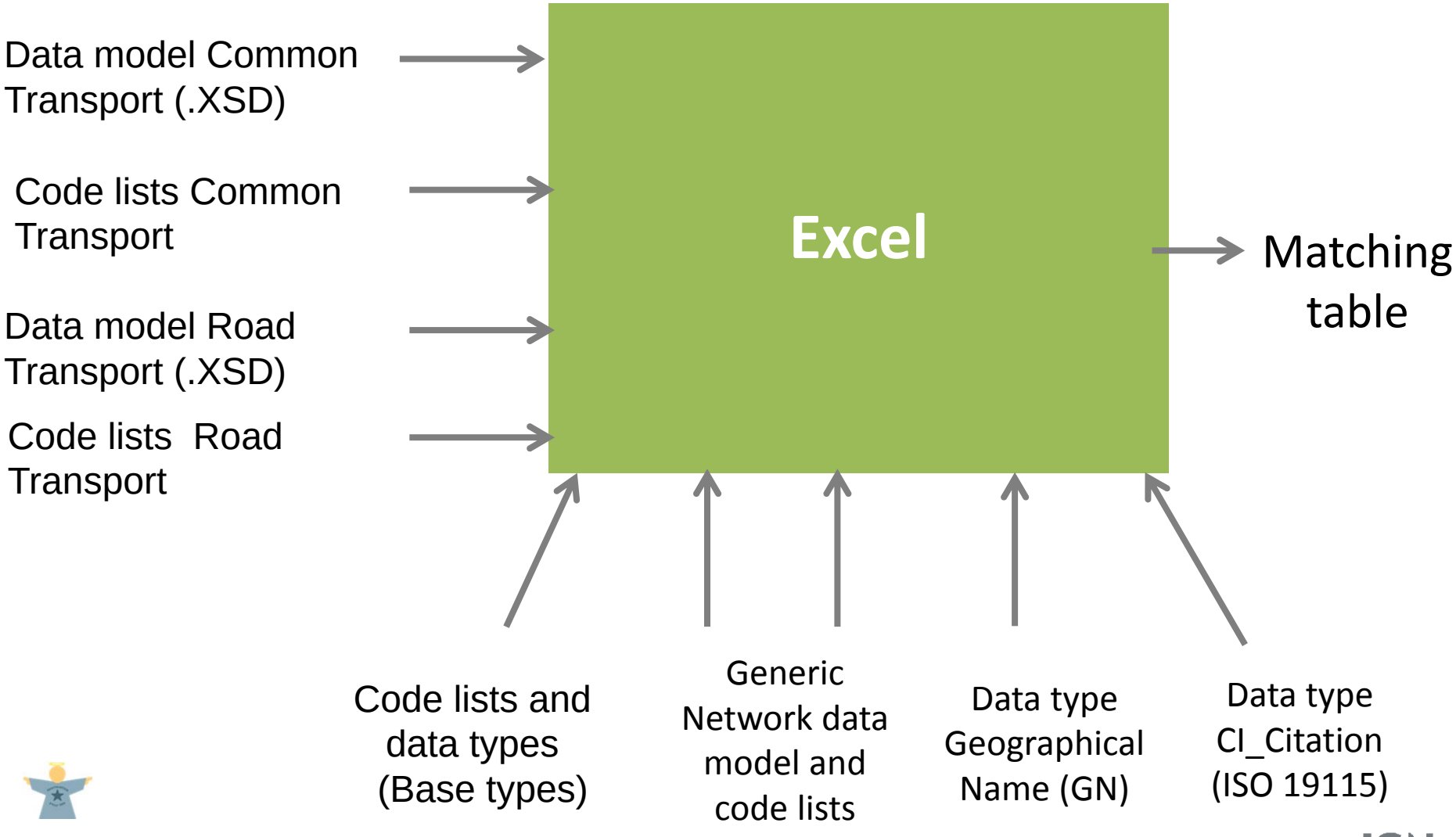
Design of matching tables: TN road example



The Road Transport uses the data type CI_Citation …. that is not simple at all and need to be imported in Excel table

⇒ 50 additional lines !!
(but only 3 have been filled by IGN)

Design of matching tables: TN road example



Design of matching tables: TN road example

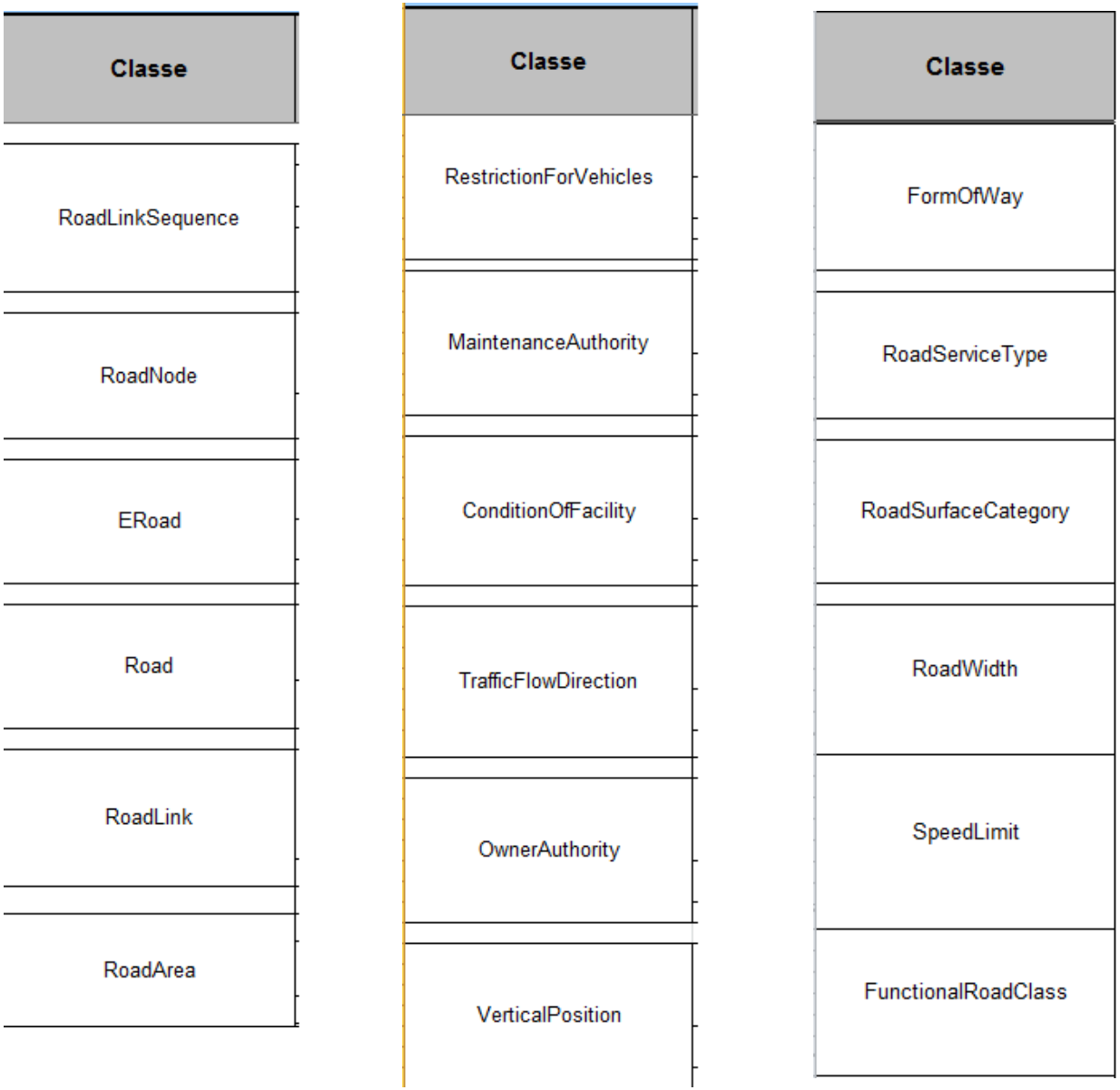
Roads		
Classe	**Classe parent**	**Attribut** **Lien** **Contrainte**
RoadLinkSequence	LinkSequence	link
	TransportLinkSequence	validFrom
		validTo
	TransportObject	geographicalName
	NetworkElement	beginLifespanVersion
		inspireId
		endLifespanVersion
		inNetwork
RoadNode	Node	geometry
		spokeEnd
		spokeStart
	TransportNode	validFrom
		validTo
	TransportObject	geographicalName
	NetworkElement	beginLifespanVersion
		inspireId
		endLifespanVersion
		inNetwork
		formOfRoadNode
ERoad	LinkSet	link
	TransportObject	geographicalName
	TransportLinkSet	validFrom
		validTo
		post
	NetworkElement	beginLifespanVersion
		inspireId
		endLifespanVersion
		inNetwork
		europeanRouteNumber

Instanciable feature types

The inheritance has to be considered in the matching table

Common parent feature type => often common transformation rules (copy/paste)

Design of matching tables: TN road example



The “real” spatial objects

The transport properties

The road properties

The matching table is ordered

INSTITUT NATIONAL
DE L'INFORMATION
GÉOGRAPHIQUE
ET FORESTIÈRE

USE OF MATCHING TABLES

Context of use

- **For each theme, there are meetings to fill the matching table**
- **The matching table is prepared in advance, before the meeting**
- **Contributors to the meeting:**
  - INSPIRE experts
    - Dominique Laurent: organises and chairs meeting, makes minutes
    - Marie Lambois : involved in test of annex I themes, prepares matching table
  - Source data expert
  - Transformation expert

Example 1 : scope of GN

- **IGN has named places in source data in the theme POI but many other features have also a name**
  - Administrative units
  - Hydrography (rivers, lakes, ….)
  - Transport (some roads, ….)

- **Decision was taken to provide for INSPIRE all features having a name**

Example 1 : scope of GN

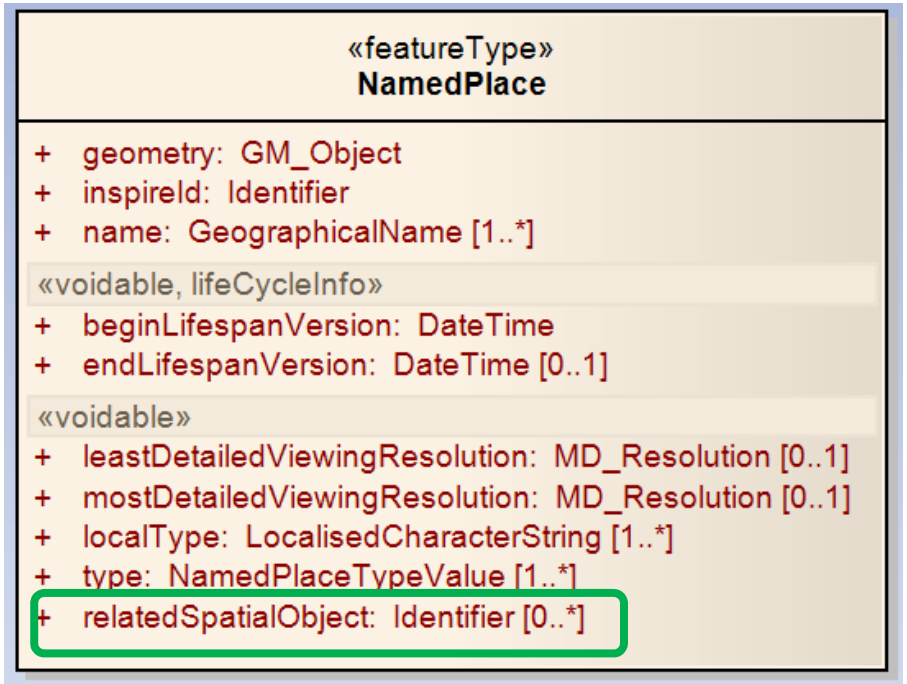
- **Consequence 1: duplication of tables**
  - Many source features have a name => to be included in theme GN
  - These source features come from different feature types, with different structures
  - The mapping with INSPIRE requires different rules
  - => 3 matching tables
    - POI
    - Administrative units
    - Networks

Example 1 : scope of GN

- **Consequence 2: identifier**
  - IGN has unique and persistent identifiers : CLEABS (absolute key)
    - => initial idea was to make following mapping:
      - Namespace: **FR_IGNF_BDUniGE (country + producer + product)**
      - localId : CLEABS
  - But the same source feature may be used for several INSPIRE themes
    - A river may be in HY (PhysicalWaters and HydroNetwork), in TN (WaterTransportNetwork) and in GN
    - A municipality may be in AU, in SU and in GN
    - ⇒We decided to add the name of INSPIRE application schema to the namespace
    - ⇒Namespace: **FR_IGNF_BDUNIME_GeographicalNames**
  - In addition, we realised that our identifiers were unique only by territory (metropol / oversea) => decision to add a suffix
    - Namespace: **FR_IGNF_BDUNIME_GeographicalNames_MET**

Example 1 : scope of GN

- **Consequence 3: relatedSpatialObject**



It is necessary to document this attribute

Having in theme GN all features with names is fine for some applications (geocoding)

But not for all: for mapping, this would imply duplicated features

Very difficult to document : need to know the INSPIRE identifier of the target object.

For theme GN, the relatedSpatialObject has been filled as « placeholder » : to be reviewed carefully once all themes have been transformed

=> Most of X-theme associations have not been filled

Example 1 : scope of GN

GeographicalNames				Source :			
Element	Type complexe	Attribut Lien Contrainte	Transformation	Classe	Attribut Lien Contraint	Type	Remarques
InspireId	Identifier	namespace	Case 1 : dataset name contains "FR" or dataset contains "20" « FR_IGNF_BDUniGE_GeographicalNames_MET » Case 2 : dataset name contains "GF" « FR_IGNF_BDUniGE_GeographicalNames_GUF » Case 3 : dataset name contains "GP" « FR_IGNF_BDUniGE_GeographicalNames_GUA » Case 4 : dataset name contains "RE" « FR_IGNF_BDUniGE_GeographicalNames_REU » Case 5 : dataset name contains "MQ" « FR_IGNF_BDUniGE_GeographicalNames_MTQ » Case 5 : dataset name contains "YT" « FR_IGNF_BDUniGE_GeographicalNames_MYT » Case 5 : dataset name contains "PM" « FR_IGNF_BDUniGE_GeographicalNames_SPM »	(CDENOMME, Voie ferrée nommée ROUTENUM) where GRAPHIEP is not null .			
		versionId					
relatedSpatialObject	Identifier	localId	=	(CDENOMME, Voie ferrée nommée ROUTENUM) where GRAPHIEP is not null .	Cleabs		
		namespace	si classe Route numérotée ou nommée, alors "FR_IGNF_BDUNIGE_RoadTransportNetwork" si classe Voie ferrée nommée, alors"FR_IGNF_BDUNIGE_RailwayTransportNetwork" si classe Cours d'eau nommé, alors "FR_IGNF_BDUNIGE_HydroNetwork"				
		versionId					

The data type « Identifier » has been duplicated because it has to be used twice (with different matching rules)

Example 2 : properties of TN

- **In source data, we have TRONROUT (RoadLink)**
  - with unique identifier: CLEABS
  - with direct attributes (that have no identifier) : number of lanes, functional road class, ….
- **How to provide identifiers to these properties**
  - same namespace as other road TN features
  - localId = property name + CLEABS (road link)
  - => the property is mainly defined by its location

TN Roads						
Classe		Type complexe	Attribut Lien	Transformation	Classe	Attribut Lien
FunctionalRoadClass	inspireId	Identifier	localId	='FunctionalRoadClass_'+CLEABS	TRONROUTwhere (NATURE <> 'Bac auto and NATURE <> 'Bac piéton') and IMPORTAN <> 'Sans objet'	CLEABS
			namespace	Case 1 : dataset name contains 'FR' or dataset name contains '20' 'FR_IGNF_BDUniGE_RoadTransportNetwork_MET'		
			versionId			
NumberOfLanes	inspireId	Identifier	localId	=NumberOfLanes_'+CLEABS	TRONROUTwhere (NATURE <> 'Bac auto and NATURE <> 'Bac piéton') and (NB_VOIES<> '0' and NB_VOIES is not null)	CLEABS
			namespace	Case 1 : dataset name contains 'FR' or dataset name contains '20' 'FR_IGNF_BDUniGE_RoadTransportNetwork_MET'		
			versionId			

Example 2 : properties of TN

- **How to reference these properties, i.e. to make the association to the related RoadLink**
  - concatenation of the **RoadLink inspireId** components
  - namespace + '_'+ localId

TN Roads						
Classe		Type complexe	Attribut Lien	Transformation	Classe	Attribut Lien
VerticalPosition, ConditionOfFacility, TrafficFlowDirection, FormOfWay, RoadSurfaceCategory, RoadWidth, FunctionalRoadClass, NumberOfLanes, AccessRestriction	networkRef	NetworkReference	element	Case 1 : dataset name contains 'FR' or dataset name contains '26' 'FR_IGNF_BDUniGE_RoadTransportNetwork_MET_'+ CLEABS Case 2 : dataset name contains 'GF' 'FR_IGNF_BDUniGE_RoadTransportNetwork_GUF_'+ CLEABS' Case 3 : dataset name contains 'GP' 'FR_IGNF_BDUniGE_RoadTransportNetwork_GUA_'+ CLEABS' Case 4 : dataset name contains 'RE' 'FR_IGNF_BDUniGE_RoadTransportNetwork_REU_'+ CLEABS' Case 5 : dataset name contains 'MQ' 'FR_IGNF_BDUniGE_RoadTransportNetwork_MTQ_'+ CLEABS' Case 6 : dataset name contains 'YT' 'FR_IGNF_BDUniGE_RoadTransportNetwork_MYT_'+ CLEABS' Case 7 : dataset name contains 'PM' 'FR_IGNF_BDUniGE_RoadTransportNetwork_SPM_'+ CLEABS'	TRONROUT where (NATURE <> 'Bac auto and NATURE <> 'Bac piéton')	CLEABS

As the rule is the same for all these properties, there is some factorisation

Example 2 : properties of TN

- **The same property may apply to different feature types**

Roads				BD-L-TC
Classe	**Classe parent**	**Attribut Lien Contrainte**	**Transformation**	**Classe**
		authority	see type complexe	
ConditionOfFacility	NetworkProperty	networkRef	see type complexe	TRONROUT where ETATOBJ is not null and (NATURE <> 'Bac auto' and NATURE <> 'Bac piéton')
		inspireId	see type complexe	
		beginLifespanVersion	voidValueReason = unpopulated	
		endLifespanVersion		
	TransportProperty	validFrom	voidValueReason = unknown	
		validTo	voidValueReason = unknown	
		currentStatus	= 'underConstruction'	
ConditionOfFacility	NetworkProperty	networkRef	see type complexe	TRANSPOR where NATURE = 'Aire de service' or 'Aire de repos' or 'Péage' or 'Carrefour' or 'Parking' or 'Rond-point' or 'Echangeur'
		inspireId	see type complexe	
		beginLifespanVersion	voidValueReason = unpopulated	
		endLifespanVersion		
	TransportProperty	validFrom	voidValueReason = unknown	
		validTo	voidValueReason = unknown	
		currentStatus	= functional'	
ConditionOfFacility	NetworkProperty	networkRef	see type complexe	SURFROUT where NATURE = 'Parking' or 'Péage'
		inspireId	see type complexe	
		beginLifespanVersion	voidValueReason = unpopulated	
		endLifespanVersion		
	TransportProperty	validFrom	voidValueReason = unknown	
		validTo	voidValueReason = unknown	
		currentStatus	= functional'	

RoadLink

RoadNode

Road Service Area

INSTITUT NATIONAL
DE L'INFORMATION
GÉOGRAPHIQUE
ET FORESTIÈRE

CONCLUSIONS

Conclusions

- **The matching table : a key document**
  - starting point for transformation process
    - to document the mapping decisions
    - to define the mapping on computer
  - starting point for further activities
    - Data validation
    - Data documentation
    - Evolution of transformation rules

- **To be accompanied by meeting minutes**
  - Rationale for choice when discussions (conflicts, options, ….)
  - Imperfect matchings
    - To be documented, reviewed,…
    - Potential evolutions of our products

Conclusions

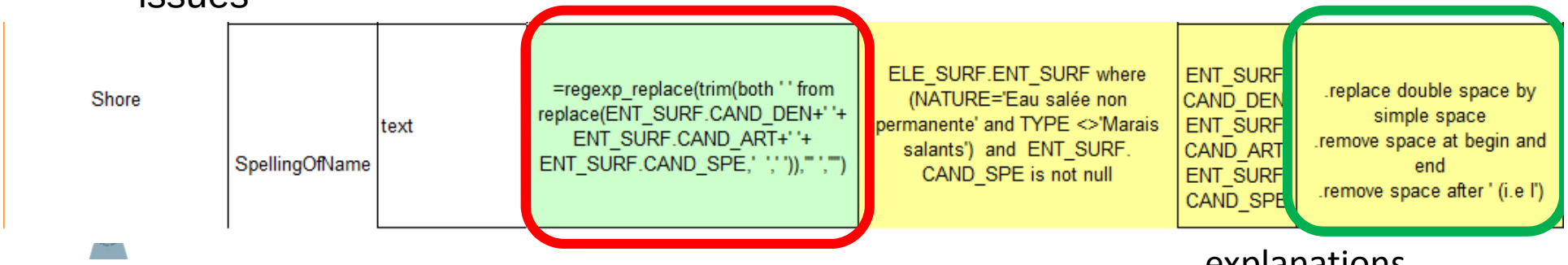
- **What is not addressed by the matching tables**
  - http URI identifiers
    - We expect more guidelines (MIG)
  - associations between themes
    - To be done once all themes transformed
  - Some common issues to all themes
    - How to deal with GML identifiers
    - How to deal with localisedCharacterString

To be done later

Conclusions

- **Schema transformation rules must be defined at “atomic” level:**
  - On feature types
  - On basic attributes (geometry, number, character string, boolean, date, …)
  - Even on values of attributes (enumerations / code lists)

- **Make the transformation rules as clear as possible**
  - by duplicating the feature type /data type lines if transformation rules are different
  - by using as much as possible a standard language (e.g. SQL)
  - by commenting the transformations that may raise understandability issues



explanations

Complex transformation (Postgre)

Conclusions

- **Filling matching tables is a (high level) sport**

outlet	gml:ReferenceType	FR_IGNF_BDUniME_HydroPhysicalWaters_MET_' + select TRON_HYD.BDC_ID from (select TRON_HYD.BDC_ID,TRON_HYD.GID from NEUD_ZON,NEUD_HYD,NEUD_EXT,GRAPH_HYD,TRON_HYD where (NEUD_ZON.NEUD_EXT.NOEUD=NEUD_HYD.BDC_ID and NEUD_HYD.GID=GRAPH_HYD.KEY3 and GRAPH_HYD.KEY1=TRON_HYD.GID)) as TRONCONS where (TRONCONS.CDO_TRON.COURDEAU.BDC_ID=DRAIN_Z.DRAIN.COURDEAU)

- Don't forget any attribute
  - Make consistent decisions
- Be resistant
  - Some themes may require several meetings
- Be clever and flexible
  - Some really complex transformations
- Be proud of yourself !

Our masterpiece: a" beautiful" transformation implying many source tables, semantic and topologic relations

INSTITUT NATIONAL
DE L'INFORMATION
GÉOGRAPHIQUE
ET FORESTIÈRE

Thanks for your attention