
DELIVERABLE

Project Acronym: E.L.F.

Grant Agreement number: CIP 325140

Project Title: The European Location Framework

D2.3.4: Transformation

Revision: final

Authors:

Anja Hopfstock (BKG)

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	x
C	Confidential, only for members of the consortium and the Commission Services	

Revision History and Statement of Originality

Revision	Date	Author/Editor	Organisation	Description
0.1	07/11/2014	Anja Hopfstock	BKG	Baseline
0.2	19/11/2014	Anja Hopfstock	BKG	Comments and amendments by Dominique Laurent (IGNF)
1.0	30/01/2015	Anja Hopfstock	BKG	Final editorial changes

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Transformation

References

Ref.	Title/Version/Publication Date/Author
[1]	ELF D2.3 Data Maintenance and Processing http://www.elfproject.eu/documentation/specification/data-maintenance-and-processing
[2]	ESDIN D6.3 Specification of the transformation between EBM/ERM/EGM and ExM data / v1.1 / 10.07.2010 / http://www.esdin.eu/sites/esdin.eu/files/D6.3%20Transformation%20Specifications%20Small%20and%20Midscale%20v1.0_0.pdf
[3]	ESDIN D10.2 Framework for specifying transformation rules / v1.0 / 24.02.2011 / http://www.esdin.eu/sites/esdin.eu/files/ESDIN%20D10_2%20Framework%20for%20Trans%20Rules%20v1.0.pdf
[4]	ELF Glossary http://www.elfproject.eu/documentation/glossary
[5]	ELF Specifications http://www.elfproject.eu/documentation/specification
[6]	ELF Schema Repository http://www.locationframework.eu/schemas/
[7]	ELF D4.1 Transformation Tools / v1.1 / 11.04.2014 https://service.projectplace.com/pp/pp.cgi/r1030046894
[8]	ShapeChange http://shapechange.net/
[9]	INSPIRE data specifications http://inspire.ec.europa.eu/index.cfm/pageid/2/list/datamodels
[10]	ELF WP2 - Excel Matching Table Guidelines v1.3 / 20.12.2013 http://www.elfproject.eu/sites/default/files/ELF_MatchingTableGuidelines_v1.3.pdf
[11]	INSPIRE KEN workshop about Transformation of Transport Network and Hydrography http://www.eurogeographics.org/content/workshop-about-transformation-themes-transport-network-and-hydrography-oct-2014

Figures

Figure 1 – Transformation in the ELF framework	6
Figure 2 - ELF schema development and transformation.....	9
Figure 3 – The context of schema mapping [3]	10
Figure 4 - Example of transformation specification.....	Feil! Bokmerke er ikke definert.
Figure 5 - Mapping of attribute values	13
Figure 6 – Transformation process	11

Table of content

1	Introduction.....	5
1.1	Purpose.....	5
1.2	Scope	5
1.3	Terminology.....	5
2	Transformation in the ELF framework.....	6
2.1	ELF organizational framework.....	6
2.2	ELF level of detail.....	7
3	Transformation concept.....	8
3.1	Schema matching	8
3.2	Schema mapping	10
3.3	Transformation process.....	11
4	Data Transformation Scenarios.....	12
4.1	Transformation of EG products	12
4.2	Transformation of national data	14
4.2.1	IGNF experience	14
5	Summary and conclusions.....	15
	Annex A.....	Feil! Bokmerke er ikke definert.

I Introduction

I.1 Purpose

The primary purpose of this document is to specify the details and requirements of the transformation process within the context of the ELF project.

This document neither specifies the location of transformation within the overall ELF workflow, nor the detailed implementation as a geo-tool.

I.2 Scope

For the purpose of this document, the “system” refers to each of the ELF WP4 transformation tools to be provided by the tools providers.

I.3 Terminology

The ELF Glossary [5] defines basic terms relevant for the ELF project. Beside this generic glossary, the following terms are used in this document:

Acronym	Term	Definition
-	Mapping tables	EXCEL mapping tables provided in XML-format can be used to set up matching tables as well as transformation tables (e.g. for ELF WP3 and WP4 Transformation) between any input and the ELF data schema.
LoD	Level of Detail	
EBM	EuroBoundaryMap	http://www.eurogeographics.org/products-and-services/euroboundarymap
ERM	EuroRegionalMap	http://www.eurogeographics.org/products-and-services/euroregionalmap
EGM	EuroGlobalMap	http://www.eurogeographics.org/products-and-services/euroglobalmap
ESDIN	European Spatial Data Infrastructure Network	http://www.esdin.eu/
NMCA	National Mapping and Cadastral Authority	
EG	EuroGeographics	www.eurogeographics.org

2 Transformation in the ELF framework

2.1 ELF organizational framework

Transformation is one component of the organisational framework needed for the maintenance of the European Location Framework. The overall data supply chain, from national data to the ELF platform, is described in the deliverable [D2.3 Data Maintenance and Processing specification](#) [1].

Transformation is the process where any source data is transformed to the ELF data specifications, see [ELF Data Specifications](#) [7].

The transformation process is implemented as ELF geo-tool, which allows to transform and publish datasets against the ELF data specifications, see deliverable [D4.1 Transformation tools](#) [8]. However it should be noted that NMCAs can implement the transformation process differently not utilizing the ELF Transformation tool.

The following picture portrays the link between transformation and the other deliverables in the ELF framework.

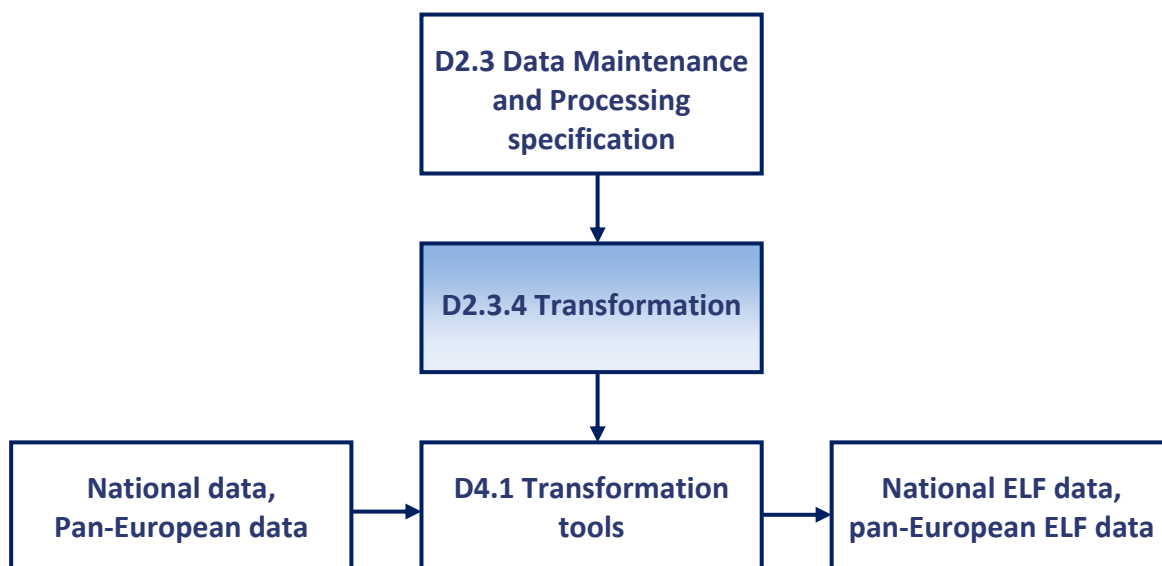


Figure 1 – Transformation in the ELF framework

2.2 ELF level of detail

ELF has defined different levels of detail (LoD), from Master LoD (LoD0, LoD1 and LoD2) via Regional LoD to Global LoD. In the project lifetime, the transformation is considered for these LoDs in the following way:

- **ELF Master:** Data for Master LoD0 and LoD1 are provided by the NMCAs and transformation needs to be defined by each NMCA for their national data sources. For Master LoD2 there are two options: either the data is transformed from national sources or ELF data for Master LoD2 is generated by generalisation. The ELF project is working on the generalisation from Master LoD1 to Master LoD2 (for more information see D2.3.5 Generalisation [1]).
- **ELF Regional:** This LoD is provided based on the current EG pan-European datasets EBM and ERM.
- **ELF Global:** This LoD is currently generalised from Regional LoD in the current production flowline of EG. Thus the transformation to ELF needs to be considered and specified based on the current EG dataset EGM. Once the generalisation process is defined within the ELF context transformation is out of scope for this LoD.

The analysis shows the necessity of transformation for different LoDs, at least for the time being. The transformation concept, described in next chapter, outlines the process independent of LoD.

3 ELF Transformation process

The ELF data specifications are developed as a set of profiles of the INSPIRE data specifications for different levels of detail comprising the parts of INSPIRE data specifications that are in the remit of the NMCAs and for which a pan-European and cross-border coverage is of interest. The data modelling adheres to the INSPIRE principles and follows a scale-independent approach. As a result, ELF provides a single data model for all levels of detail and one set of application schema files. This eases the work of service implementers and developers of the ELF geo-tools.

It requires data providers to transform their source data into the ELF/INSPIRE schema. In general this is a 3-step process as shown in Figure 2 below.

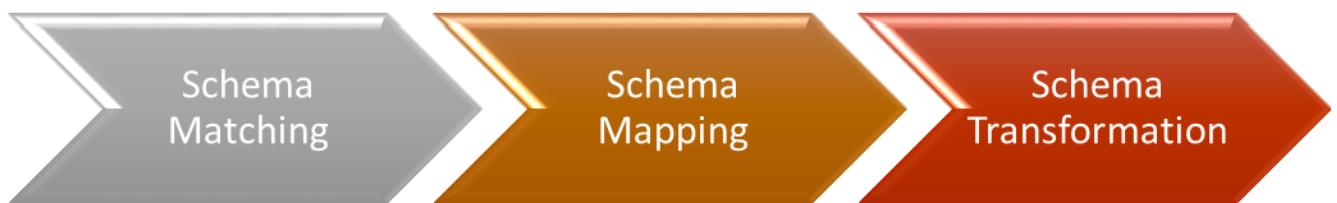


Figure 2 – Transformation Process

In the first step, the source data schema is analysed against the target schema in order to identify the matching concepts. Based on the analysis the schema mapping is used to refine the matching and to specify the transformation rules. The first two steps sometimes are not distinct steps but done in one effort. The result of the schema mapping is one of the controls for the actual schema transformation and is referred to as the Transformation Specification.

3.1 Schema matching

The objective of the schema matching is to identify the relationship between the source schema and the target schema. This task is usually carried out by data specification experts who analyse the schemas and identify the matching and missing concepts at all levels of the schemas (feature type, data type, attributes, associations).

The schema matching can be done differently: some experts prefer a visual approach comparing the data models but there are also tools to support this task. INSPIRE and ELF provide Excel mapping tables to support the matching and its documentation. These tables are automatically derived from the source data model. During this process the model is flattened and all types (feature and data) that can be instantiated are included in the table. Due to the flattening process no supertypes are actually included in the tables as types, but their attributes, which are inherited by the subtypes, are. Note that code list and enumeration classes are not yet included.

The matching tables have been split into 2 sections, each with several columns:

- 1) Left hand side which contains the target application schema data
- 2) Right hand side that is empty apart from column headings where concepts of the source schema can be added

In the ELF project schema matching has already been applied in the development of the ELF data specification. All ELF NMCA partners have been asked to perform a schema matching of their national data specifications at Master LoD against the INSPIRE data specifications. For Regional and Global

LoDs the ESDIN mapping tables [2] have been revised and amended considering the changes in the INSPIRE data specifications for Annex I, the new data specifications for INSPIRE Annex II-III as well as new users requirements from the EuroGeographics contract with EuroStat. The analysis of the matching tables revealed the matching concepts between INSPIRE and NMCA data holdings. This results in the common part of ELF/INSPIRE schemas. Additional content in ELF with respect to INSPIRE is coming from user requirements and requirements identified by ELF WP7. Figure 3 illustrates the approach. The ELF project developed guidelines how to perform the schema matching [10].

Project partners can re-use the results for the common part of ELF/INSPIRE in the preparation of the schema mapping.

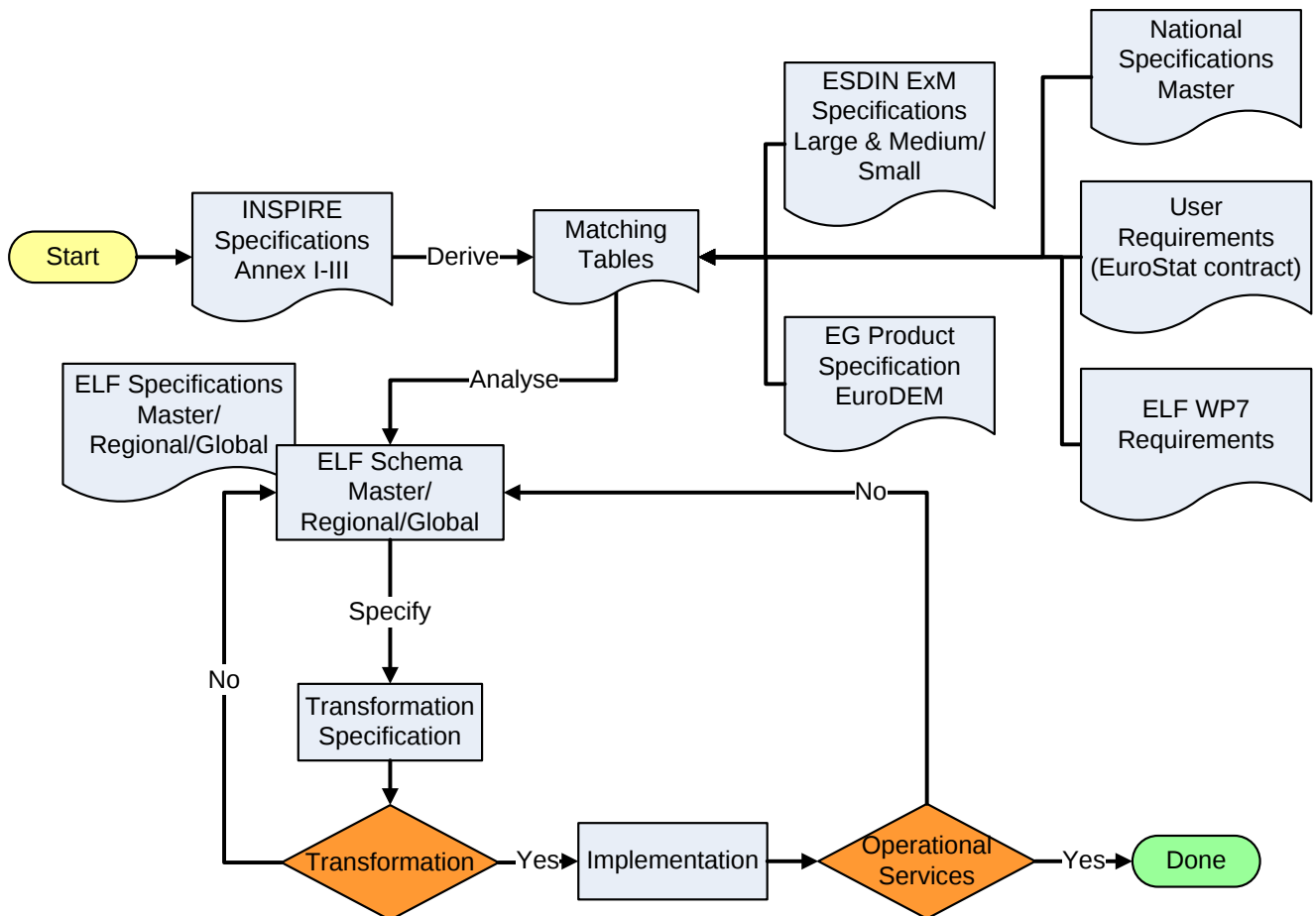


Figure 3 - ELF schema development and transformation

3.2 Schema mapping

The ESDIN framework for transformations [3] considers four transformation domains – model, coordinate, generalisation and language transformations. The ELF project has utilised the proposed ESDIN approach but is focusing on model transformation which mainly consists of derive, modify, classification and selection operations. The method used for model transformation is based on schema mapping. A schema mapping is a set of detailed rules (conditions of existence and operators) which modifies a spatial feature taken from a source dataset and to make it available in the target data schema. The resulting transformation specification is one of the controls for the transformation process (see section 3.3).

The documentation of the schema matching from the first task can be re-used for the schema mapping. As reported in the INSPIRE KEN workshop [11] the Excel mapping tables are utilised also for schema mapping but are adapted to the needs of the data provider.

In general, for each target application schema there is an Excel mapping table providing the description framework of the transformation specification. It contains the descriptions of the target and source application schemas as result of the schema matching. In order to document how to transform the source data into the target schema an additional column “Transformation Specification” can be added in between the target and source schema documentation. This column is used to document the mapping in terms of filtering conditions and predefined transformation operators with some explicit parameters.

These documents are a means for communication between domain and IT-experts. *“A domain expert knows the source specification and target specification and uses this knowledge to specify a mapping between the two. A database expert (or general IT staff) will have to implement this mapping specification (mapping description) in the database. The mapping description can either be executed on-the-fly (thus, no physical representation of the target data model) or as a batch, e.g., once a day where data is replicated from the source data model to the target data model”* ([3], pp.7-8).

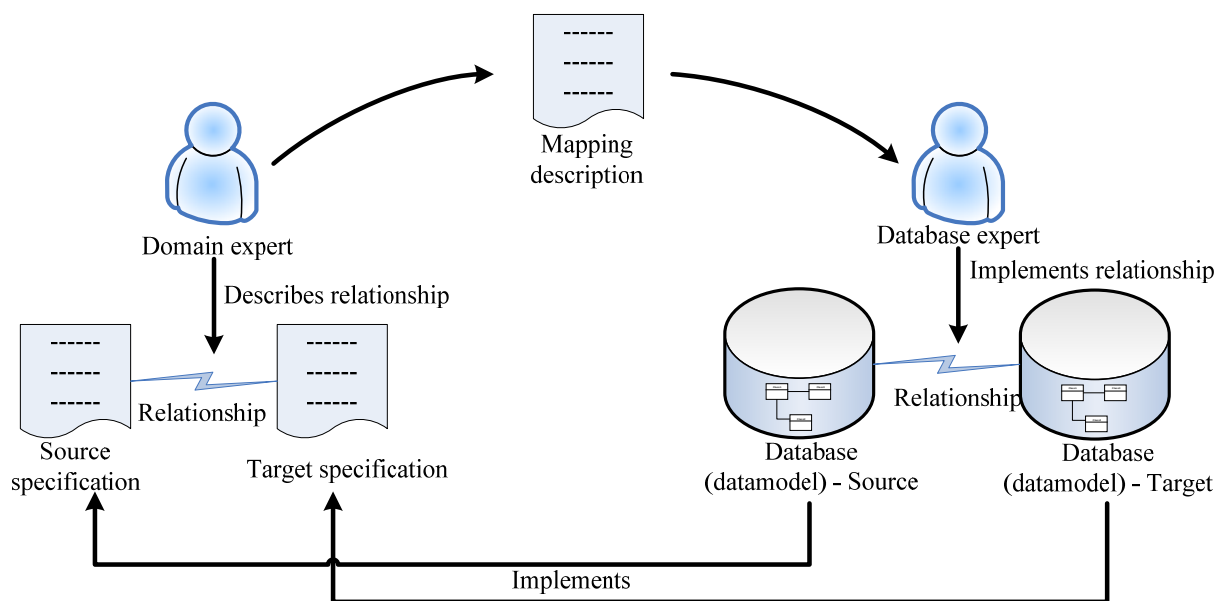


Figure 4 – The context of schema mapping [3]

3.3 Schema Transformation

The **transformation process** consists of 4 basic process steps:

- 3) Load data
- 4) Transform data
- 5) Validate schema
- 6) Return data and validation report

The **inputs** for this process are source data, either national data or EG pan-European data as described in section 2.2 above. The **outputs** are ELF data (Master, Regional, Global) and validation reports.

The **controls** for this process are transformation specifications and the ELF GML application schema files (XSD) for the main process step (2).

- Transformation specification as result of the schema mapping (see 3.2)
- ELF GML application schema made available on the ELF Schema Repository [6] define the target schema

The **resources** for transformation consist of the transformation tools themselves and the NMCA staff assigned to the transformation process.

The **output** of the transformation consists of the transformed data and a validation report. Some systems include a validation of the output against the target schema. Validation of target data itself is part of the ELF data quality process (for more information see D2.3.6 Data Quality [1]).

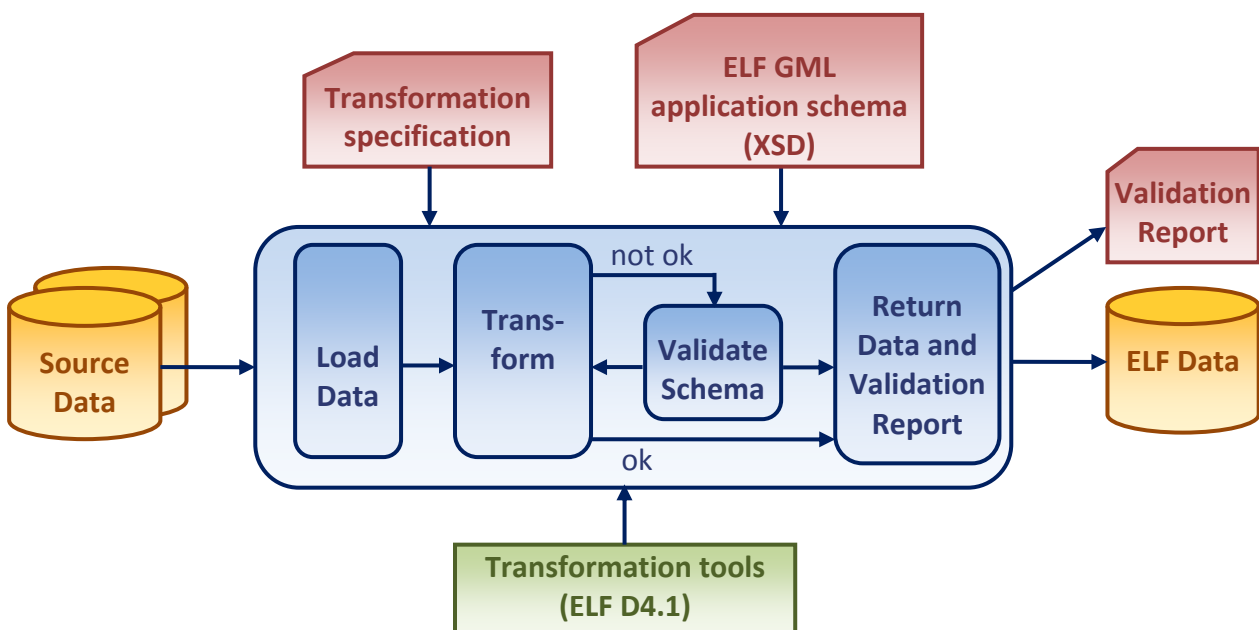


Figure 5 – Transformation process

4 Data Transformation Scenarios

4.1 Transformation of EG products

The existing EuroGeographics pan-European data sets are considered a key data component of the ELF authoritative reference geo-information. The envisaged ELF Regional and Global services deliver ELF compliant data that requires a transformation of the existing EG products into the ELF data specifications.

The ESDIN “Specification of the transformation between EBM/ERM/EGM and ExM data” [2] provided the starting point and have been revised according to the ELF data specifications [5].

The mapping tables are filled by aligning the target and source application schemas by feature type, attribute and attribute value and defining the required mapping expressions.

The following basic expressions and constructs are used for e.g. the mapping descriptions to specify the transformations between EuroGeographics products EBM/ERM/EGM data schemas and the ELF data schema:

- Basic expressions (for further explanations see [3], 4.1):
 - o Pathname, e.g. WatcrsL.HOC = ManMade
 - o Conditions, e.g. IF Lake / Pond THEN setValue('natural') IF Reservoir THEN setValue('manMade')
- Basic constructs (for further explanations see [3], 4.2):
 - o Comparison, e.g. WatcrsL.LOC <> 984
 - o String, e.g. Trim(ICC)
 - o Logical, e.g. select (LLE <> 0 OR LLE <> 984)
 - o Identifier, e.g. NewGuid
 - o Geometric functions, e.g. Length(WatcrsL.SHAPE)
 - o Miscellaneous:
 - Function <copy> indicates an attribute to be mapped 1:1.
 - Function <missing> indicates a feature class missing in the data set, e.g. “No matching feature type in EBM”
 - Function setValue() is a function for inserting a constant value, e.g. setValue("1:250000")
 - Function rename() is a function for putting a new name for a feature or an attribute in the target dataset.

An example is shown in Figure 6.

Type	Attribute	Association	Multiplicity	Voidable / Non-Voidable	Transformation Specification	Type	Documentation	Attribute	Association	Attribute / Association role / Enumerations	Values / Enumerations	Multiplicity	Voidable / Non-Voidable	Status	Remarks
AerodromeArea Supertypes: AerodromeAreaTransportAreaNetworkAreaNetworkElementTransportObject	beginLifespanVersion		1	voidable	copy or setValue(VoidValueReason = unpopulated)	GB005 Airfda GB005 AirfdaC		beginLifespanVersion							
	inspireId		0..1	voidable	copy or NewGUID			inspireD							
	endLifespanVersion		0..1	voidable	setValue(VoidValueReason = unpopulated)										
	inNetwork		1..*	voidable	setValue(VoidValueReason = unpopulated)										
	geometry		1		copy			SHAPE	Feature geometry	polygon					
	geographicalName		0..1	voidable	IF Airfda.IKO = Airfda.IKO THEN copy			Airfda.NAMIN1 / Airfda.NAMIN2	name				mandatory		
	validFrom		1	voidable	setValue(VoidValueReason = unpopulated)										
	validTo		0..1	voidable	setValue(VoidValueReason = unpopulated)										
	area		1		copy			ARA	area						
	aerodromeNode		1		set link using IKO										

Figure 6 - Example of transformation specification

A separate worksheet “Mapping tables” describes the mapping of code lists and enumerations in more detail. The INSPIRE/ELF code list and enumeration values are aligned with the corresponding source feature domain values. Missing information is indicated in red whereas additional information is highlighted in blue. A sample of mapping coded values is given in Figure 7.

Target feature type	Target feature attribute	Target feature value	Source feature type	Source feature attribute	Source feature value	Source value description
ConditionOfAirFacility	currentStatus		GB005 AirfdaP GB005 AirfdaC	EXS	Existence Category	
		decommissioned			<missing>	
		disused			Abandoned / disused	6
		functional			Operational	28
		projected			<missing>	
		underConstruction			Under construction	5
					Unknown	0
					Unpopulated	997

Figure 7 - Mapping of attribute values

BKG has started the data transformation using “GoPublisher” by Snowflake (see also ELF D4.1 Transformation Tools [7]) for EBM and part of ERM.

Table 1 – Status of transformation for EG products

Source	Target	Example
EBM	AU http://www.locationframework.eu/schemas/AdministrativeUnits/0.1/	http://www.locationframework.eu/schemas/AdministrativeUnits/0.1/example/
ERM	TN Air http://www.locationframework.eu/schemas/AirTransportNetwork/0.1/	http://www.locationframework.eu/schemas/AirTransportNetwork/0.1/examples/
	PS http://www.locationframework.eu/schemas/ProtectedSites/0.1/	http://www.locationframework.eu/schemas/ProtectedSites/0.1/examples/

Lessons learned

- Understanding of source data structure required
- Understanding of INSPIRE/ELF XML schema structure required
- Implementation of complex feature and attribute mappings possible but sometimes challenging
- Transformation itself very fast

Issues

- Handling of GML
- What is the best workflow?

4.2 Transformation of national data

4.2.1 IGNF experience

IGN France has begun to transform its Master data for ELF, using two different methods:

- By home-made solution based on a two-steps process:
 - o Data model transformation from source database to pseudo-INSPIRE database
 - o Format transformation to GML using WFS (GeoServer and then deegree)
- By requesting the help of Snowflake to perform the data transformation with its tool Go Publisher. In this case, the transformation is performed in one-step by populating the GML target structure from source data, using relevant transformation rules.

A priori, the first solution should be the definitive one but its development takes time; it is why IGN has also envisaged the second option.

Status

- Implementation started with
 - o Master level 0 : theme AD (transformed data on test areas)
 - o Master level 1 :
 - Themes AU ,GN, TN road and railway (transformed data on test areas)
 - Theme BU (matching table on-going)
 - o Master level 2 :
 - Theme HY PhysicalWaters and HydroNetwork (transformed data on test areas)
 - Theme TN Water (matching table ready)

Lessons learned

- Transformation may be performed only on basic concepts (attributes that are character strings, dates, Boolean, number or even values of attribute if codelists): this is why IGNF has adapted the ESDIN matching table by adding a sheet devoted to the data types
- The help of Snowflake was provided in the following way:
 - o IGNF provides source data and matching table
 - o Snowflake makes the transformation and sends back the transformed data in GML format
 - o Positive consequence was to push IGNF to fill the matching table in a very detailed and standardized way
 - o Negative point is that IGNF has not yet succeeded in using the GML files provided by Snowflake as input data for WFS: it seems that WFS can only use database as input (more investigation still required).

- IGNF has data that is generally in quite simple data models
 - o AD components and TN properties are simple attributes in source data whereas they are feature types in INSPIRE/ELF models
 - o In source data, the attributes of “aggregate” features, such as roads, upper level administrative units are carried by the simple objects that compose the aggregate ones; this modelling is considered as quite convenient for data delivery. In INSPIRE models, either these aggregate features have their own geometry (upper level administrative units) or are defined as sets of simple objects (roads, railway lines).
 - o In these cases, the transformation from source to target becomes rather complex. More especially, great care has to be taken in the transformation of inspire identifiers and of associations.

5 Summary and conclusions

The documentation of the schema matching and schema mapping is not only important for the transformation process within one organisation but moreover for the data harmonisation and the future maintenance of the ELF data specification. It documents the diverse national source data sets / data model and the chosen approach to transformation. Sharing the documentation between ELF data providers is recommended as it provides the opportunity to share best practice, compare the approach and to discuss any mapping issues.