Examples of Data Transformation

Authors: Giacomo Martirano, Stefania Morrone, Fabio Vinci (EPSILON ITALIA).
www.epsilon-italia.it

Introduction

This self-learning module provides an example of transformations of a source dataset into a dataset compliant to the technical requirements of the relevant Implementing Rules and Technical Guidelines of INSPIRE.

It shows, step by step, a schema transformation process, starting from the analysis of the source dataset and of its data model and the study of the applicable INSPIRE Data Specification.

The module shows the use of the mapping table as useful tool to document the mapping process between the elements of the source dataset and the INSPIRE data model elements and explains how to identify and solve some common mapping problems.

Through the use of a selected tool, the transformation process is practically explained, showing also the “live” validation of the mapping being performed against the relevant INSPIRE application schema. At the end, a demonstration is given of how to generate a harmonized GML dataset.

Learning outcomes:

After the module, the participant will be able to

- Identify and understand the source and target data models
- Fill in a mapping table
- Perform a data transformation from a non-harmonized source dataset into a harmonized dataset
- Export a harmonized dataset into a harmonized GML dataset file.

Intended Audience:

GIS and ICT professionals aiming to harmonize their datasets against INSPIRE Data Specifications.

Pre-requisites:

- Basic knowledge of INSPIRE Directive.
- Module “Procedures for Data and Metadata Harmonization”.

Referenced files:

1. **Com2011.shp**: a sample dataset of Italian municipalities.

2. **ID_table.csv**: table associating dataset of municipalities to related boundaries.

3. **AU_Mapping_Table.xls**: INSPIRE mapping table (extended and filled in).

4. **AdministrativeUnit.halez**: Hale project covered by present training module. The *.halez* file contains the complete project i.e. source and target schemas, source data and alignment.

5. **AdministrativeUnit.gml**: INSPIRE compliant gml dataset file, obtained by means of the data transformation described in this training module.

Referenced transformation tool:

- HUMBOLDT Alignment Editor (HALE) downloadable at **http://www.esdi-community.eu/projects/hale/files**

Summary

- Source Data Model
- Target Data Model
- Mapping table: instruction for use
- Common mapping problems
- Open source Hale transformation tool
- Data Transformation
- Data Validation
- Creation and validation of a harmonized GML dataset.

Source Data Model

Source Data Model

Here follows a list of the attributes of the dataset of Italian municipalities (com2011.shp) which represents our source data model.

Attributi dei Comuni: *shapefile* poligonale "com2011"

Campo	*Definizione*
COD_REG	Codice ISTAT della Regione
COD_PRO	Codice ISTAT della Provincia
PRO_COM	COD_PRO e COD_COM concatenati
NOME_COM	Denominazione del Comune
NOME_TED	Denominazione, in lingua tedesca, dei Comuni della Provincia autonoma di Bolzano/*Bozen*

Summary

- Introduction
- Source Data Model
- Target Data Model
- Mapping table: instruction for use
- Common mapping problems
- Open source Hale transformation tool
- Data Transformation
- Data Validation
- Creation and validation of a harmonized GML dataset.

Target Data Model

Among the Data Themes identified by the INSPIRE Directive, the Administrative Units one is the most suitable for the transformation of a dataset of municipalities.

Our Target Data Model is described in the INSPIRE Data Specification for the spatial data theme Administrative Units and can be downloaded from the INSPIRE website.

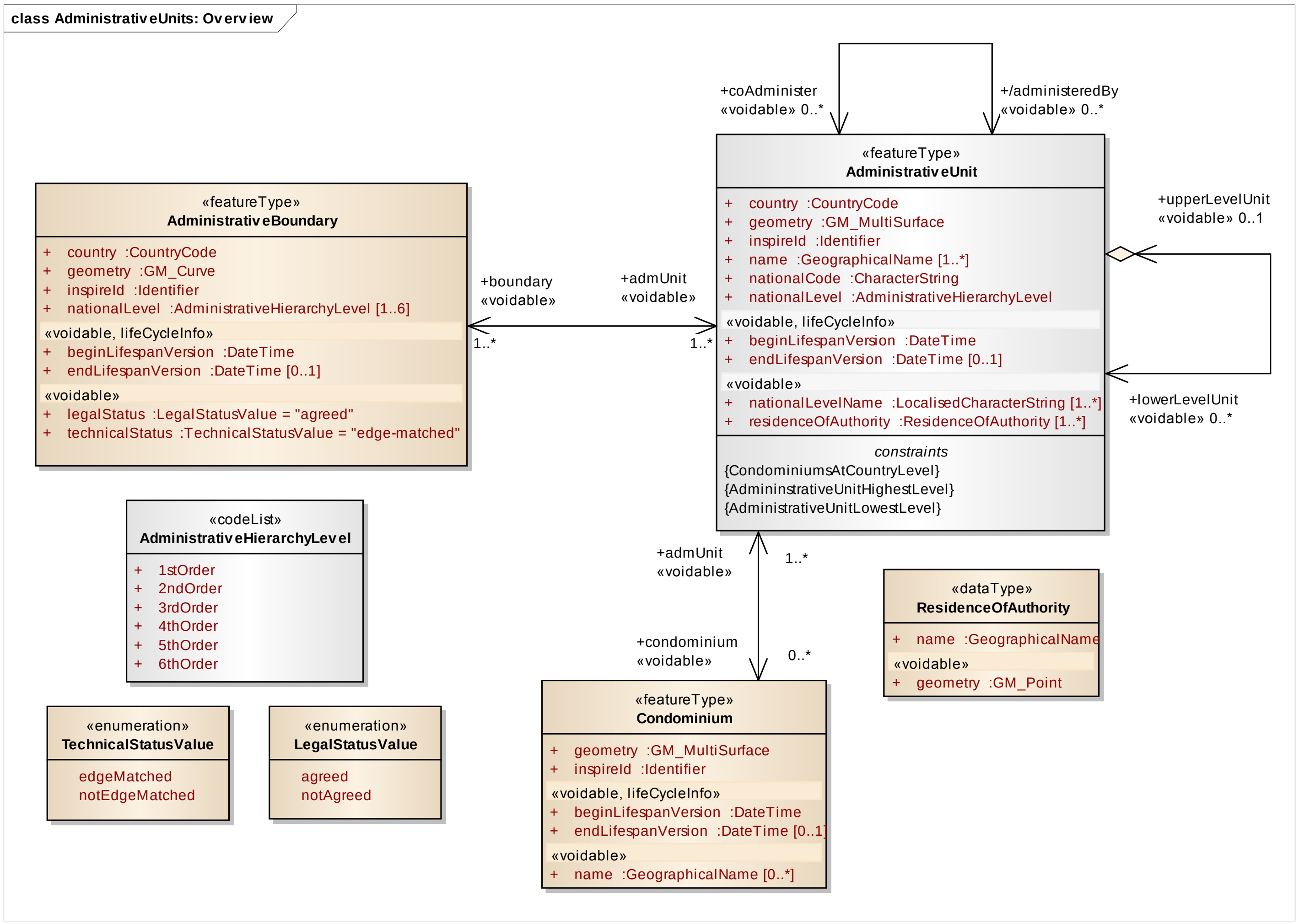
INSPIRE
Infrastructure for Spatial Information in Europe

D2.8.I.4 INSPIRE Data Specification on Administrative units – Guidelines

Title	D2.8.I.4 INSPIRE Data Specification on *Administrative units* – Guidelines
Creator	INSPIRE Thematic Working Group Administrative units
Date	2010-04-26
Subject	INSPIRE Data Specification for the spatial data theme *Administrative units*
Publisher	INSPIRE Thematic Working Group Administrative units
Type	Text
Description	This document describes the INSPIRE Data Specification for the theme *Administrative units*
Contributor	Members of the INSPIRE Thematic Working Group Administrative units
Format	Portable Document Format (pdf)
Source	
Rights	public
Identifier	INSPIRE_DataSpecification_AU_v3.0.1.pdf
Language	En
Relation	Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)
Coverage	Project duration

Target Data Model

At the INSPIRE website the Data model is available in different formats: UML, HTML, Mapping table and GML Application schemas (XSD files).



Summary

- Introduction
- Source Data Model
- Target Data Model
- Mapping table: instruction for use
- Common mapping problems
- Open source Hale transformation tool
- Data Transformation
- Data Validation
- Creation and validation of a harmonized GML dataset.

Mapping table: instructions for use

Attributes of the com2011 shapefile shall now be mapped into corresponding attributes of the Administrative Unit Feature Type.
In this process the use of mapping tables available at the INSPIRE website could be very handy.

In the case of our exercise we need to download the AdministrativeUnits Mapping Table.xml file, which contains a list of the feature types, data types and code lists associated to our target model and related attributes. Associations between feature types are described as well.

More detailed description of how to download mapping tables can be found in the module “Procedures for Data and Metadata Harmonization”.

Mapping table: instructions for use

Application Schema 'AdministrativeUnits' (version 3.0)								
Type	**Documentation**	**Attribute Association role Constraint**	**Attribute / Association role / Constraint**	**Values / Enumerations**	**Multiplicity**	**Voidable / Non-Voidable**		**Type**
AdministrativeBoundary	-- Name -- administrative boundary A line of demarcation between administrative units.							
		beginLifespanVersi	-- Name -- begin	DateTime	1	voidable		
		country	-- Name -- country	CountryCode* BE*	1			
		endLifespanVersio	-- Name -- end lifespan	DateTime	0..1	voidable		
		geometry	-- Name -- geometry	GM_Curve	1			
		inspireId	-- Name -- inspire id	Identifier	1			
		legalStatus	-- Name -- legal status	LegalStatusValue*	1	voidable		
		nationalLevel	-- Name -- national	AdministrativeHierarchyL	1..6			
		technicalStatus	-- Name -- technical	TechnicalStatusValue	1	voidable		
		admUnit	-- Name -- adm unit	AdministrativeUnit	1..*	voidable		
AdministrativeUnit	-- Name -- administrative unit Unit of administration where a Member State has and/or exercises jurisdictional rights, for local, regional and national governance.							
		beginLifespanVersi	-- Name -- begin	DateTime	1	voidable		
		country	-- Name -- country	CountryCode* BE*	1			
		endLifespanVersio	-- Name -- end lifespan	DateTime	0..1	voidable		
		geometry	-- Name -- geometry	GM_MultiSurface	1			
		inspireId	-- Name -- inspire id	Identifier	1			
		name	-- Name -- name	GeographicalName	1..*			
		nationalCode	-- Name -- national	CharacterString	1			
		nationalLevel	-- Name -- national	AdministrativeHierarchyL	1			
		nationalLevelName	-- Name -- national level	LocalisedCharacterString	1..*	voidable		
		residenceOfAuthori	-- Name -- residence of	ResidenceOfAuthority	1..*	voidable		
		condominium	-- Name --	Condominium	0..*	voidable		
		boundary	-- Name -- boundary	AdministrativeBoundary	1..*	voidable		
		lowerLevelUnit	-- Name -- lower level	AdministrativeUnit	0..*	voidable		

Mapping table: instructions for use

Application Schema 'AdministrativeUnits' (version 3.0)								
Type	**Documentation**	**Attribute Association role Constraint**	**Attribute / Association role / Constraint**	**Values / Enumerations**	**Multiplicity**	**Voidable / Non-Voidable**		**Type**
AdministrativeBoundary	-- Name -- administrative boundary A line of demarcation between administrative units.							
		beginLifespanVersi	-- Name -- begin	DateTime	1	voidable		
		country	-- Name -- country	CountryCode* BE*	1			
		endLifespanVersio	-- Name -- end lifespan	DateTime	0..1	voidable		
		geometry	-- Name -- geometry	GM_Curve	1			
		inspireId	-- Name -- inspire id	Identifier	1			
		legalStatus	-- Name -- legal status	LegalStatusValue*	1	voidable		
		nationalLevel	-- Name -- national	AdministrativeHierarchyL	1..6			
		technicalStatus	-- Name -- technical	TechnicalStatusValue	1	voidable		
		admUnit	-- Name -- adm unit	AdministrativeUnit	1..*	voidable		
AdministrativeUnit	-- Name -- administrative unit Unit of administration where a Member State has and/or exercises jurisdictional rights, for local, regional and national governance.							
		beginLifespanVersi	-- Name -- begin	DateTime	1	voidable		
		country	-- Name -- country	CountryCode* BE*	1			
		endLifespanVersio	-- Name -- end lifespan	DateTime	0..1	voidable		
		geometry	-- Name -- geometry	GM_MultiSurface	1			
		inspireId	-- Name -- inspire id	Identifier	1			
		name	-- Name -- name	GeographicalName	1..			
		nationalCode	-- Name -- national	CharacterString	1			
		nationalLevel	-- Name -- national	AdministrativeHierarchyL	1			
		nationalLevelName	-- Name -- national level	LocalisedCharacterString	1..*	voidable		
		residenceOfAuthori	-- Name -- residence of	ResidenceOfAuthority	1..*	voidable		
		condominium	-- Name --	Condominium	0..*	voidable		
		boundary	-- Name -- boundary	AdministrativeBoundary	1..*	voidable		
		lowerLevelUnit	-- Name -- lower level	AdministrativeUnit	0..*	voidable		

Mapping table: instructions for use

Application Schema 'AdministrativeUnits' (version 3.0)												Sou	
Feature Type / Data Type	Documentation	Attribute Association role Constraint	Attribute / Association role / Constraint documentation	Data Type / Values / Code List - Enumeration	Multiplicity	Voidable / Non-Voidable	Data Type Attribute	Data Type Attribute documentation	Data Type / Values / Code Lists / Enumerations	Multiplicity	Voidable / Non-Voidable	"File name" or URL	Name of attribute
Administrative Unit	-- Name -- administrative unit Unit of administration where a Member State has and/or exercises jurisdictional rights, for local, regional and national governance.												
		country	-- Name -- country Two character	CountryCode* BE* BG* CZ* DK* DE*	1								
		beginLifespanVersion	-- Name -- begin lifespan version	DateTime	1	voidable							
		endLifespanVersion	-- Name -- end lifespan version	DateTime	0..1	voidable							
		geometry	-- Name -- geometry	GM_MultiSurface	1							com2011.shp	shape
		inspireId	-- Name -- inspireId External object identifier of the spatial object.NOTE An external object identifier is a unique object	Identifier	1		localId	A local identifier,	CharacterString	1		com2011.shp	pro_com
							namespace	Namespace uniquely	CharacterString	1			
							versionId	The identifier of the	CharacterString	0..1	voidable		
		name	-- Name -- name Official national	GeographicalName	1..*		See Geographical name data type						
		nationalCode	-- Name -- national code Thematic	CharacterString	1							com2011.shp	pro_com
		nationalLevel	-- Name -- national level Level in the national administrative hierarchy, at which the administrative unit is	AdministrativeHierarchyLevel* 1stOrder* 2ndOrder* 3rdOrder* 4thOrder* 5thOrder*	1								
		nationalLevelName	-- Name -- national level name Name	LocalisedCharacterString	1..*	voidable							
		residenceOfAuthority	-- Name -- residence of authority	ResidenceOfAuthority	1..*	voidable	See ResidenceOfAuthority data type						

Mapping table: instructions for use

Application Schema 'AdministrativeUnits' (version 3.0)											
Feature Type / Data Type	Documentation	Attribute Association role Constraint	Attribute / Association role / Constraint documentation	Data Type / Values / Code List - Enumeration	Multiplicity	Voidable / Non-Voidable	Data Type Attribute	Data Type Attribute documentation	Data Type / Values / Code Lists / Enumerations	Multiplicity	Voidable / Non-Voidable
Administrative Unit	-- Name -- administrative unit Unit of administration where a Member State has and/or exercises jurisdictional rights, for local, regional and national governance.										
		country	-- Name -- country Two character	CountryCode* BE* BG* CZ* DK* DE*	1						
		beginLifespanVersion	-- Name -- begin lifespan version	DateTime	1	voidable					
		endLifespanVersion	-- Name -- end lifespan version	DateTime	0..1	voidable					
		geometry	-- Name -- geometry Geometric	GM_MultiSurface	1						
		inspireId	-- Name -- inspireId External object identifier of the spatial object.NOTE An	Identifier	1		localId	A local identifier	CharacterString	1	
							namespace	Namespace uniquely	CharacterString	1	
							versionId	The identifier	CharacterString	0..1	voidable
		name	-- Name -- name Official national	GeographicalName	1..*		See Geographical name data type				
		nationalCode	-- Name -- national code Thematic	CharacterString	1						
		nationalLevel	-- Name -- national level Level in the national administrative hierarchy, at which the administrative unit is	AdministrativeHierarchyLevel* 1stOrder* 2ndOrder* 3rdOrder* 4thOrder* 5thOrder*	1						
		nationalLevelName	-- Name -- national level name Name	LocalisedCharacterString	1..*	voidable					
		residenceOfAuthority	-- Name -- residence of authority	ResidenceOfAuthority	1..*	voidable	See ResidenceOfAuthority data type				
		condominium	-- Name -- condominium	Condominium	0..*	voidable	See condominium feature type				
		boundary	-- Name -- boundary The administrative boundaries between	AdministrativeBoundary	1..*	voidable	See Administrativeboundary feature type				
		lowerLevelUnit	-- Name -- lower level unit Units	AdministrativeUnit	0..*	voidable	See AdministrativeUnit feature type				
		upperLevelUnit	-- Name -- upper level unit A unit	AdministrativeUnit	0..1	voidable	See AdministrativeUnit feature type				
		administeredBy	-- Name -- administered by	AdministrativeUnit	0..*	voidable	See AdministrativeUnit feature type				
		coAdminister	-- Name -- co administer	AdministrativeUnit	0..*	voidable	See AdministrativeUnit feature type				

Mapping table: instructions for use

A3 | GeographicalName

Data Type	Documentation	Data Type Attribute / Constraint	Data Type Attribute / Constraint documentati	Data Type / Values / Code List - Enumerations	Multiplicity	Voidable / Non-Voidab	Data Type Attribute	Data Type Attribute documen	Data Type / Values / Code List - Enumerations	Multiplicity	Voidable / Non-Voidab	"File name" or URL	Name of attribute	Example of one data source value
Application Schema 'LandCoverVector' (version 3.0)												Source location		
GeographicalName	Proper noun applied to a real world entity.	grammaticalGender	Class of nouns reflected in the behaviour of	GrammaticalGenderValue - common- feminine - masculine -	0..1	voidable								
		grammaticalNumber	Grammatical category of nouns that	GrammaticalNumberValue - dual - plural - singular	0..1	voidable								
		language	Language of the name, given as a three letters	CharacterString	1	voidable								
		nameStatus	Qualitative information enabling to	NameStatusValue - historical - official - other - standardised	1	voidable								
		nativeness	Information enabling to acknowledge if	NativenessValue - endonym -exonym	1									
		pronunciation	Proper, correct or standard (standard within the linguistic community concerned)	PronunciationOfName	1	voidable	pronunciationIPA		CharacterString	0..1	voidable	com2011.shp	nome_com	Morano Calabro
							pronunciationSoundLink		URI	0..1	voidable			
		sourceOfName	Original data source from which the	CharacterString	1	voidable								
		spelling	A proper way of writing the geographical name	SpellingOfName	1..*		text	Way the name is written.	CharacterString	1		com2011.shp	nome_com	Morano Calabro
							script	Set of graphic symbols	CharacterString	1	voidable			
							transliterationScheme	Method used for the names	CharacterString	0..1	voidable			
ResidenceOfAuthority	Data type representing the name and position of a residence of authority.													
		name	Name of the residence of authority.	GeographicalName	1		See Geographical name data type							
		geometry	Position of the residence of authority.	GM_Point	1	voidable								

Matching Table | Data Type

Mapping table: instructions for use

Application Schema 'AdministrativeUnits' (version 3.0)						Source Location of information				
Feature Type / Data Type	Documentation	Attribute Association role Constraint	Data Type Attribute	Multiplicity	Voidable / Non-Voidable	"File name" or URL	Name of attribute	Example of one data source value	Example of one data target value	Void Reason
Administrative Unit	-- Name -- administrative unit Unit of administration where a Member State has and/or exercises jurisdictional rights, for local, regional and national governance.									
		country							http://inspire.ec.europa.eu/codelist/CountryCode/IT	
		beginLifespanVersion							2013-11-20T14:12:20	
		endLifespanVersion							2015-11-20T14:12:20	
		geometry				com2011.shp	shape			
		inspireId	localId	1		com2011.shp	pro_com	78083	78083	
			namespace	1					AU.IT.ISTAT	
			versionId	0..1	voidable					
		name	See Geographical name data type			See Geographical name data type				
		nationalCode				com2011.shp	pro_com	78083	78083	
		nationalLevel							http://inspire.ec.europa.eu/codelist/AdministrativeHierarchyLevel/4thOrder	
		nationalLevelName							Comune	
		residenceOfAuthority	See ResidenceOfAuthority data type							Unknown
		condominium	See condominium feature type							
		boundary	See Administrativeboundary feature type			com2011Boundary	nome_com + nome_com_ad	Morano Calabro + Rotonda	#Morano Calabro_Rotonda	
		lowerLevelUnit	See AdministrativeUnit feature type							
		upperLevelUnit	See AdministrativeUnit feature type						#Provincia_78	
		administeredBy	See AdministrativeUnit feature type							

Mapping table: instructions for use

A3 | GeographicalName

Application Schema 'LandCoverVector' (version 3.0)						Source location of information					
Data Type	Data Type Attribute / Constraint	Multiplicity	Data Type Attribute	Data Type / Values / Code List - Enumerations	Multiplicity	"File name" or URL	Name of attribute	Example of one data source value	Example of one data target value	Void Reason	Remarks
GeographicalName	grammaticalGender	0..1							common		
	grammaticalNumber	0..1							singular		
	language	1							ita		
	nameStatus	1							official		
	nativeness	1							endonym		
	pronunciation	1	pronunciationIPA	CharacterString	0..1	com2011.shp	NOME_COM	Morano Calabro	Morano Calabro		
			pronunciationSoundLink	URI	0..1						
	sourceOfName	1								Unknown	
	spelling	1..*	text	CharacterString	1	com2011.shp	NOME_COM	Morano Calabro	Morano Calabro		
			script	CharacterString	1				Latn		
			transliterationScheme	CharacterString	0..1						
ResidenceOfAuthority											
	name	1	See Geographical name data type								
	geometry	1									

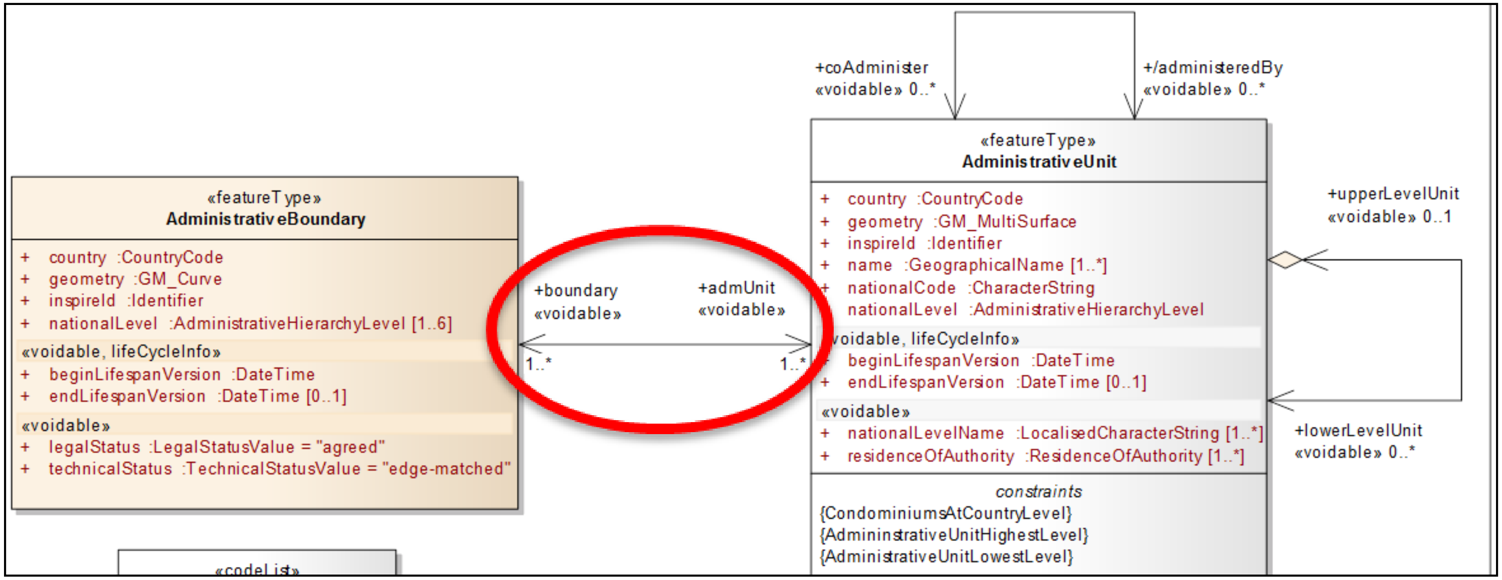
Matching Table | Data Type

Summary

- Introduction
- Source Data Model
- Target Data Model
- Mapping table: instruction for use
- Common mapping problems
- Open source Hale transformation tool
- Data Transformation
- Data Validation
- Creation and validation of a harmonized GML dataset.

Common Mapping problems

While filling in the mapping table we realise we are not able to map the association between the Administrative Unit feature type and the Administrative Boundary feature type, as we have no information in our source dataset (no shapefile nor table) linking municipalities to related boundaries.



Common Mapping problems

In the case of this example we created the dataset of boundaries using GIS tools capabilities on com2011.shp file (splitting the perimeter of the polygons into polylines). For convenience of use we also created a table to associate municipalities to their boundaries i.e. Administrative Unit Ids (PRO_COM field) to related boundaries IDs (ID_Boundary field).

N_ID_Boundary	PRO_COM	NOME_COM	ID_Boundary
6	76070	Rotonda	Boundary_78083_76070
3	78136	Saracena	Boundary_78083_78136
7	76028	Chiaromonte	Boundary_78083_76028
5	76097	Viggianello	Boundary_78083_76097
1	78033	Castrovillari	Boundary_78083_78033
4	78084	Mormanno	Boundary_78083_78084
2	78111	San Basile	Boundary_78083_78111
6	78083	Morano Calabro	Boundary_78083_76070
3	78083	Morano Calabro	Boundary_78083_78136
7	78083	Morano Calabro	Boundary_78083_76028
5	78083	Morano Calabro	Boundary_78083_76097
1	78083	Morano Calabro	Boundary_78083_78033
4	78083	Morano Calabro	Boundary_78083_78084
2	78083	Morano Calabro	Boundary_78083_78111

Summary

- Introduction
- Source Data Model
- Target Data Model
- Mapping table: instruction for use
- Common mapping problems
- Open source Hale transformation tool
- Data Transformation
- Data Validation
- Creation and validation of a harmonized GML dataset.

HALE transformation tool

The mapping between source and target properties defined in the matching tables can now be used to set the encoding rules needed to obtain an harmonized dataset by means of a software transformation tool.

Among the many available software, in the case of this example focus has been given to open source tool HALE (HUMBOLDT Alignment Editor), to define and evaluate conceptual schema mapping and to transform geodata based on these mapping.

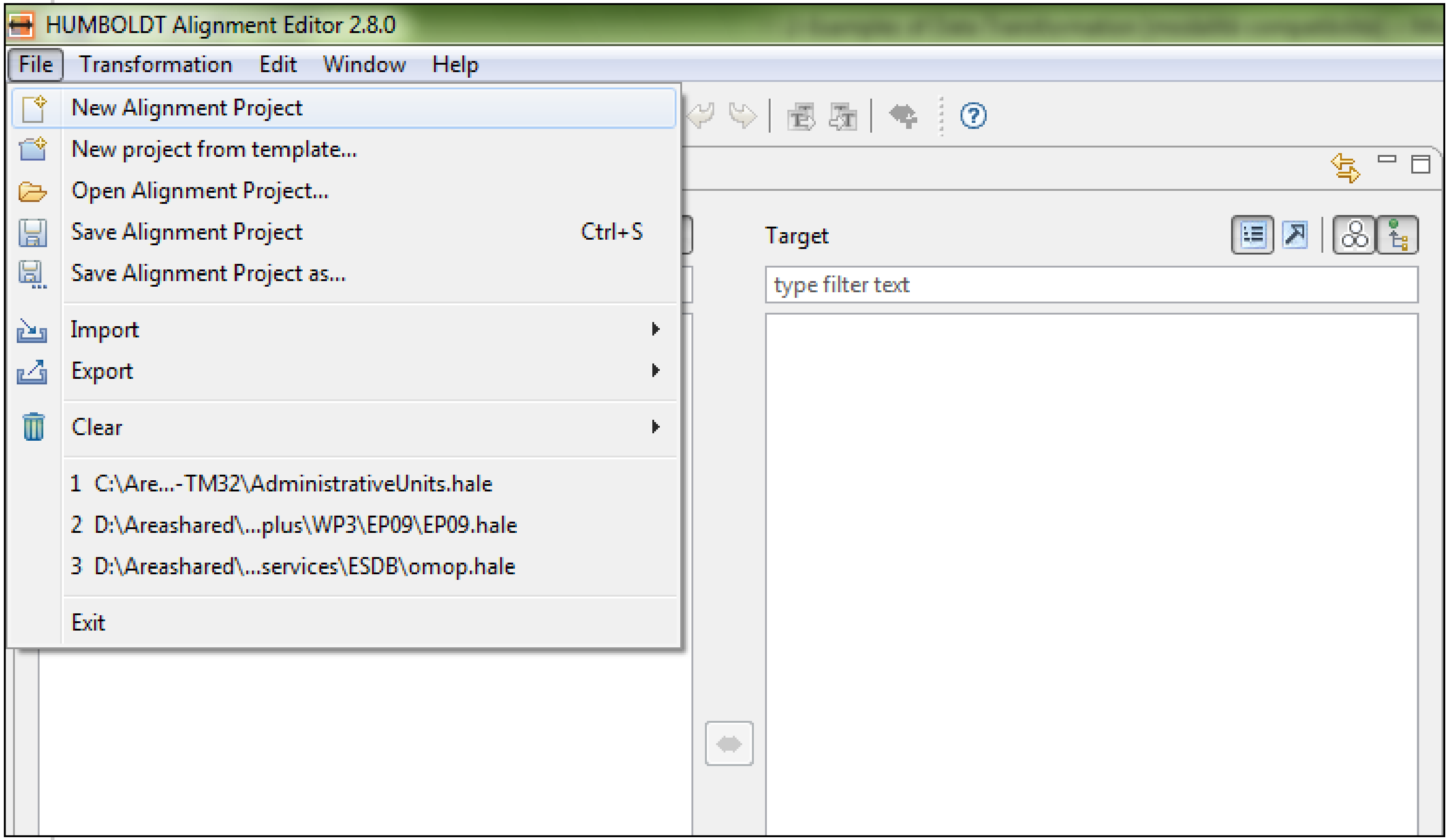
http://hale.igd.fraunhofer.de/2.8.0/help/index.jsp

HALE transformation tool

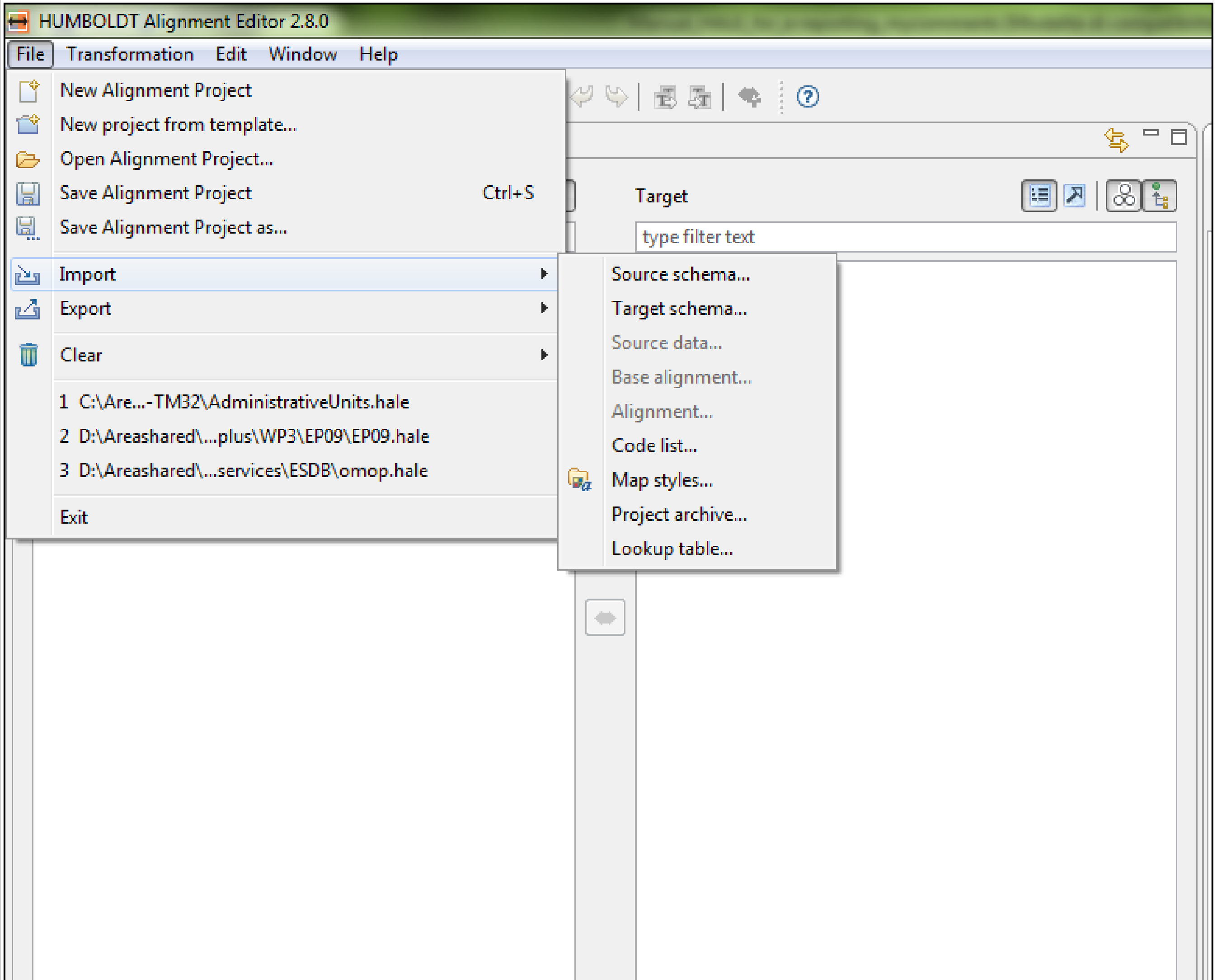
The general workflow for transforming source data according to target schema using the HALE tool is as follows:

- Load the schema of the source data
- Load the target schema
- Load the source dataset
- Operate the mapping
- Save the transformed data to a GML file.
- Validate harmonized GML dataset.

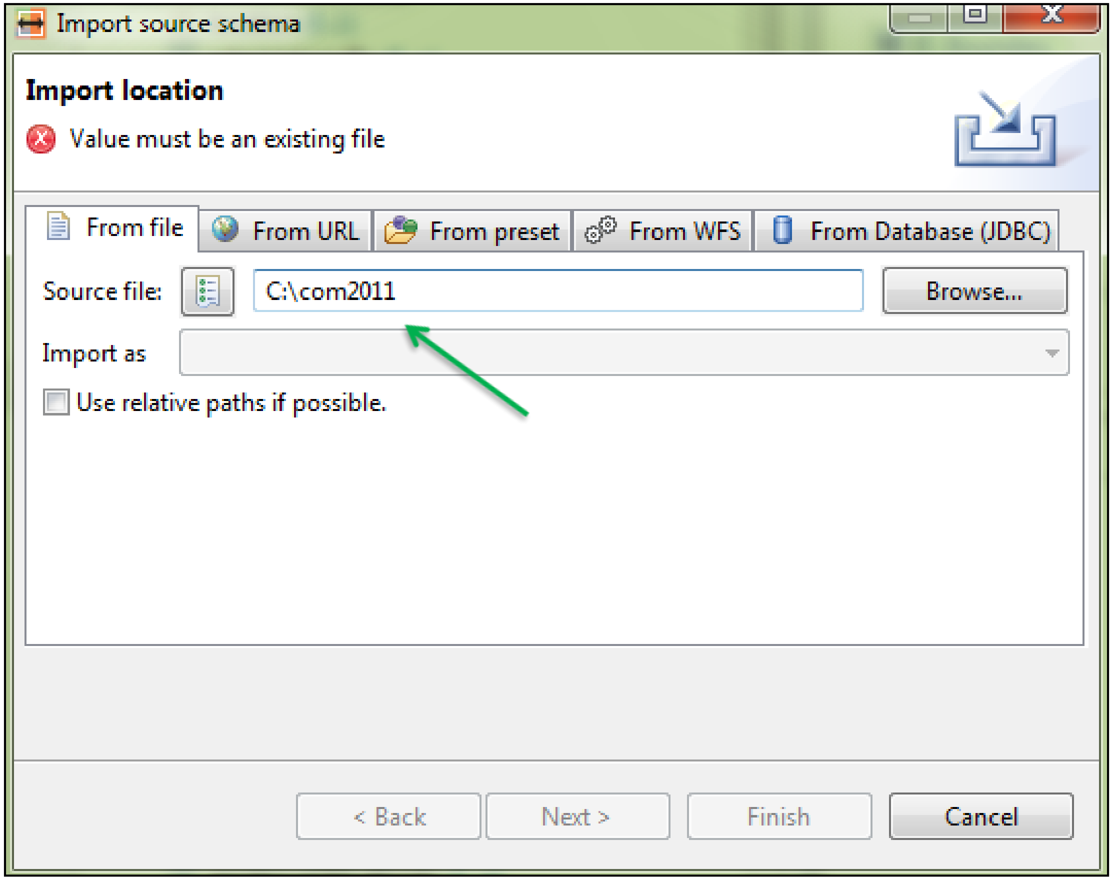
Set up the project in HALE



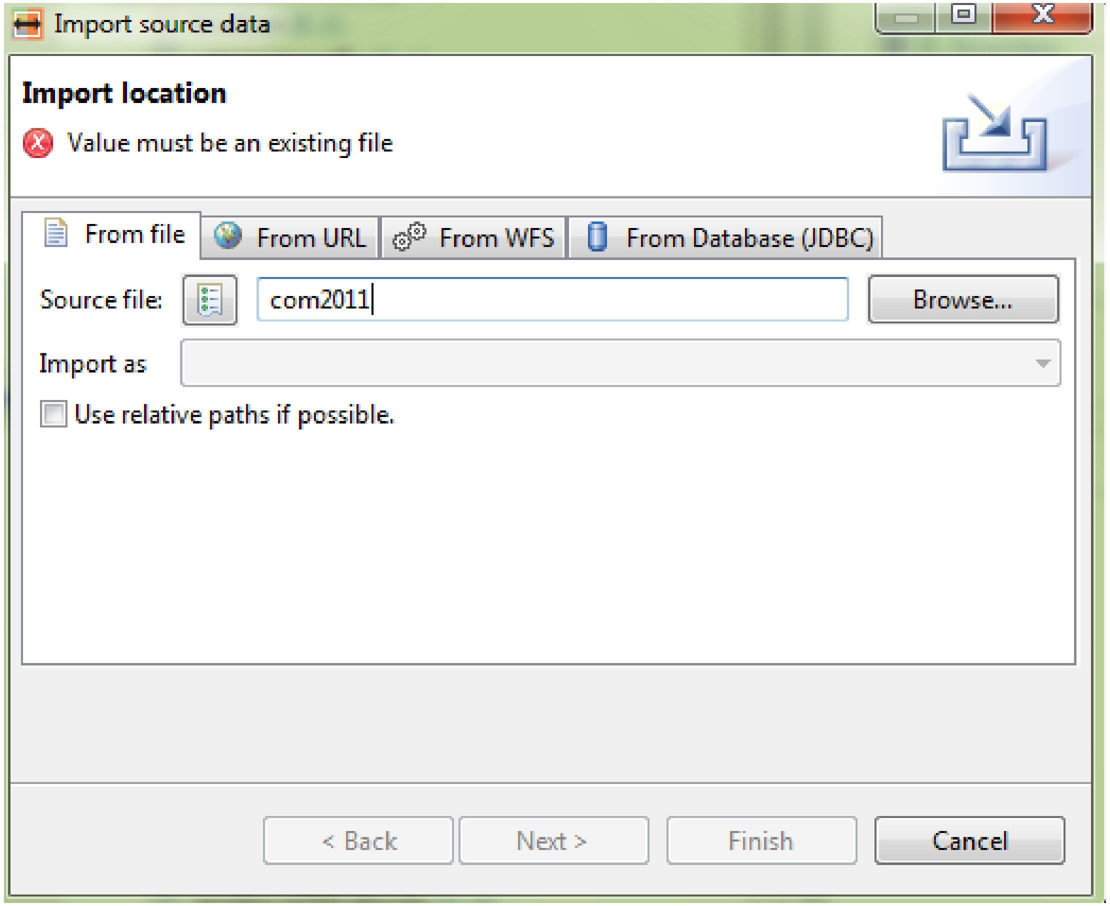
Import source and target schema



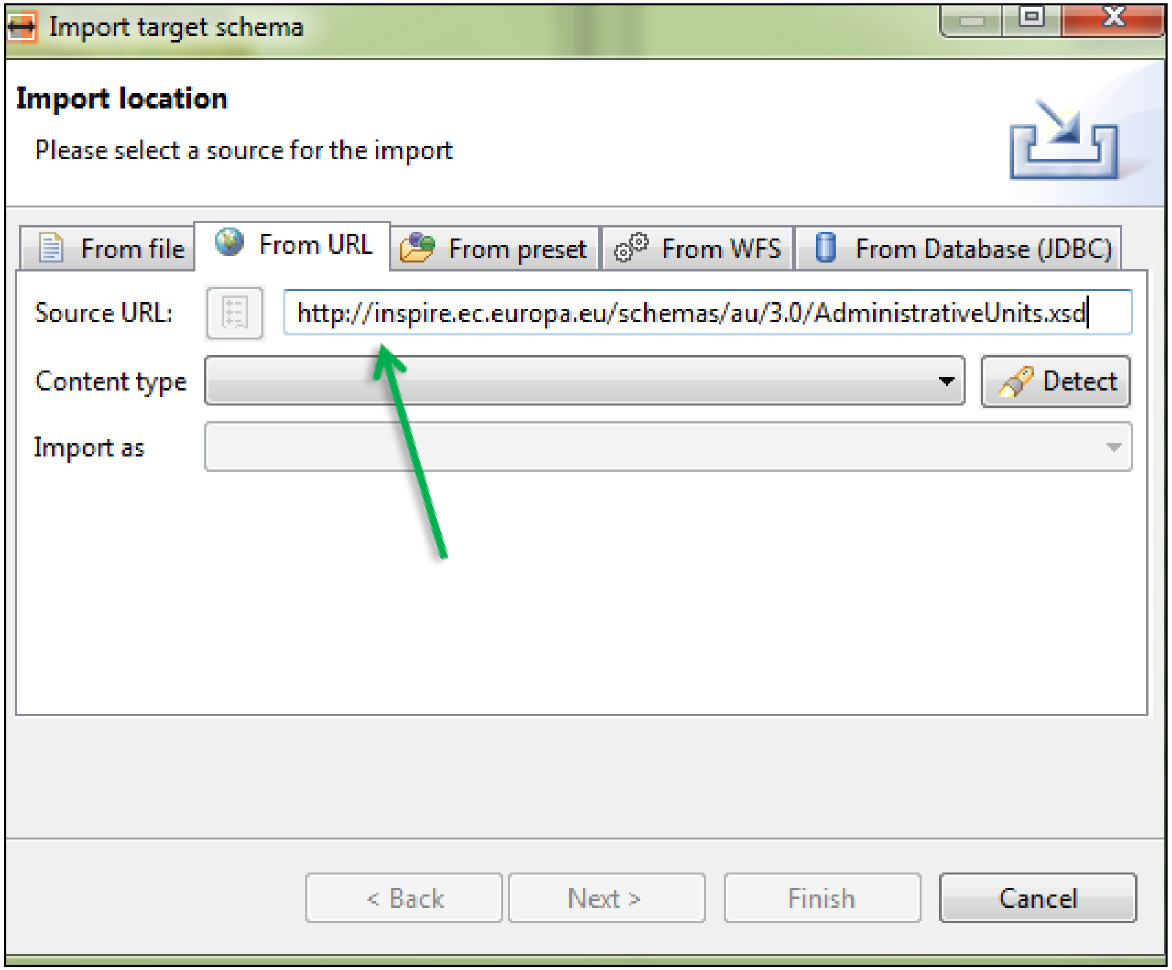
Import source schema



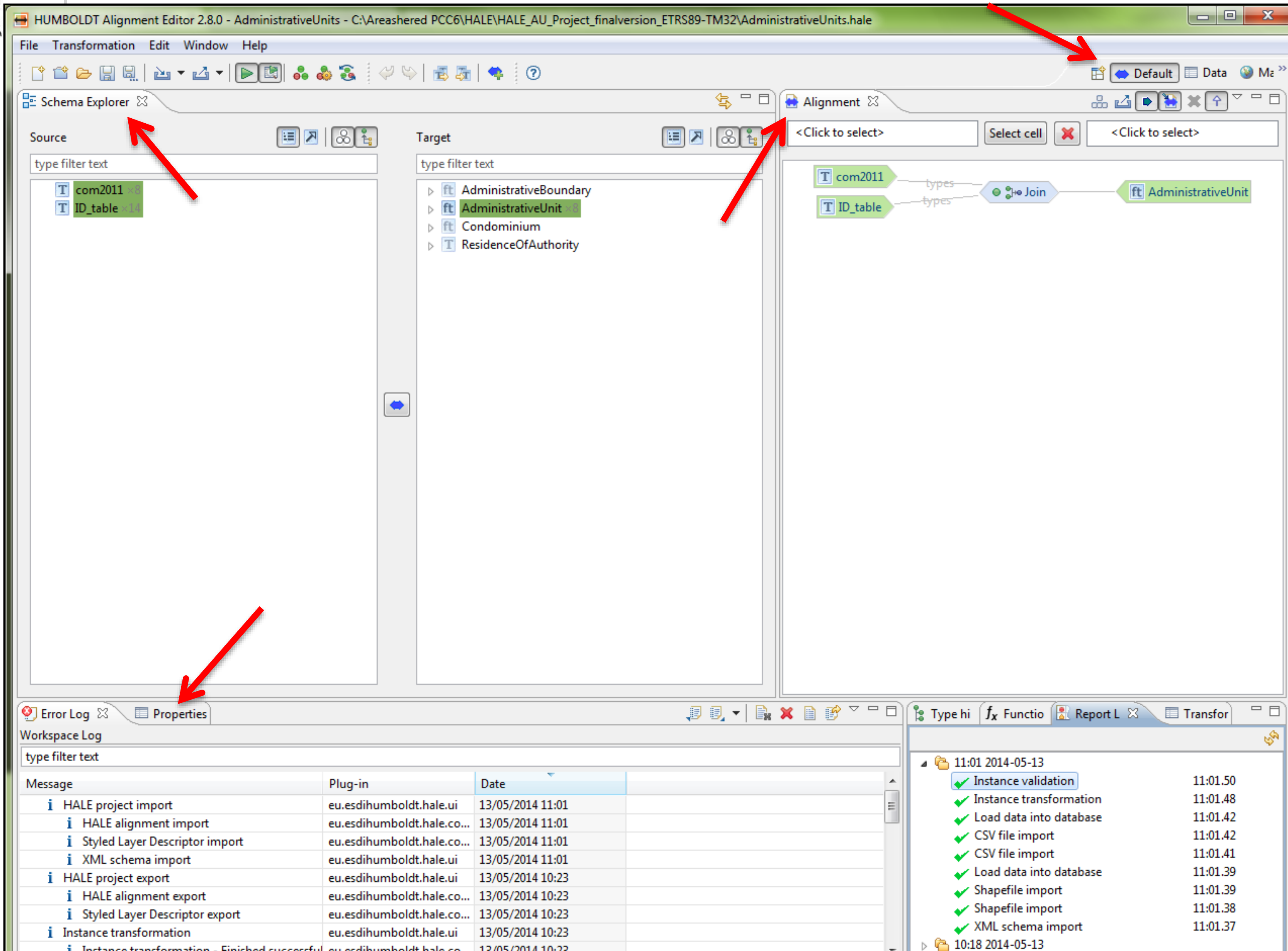
Import source data



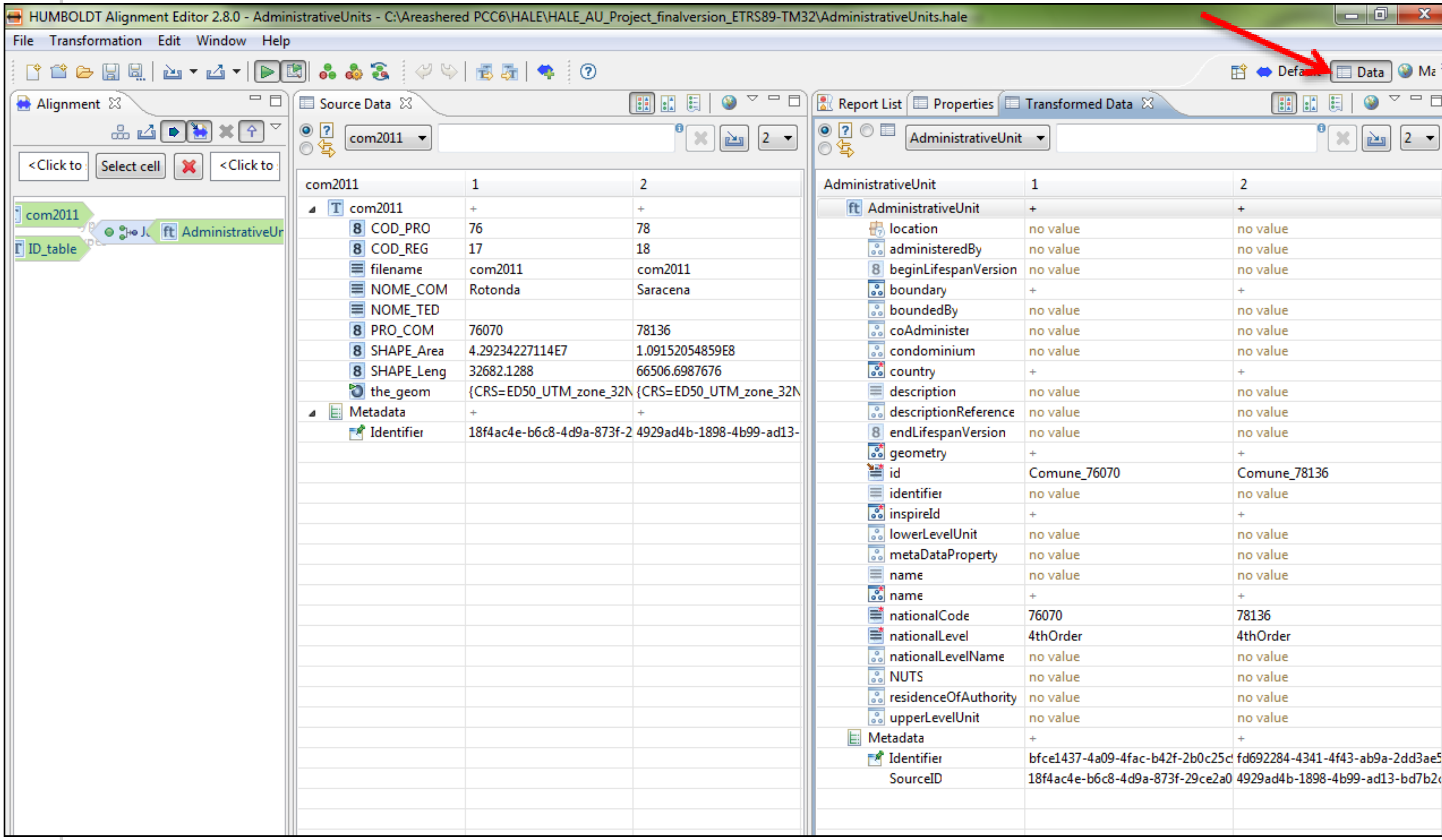
Import target schema



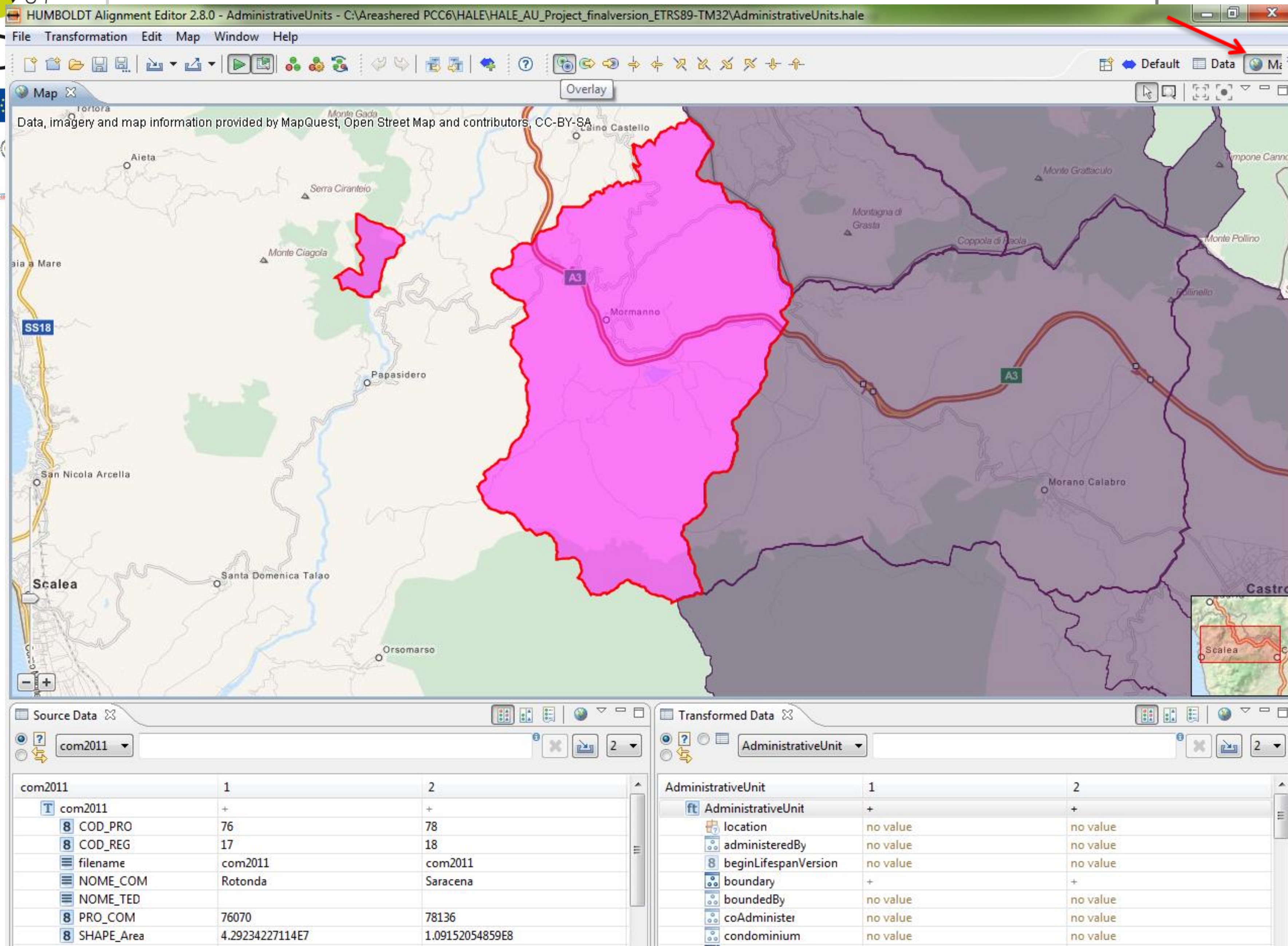
The Hale Workbench: the Default perspective



The Hale Workbench: the Data perspective



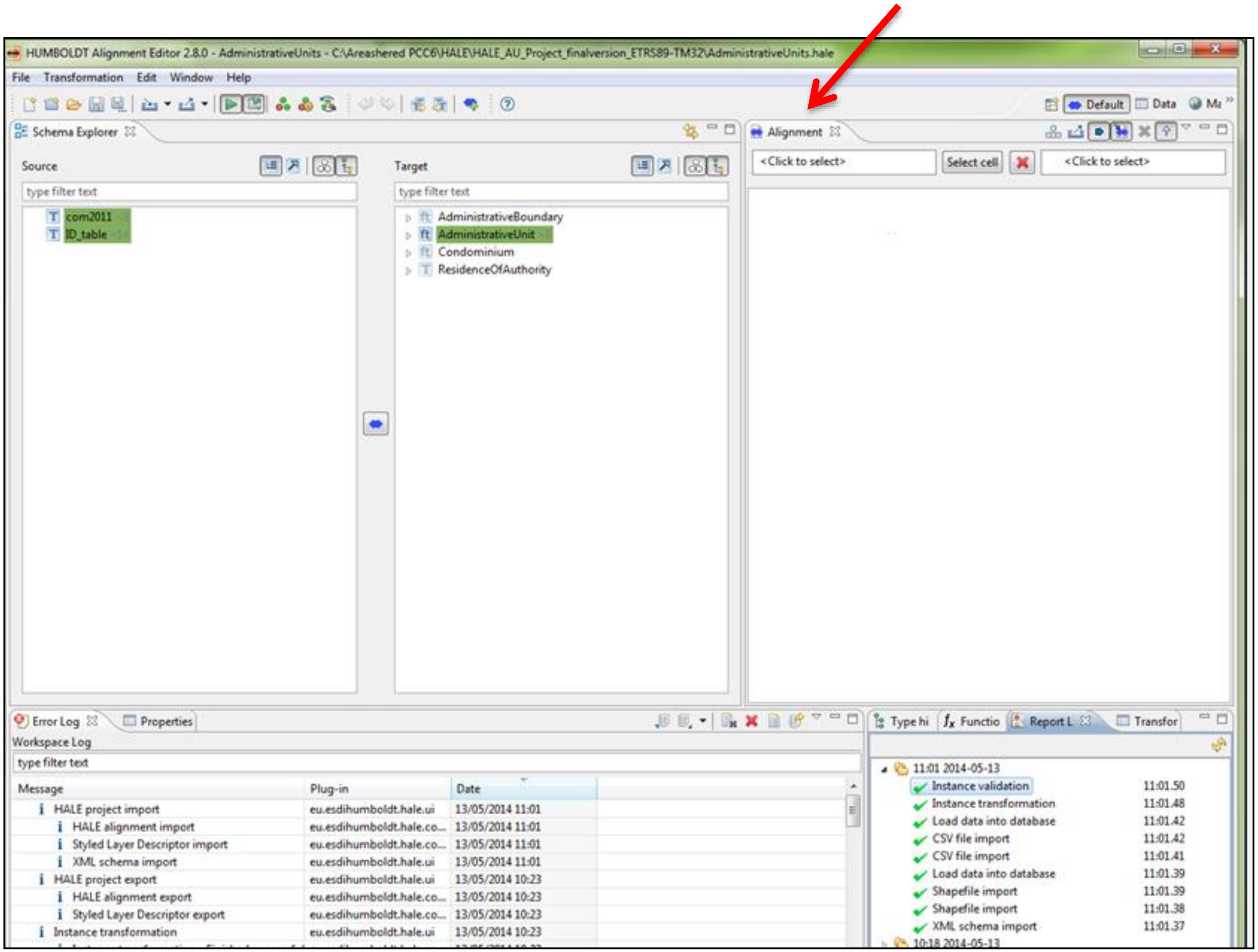
The Hale Workbench: the Map perspective



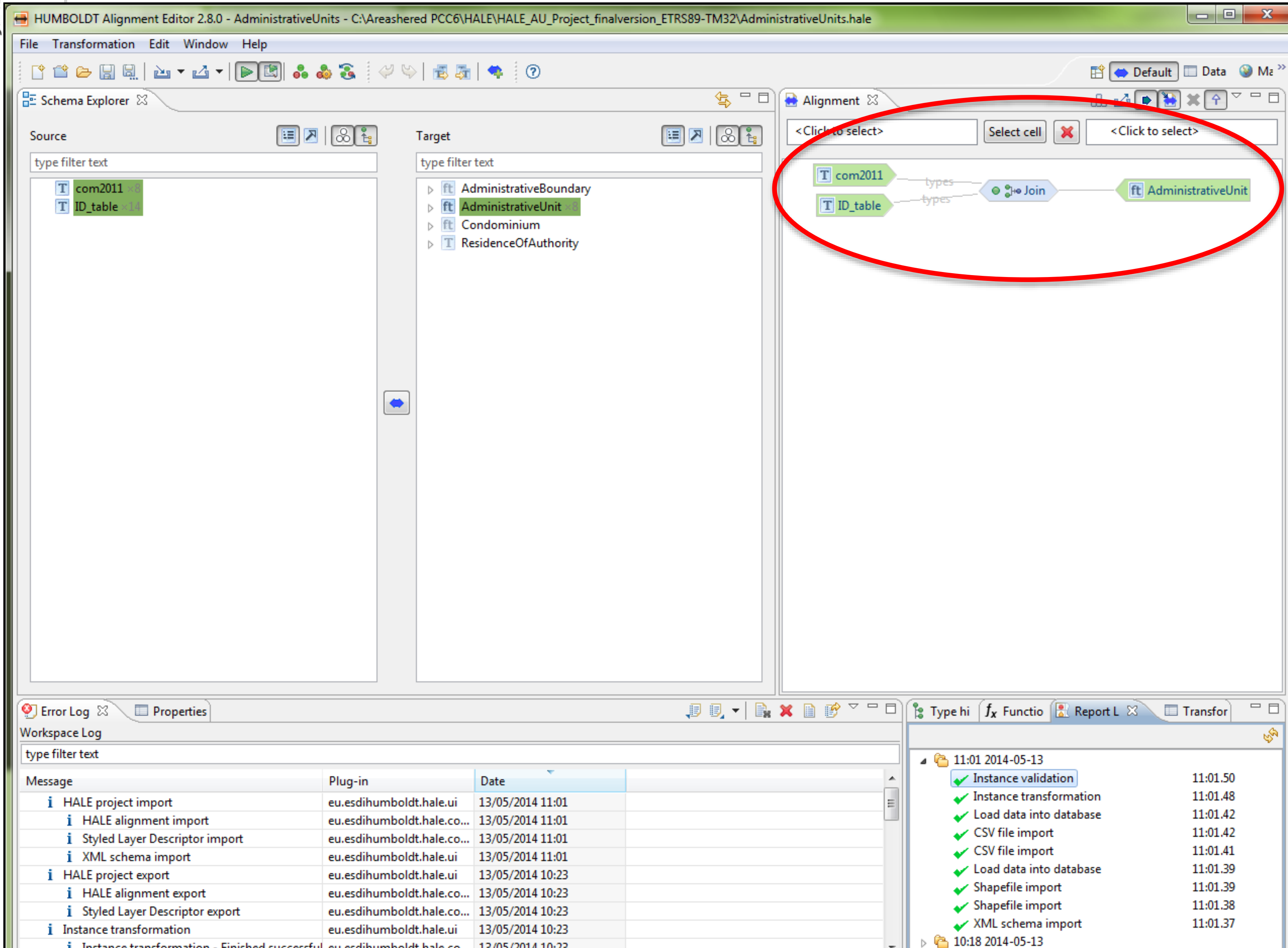
Summary

- Introduction
- Source Data Model
- Target Data Model
- Mapping table: instruction for use
- Common mapping problems
- Open source Hale transformation tool
- Data Transformation
- Data Validation
- Creation and validation of a harmonized GML dataset.

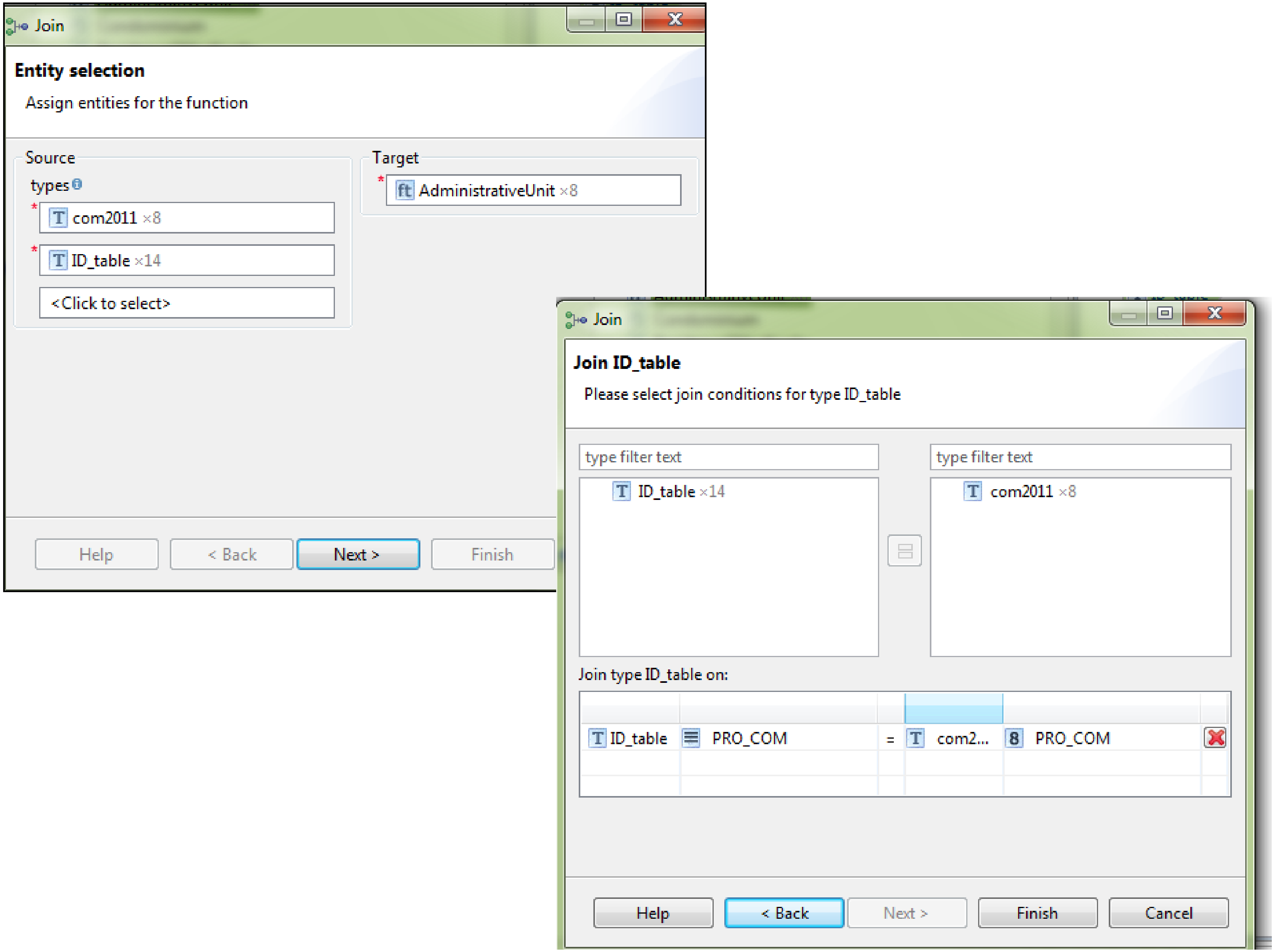
Data Transformation



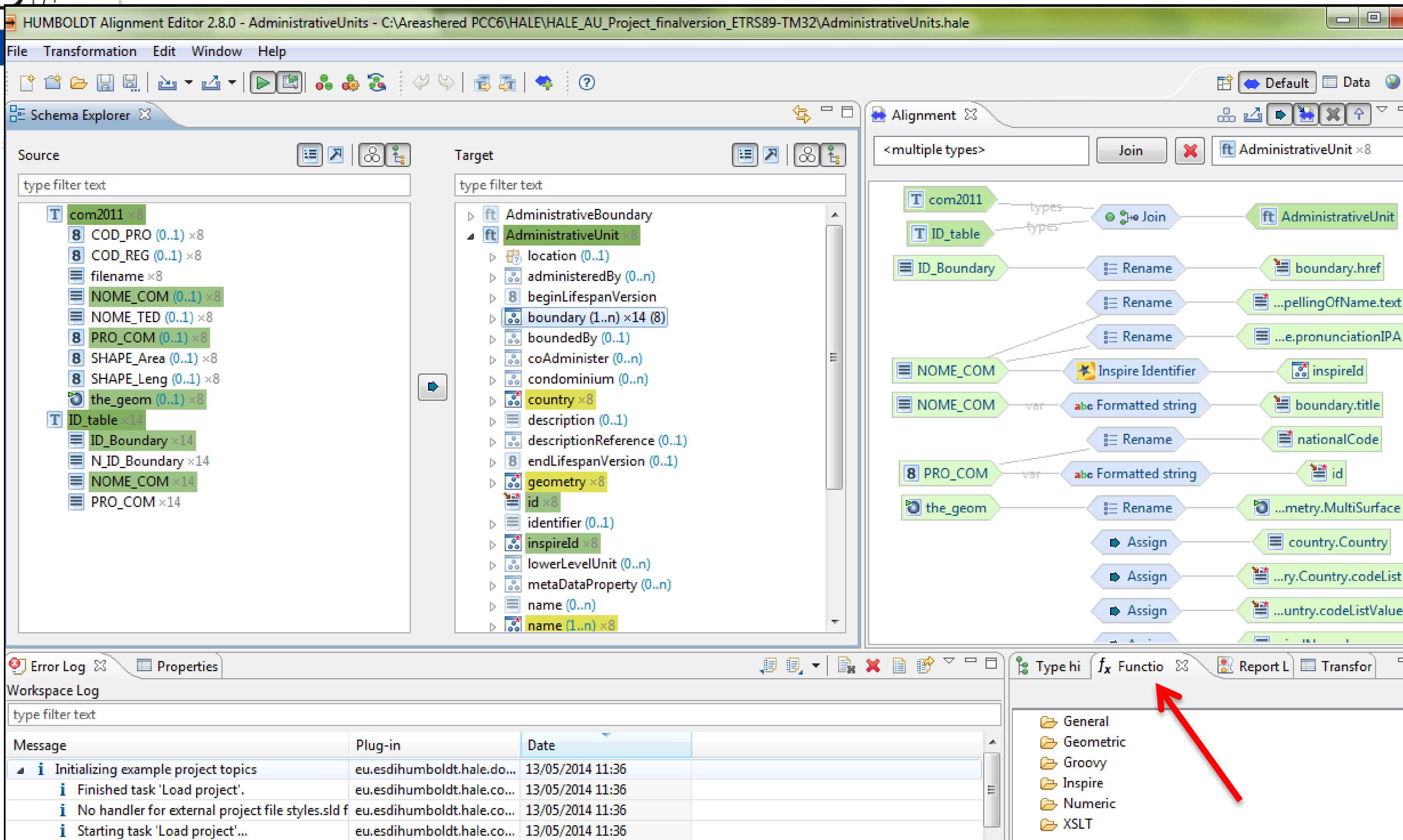
Type Relations



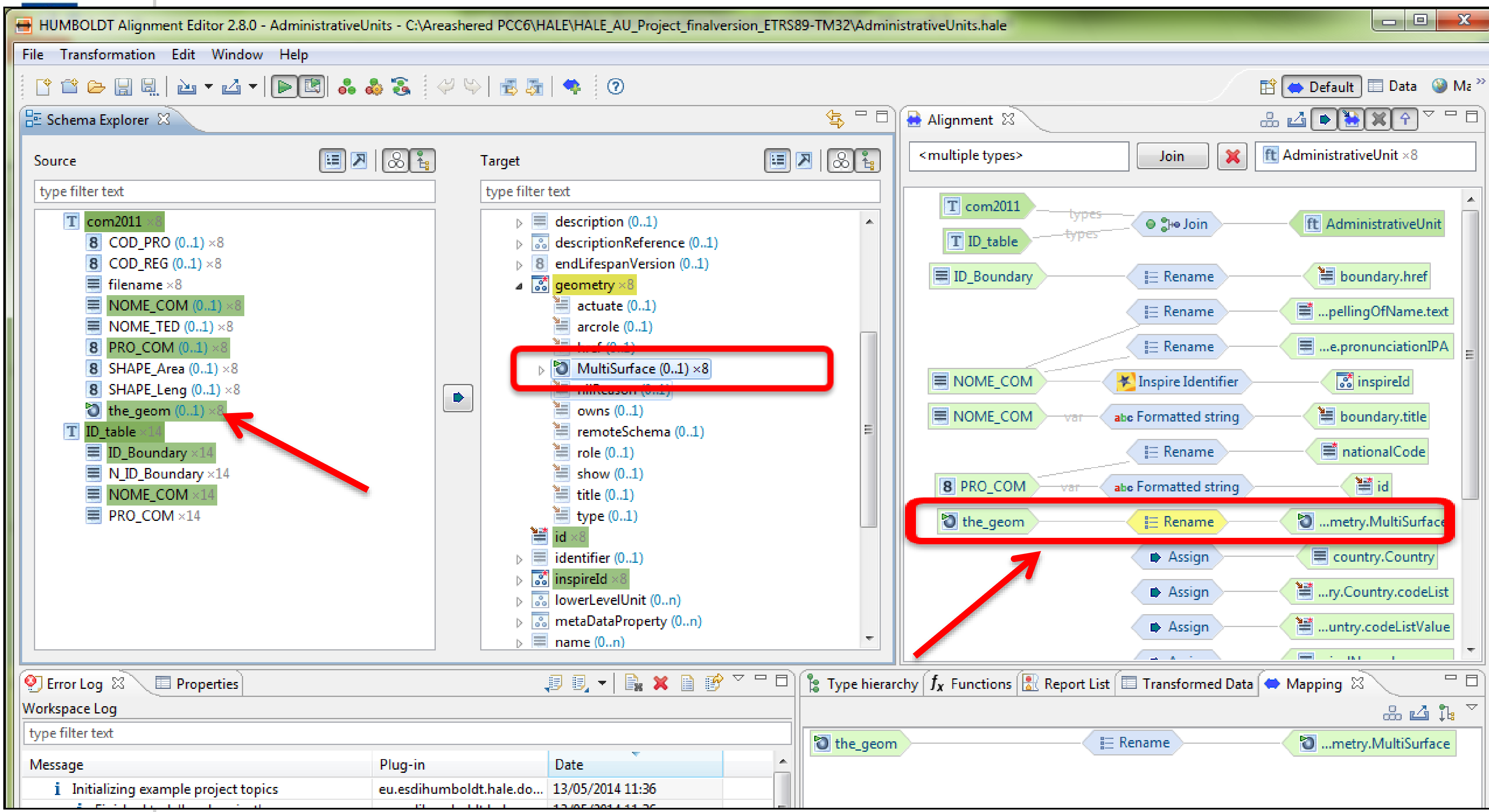
The Join function



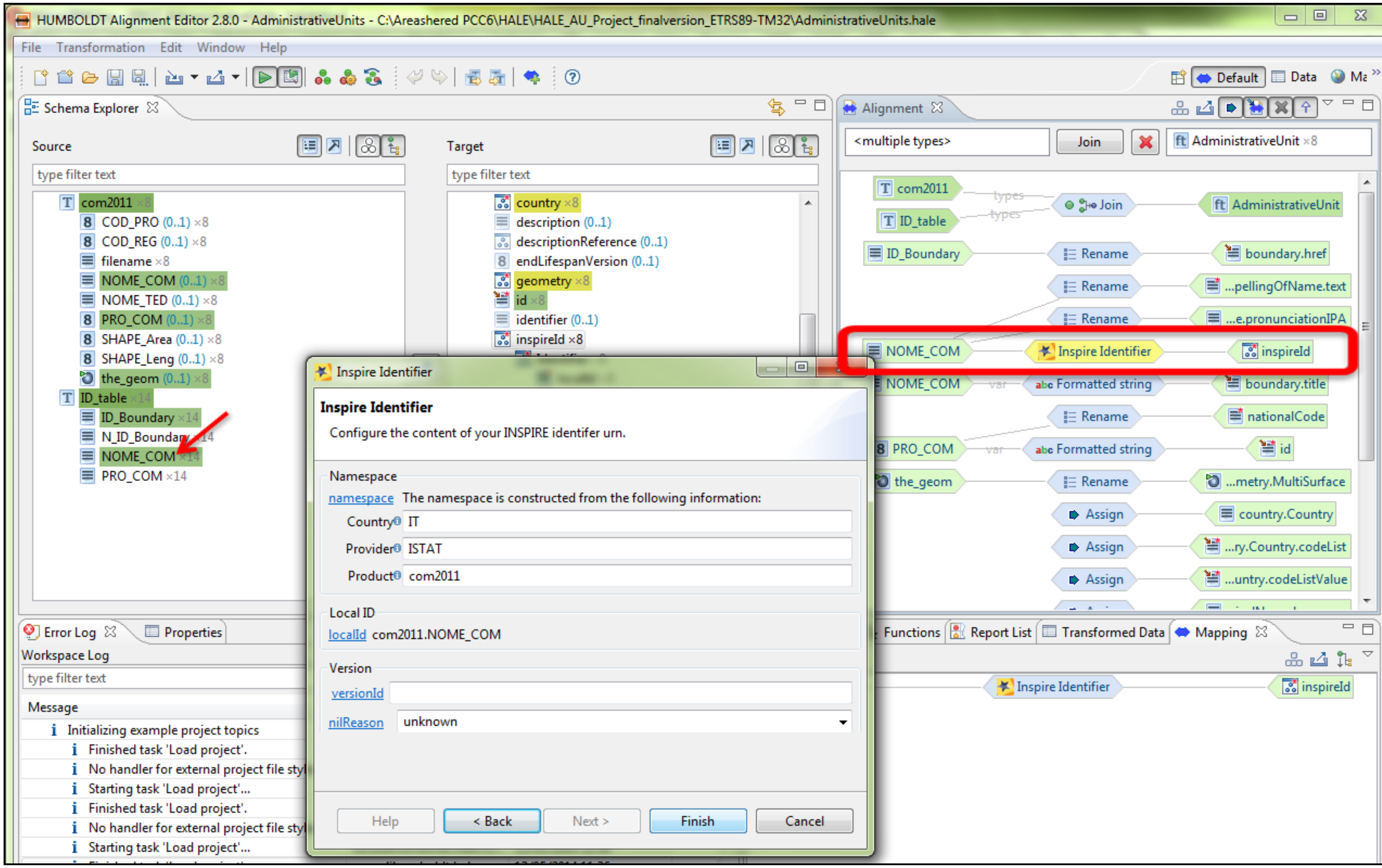
Property Relations



Mapping the geometry property

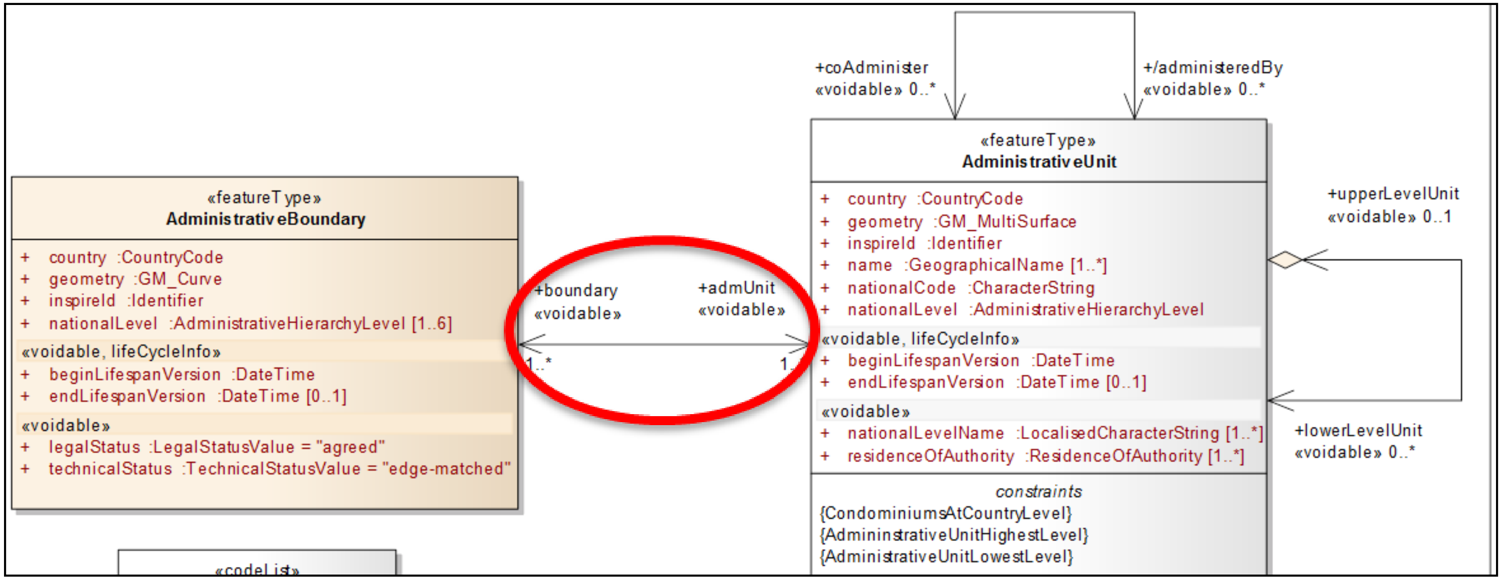


The INSPIRE Identifier

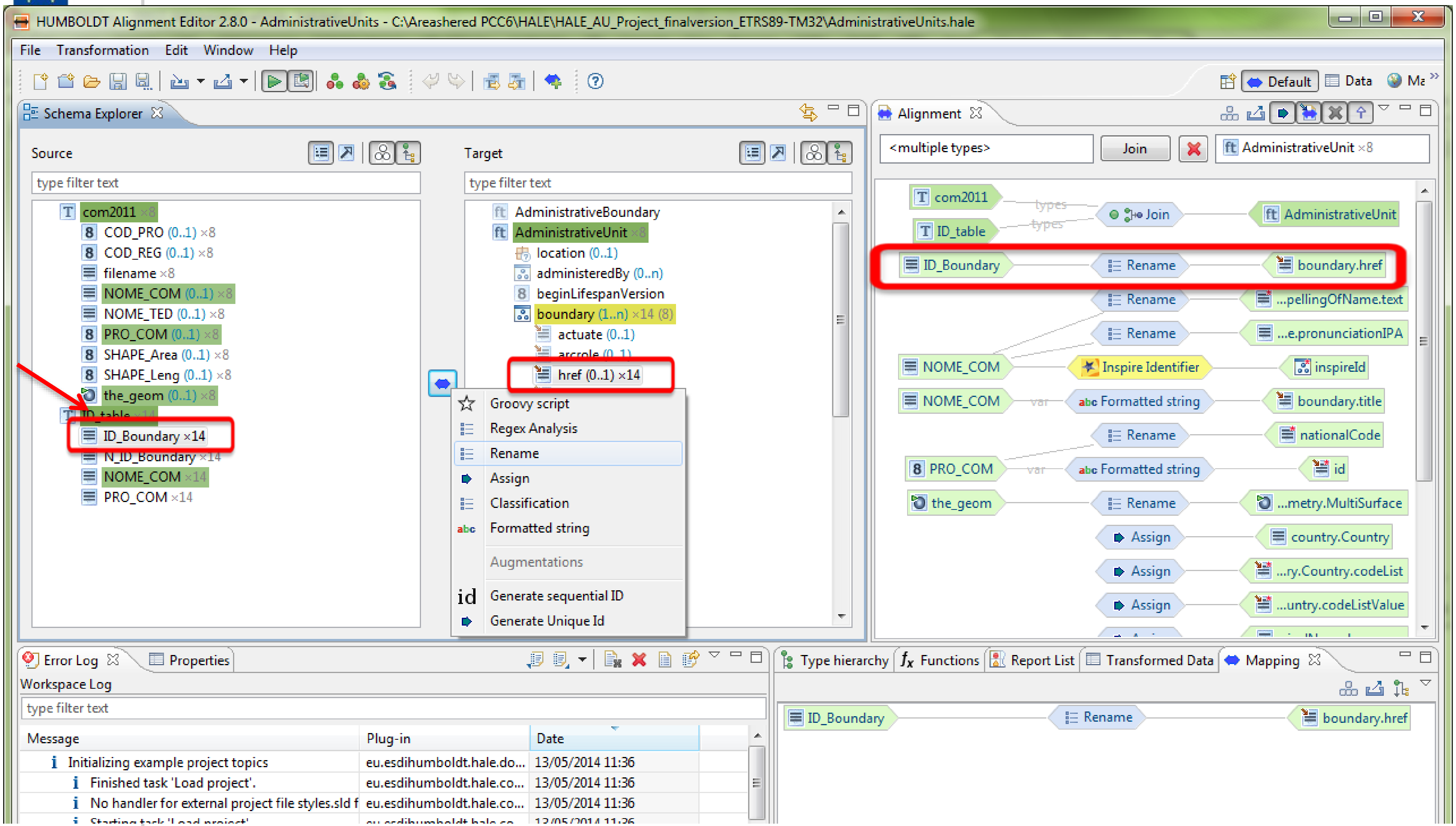


Mapping the association between municipality and its boundaries

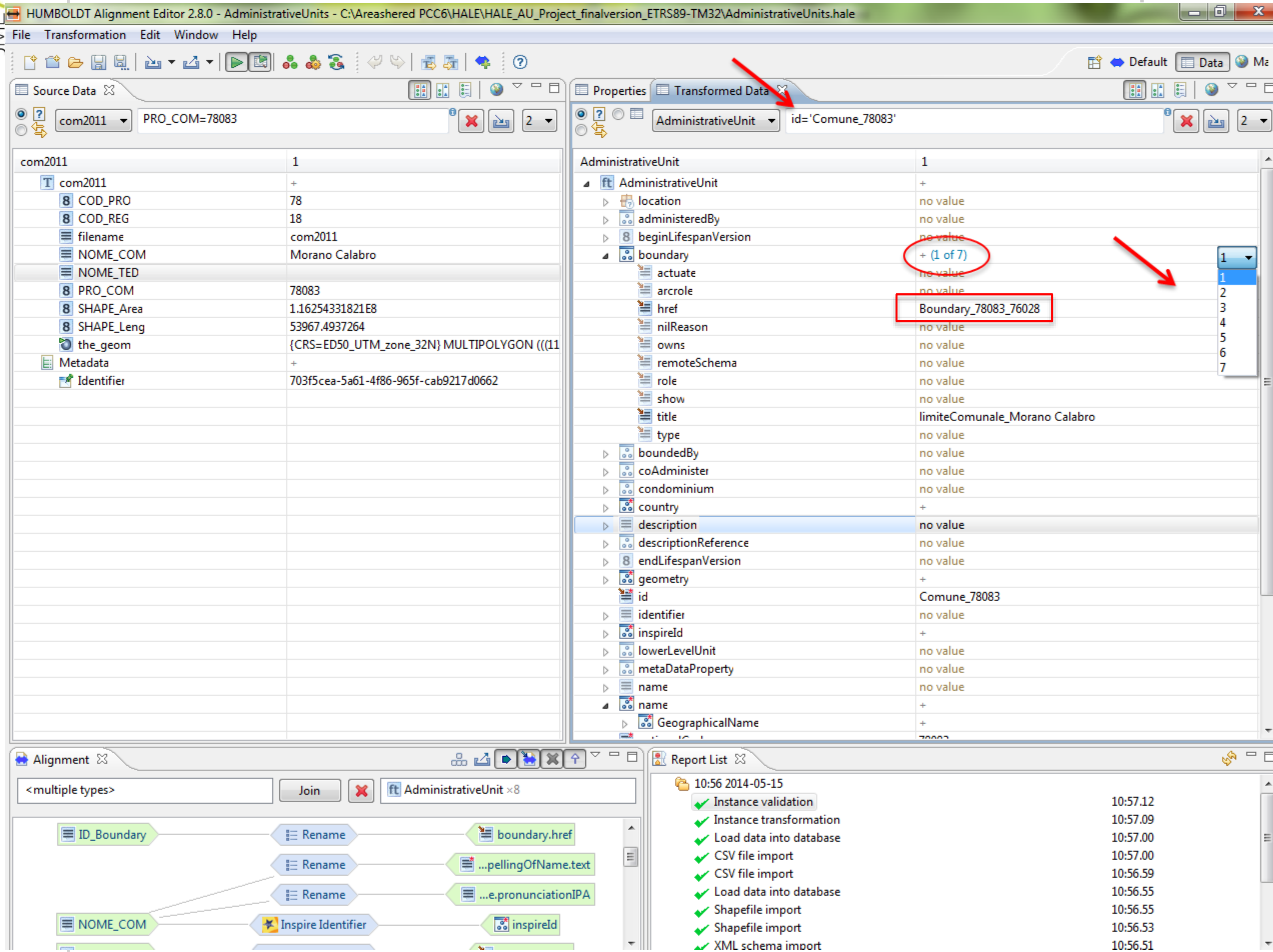
To map the association between the Administrative Unit feature type and the Administrative Boundary feature type (i.e. to link each municipality to its boundaries) we have to link the *href* attribute of the *boundary* element in the Administrative Unit feature type to the ID of the boundary contained in the ID_table.



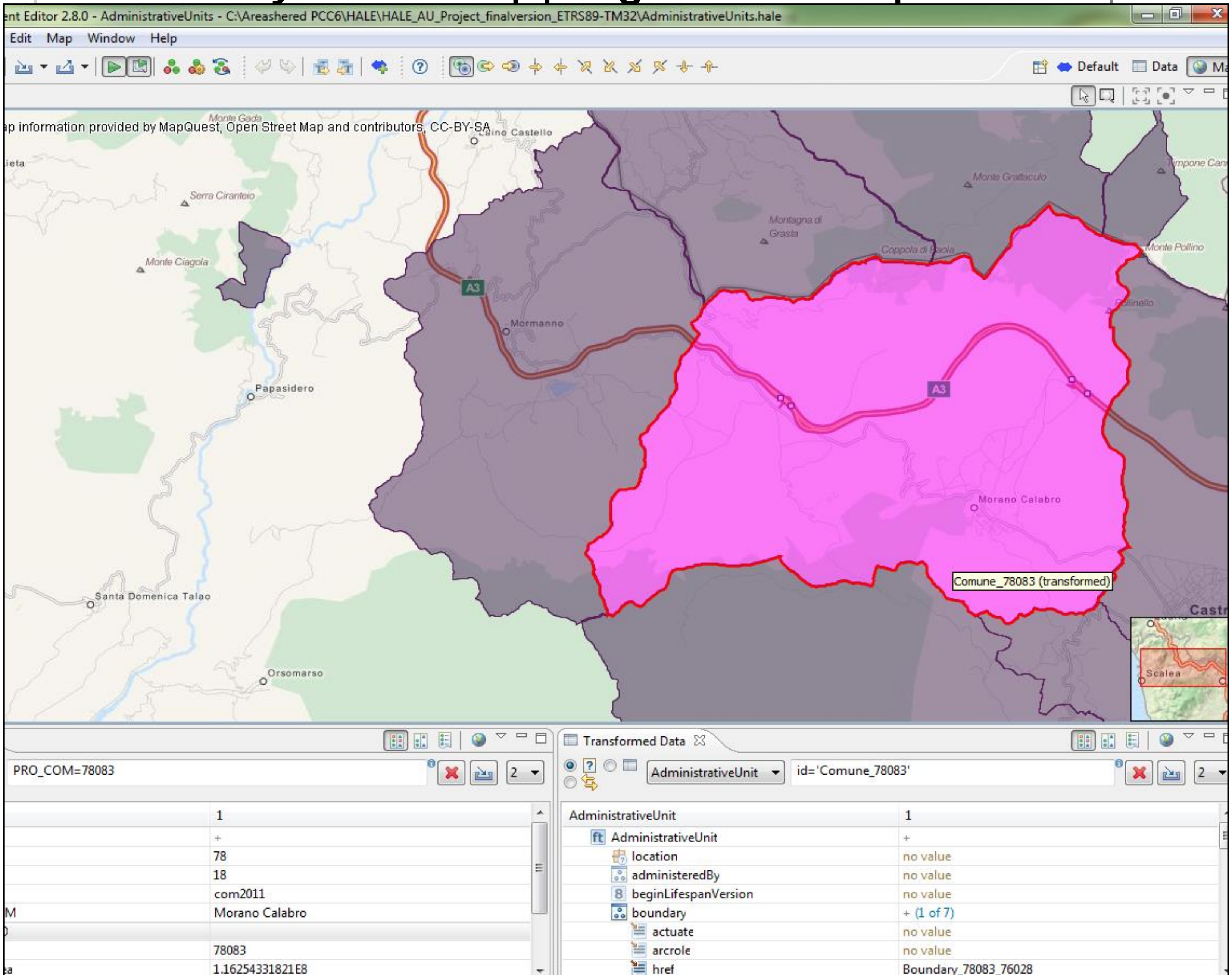
Mapping the association between municipality and its boundaries



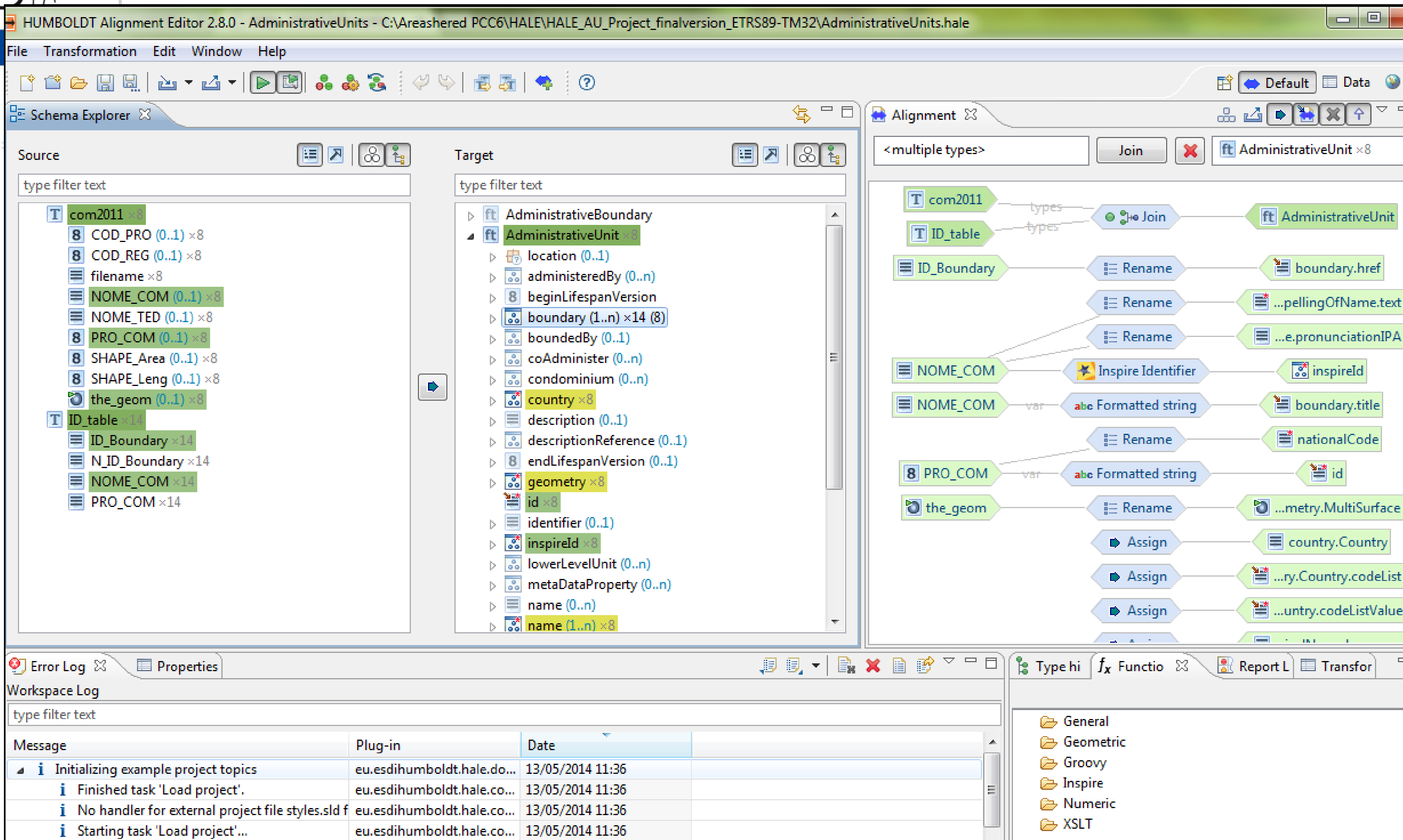
Boundary field mapping in the Data View



Boundary field mapping in the map view



Complete Hale project



Summary

- Introduction
- Source Data Model
- Target Data Model
- Mapping table: instruction for use
- Common mapping problems
- Open source Hale transformation tool
- Data Transformation
- Data Validation
- Creation and validation of a harmonized GML dataset.

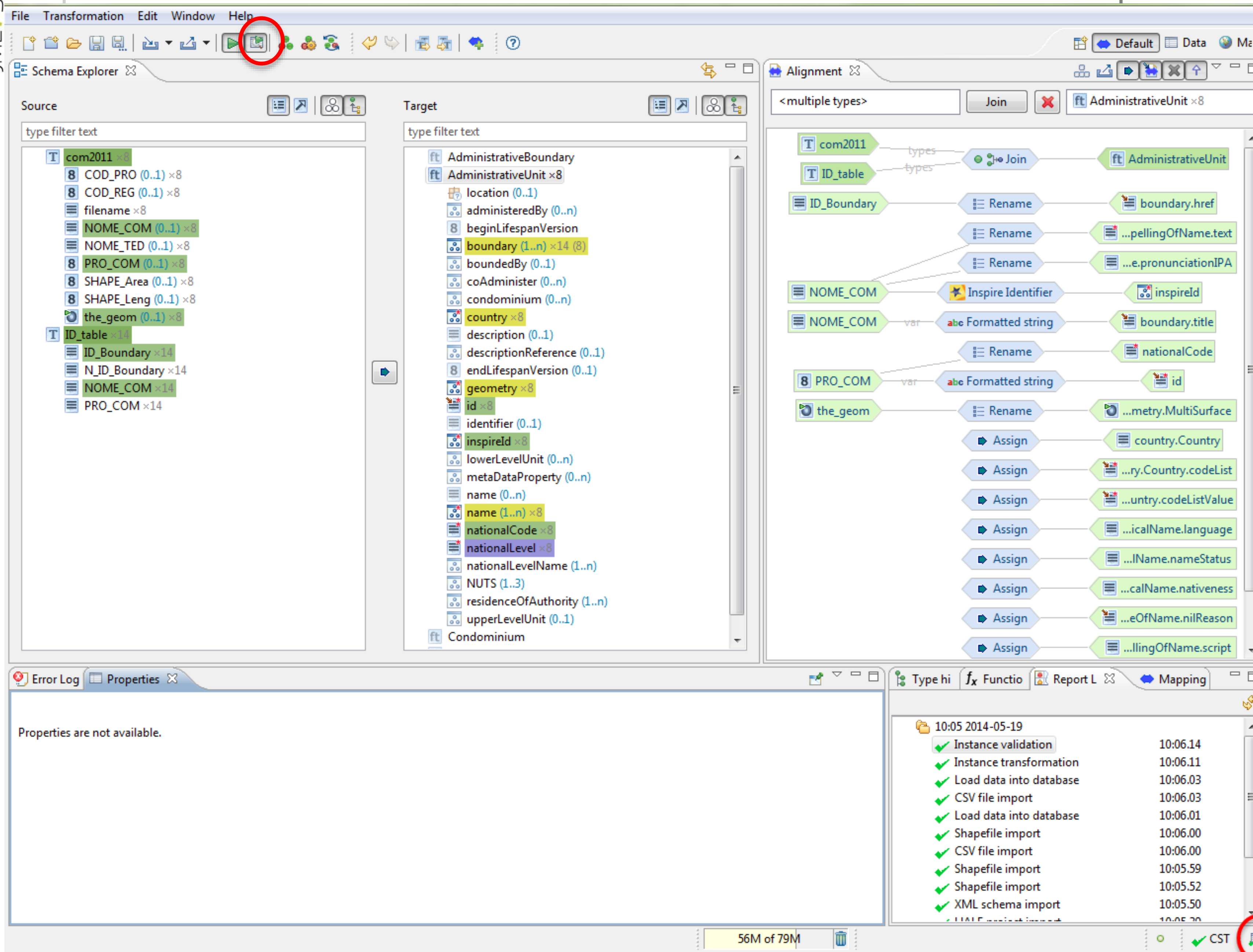
Data Validation in Hale

When targeting a specific schema, HALE verifies that your transformation result follows the target schema's structure (i.e. performs a schema validation). Moreover HALE checks whether your mapping follows some of the constraints defined by the schema, like mandatory properties and restrictions on property values.

Validation of instances in HALE is currently supported for XML based schemas and can be performed:

- on the transformed instances available in the mapping phase ('live' validation)
- on the exported transformation result (validation of 'exported data')

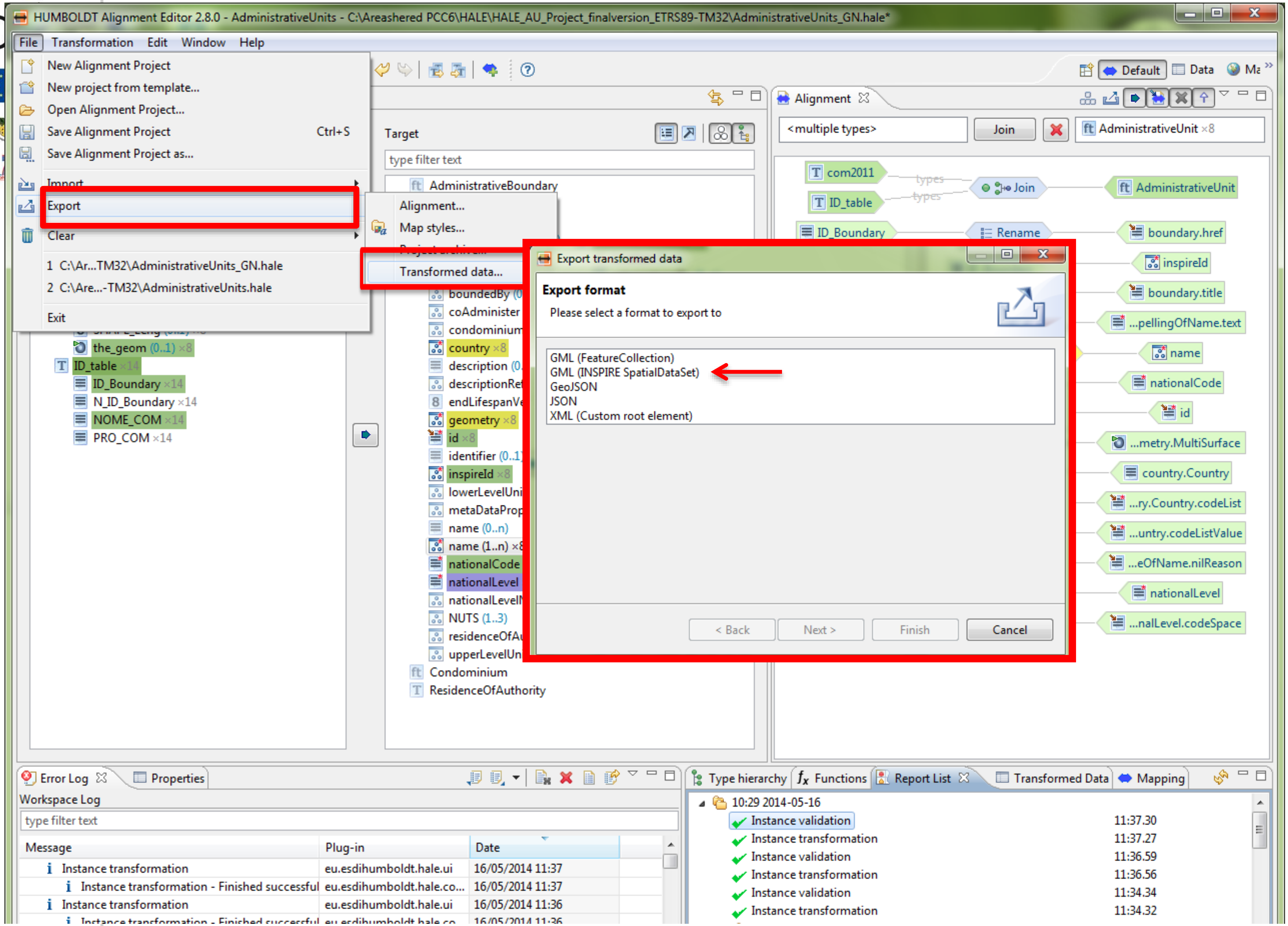
Hale: Live Validation



Summary

- Introduction
- Source Data Model
- Target Data Model
- Mapping table: instruction for use
- Common mapping problems
- Open source Hale transformation tool
- Data Transformation
- Data Validation
- Creation and validation of a harmonized GML dataset.

Export transformed data into a GML dataset file



Validate GML dataset file