



INSTITUTO NACIONAL DE ESTATÍSTICA  
STATISTICS PORTUGAL

# » Harmonização de CDG's com Hale



DMSI/GEO

 (05-07-2016)  








# Harmonização

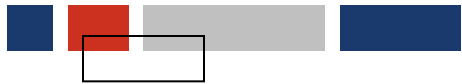
- **O que é a Harmonização?**
- Para **Hale**
  - *“The alignment is the mapping between source and target schemas. It defines relations between source and target entities (types or properties). Based on the defined relations a transformation is derived.”*
- Segundo **Dean M. Hintz**
  - *“Core to the harmonization workflow is the transformation process which reshapes source schema and geometry to match the required destination structure.” (Hintz,2012,1)*

- **Antes de começar a trabalhar com Hale ou outra ferramenta de harmonização**
- Ler (e reler) as especificações dos dados!
  - ***Data Specification - Technical Guideline***
  - Modelos UML
- Preparar com antecedência a Matching Table
- Procurar exemplos de CDG's já harmonizados nos ***Thematic Clusters***.
- Partilhar conhecimentos e experiências com a comunidade!

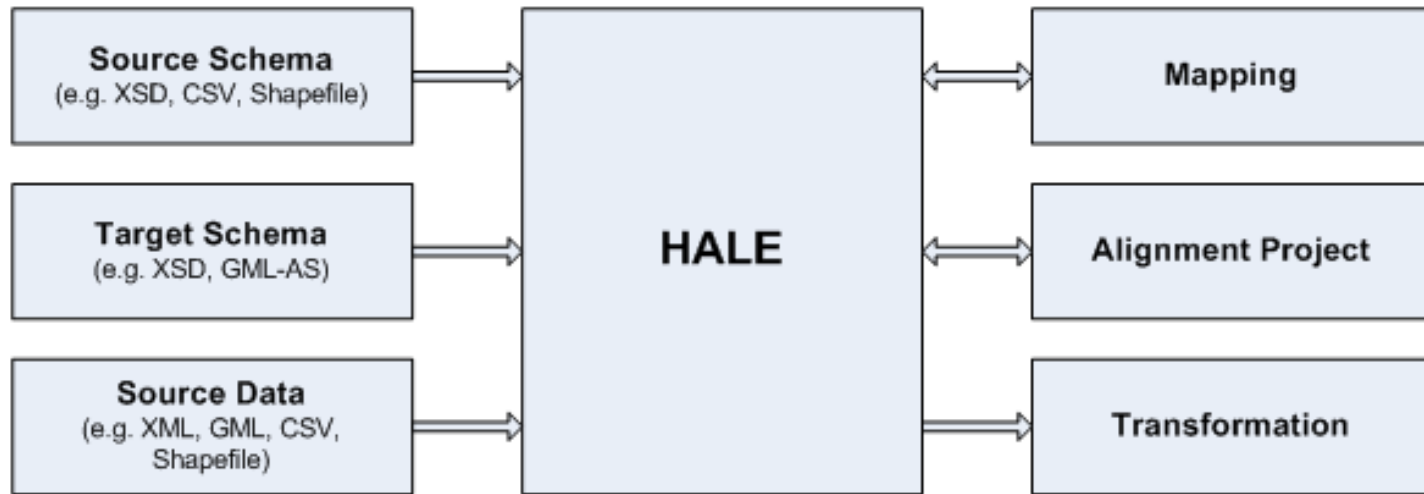


# Hale

- **HUMBOLDT** **A**lignment **E**ditor 
- Software utilizado para criar mapping entre schemas diferentes e aplicar a transformação resultante
- *Software Open Source*
- Fornece uma interface gráfica rica, textual e especificamente aprovado para especialistas em IG
- *Feedback* instantâneo sobre o processo de harmonização de dados 
- [Download](#) versão 2.9.4 (2015-11-01) versão 32 e 64 bit para windows, Mac OS, Linux

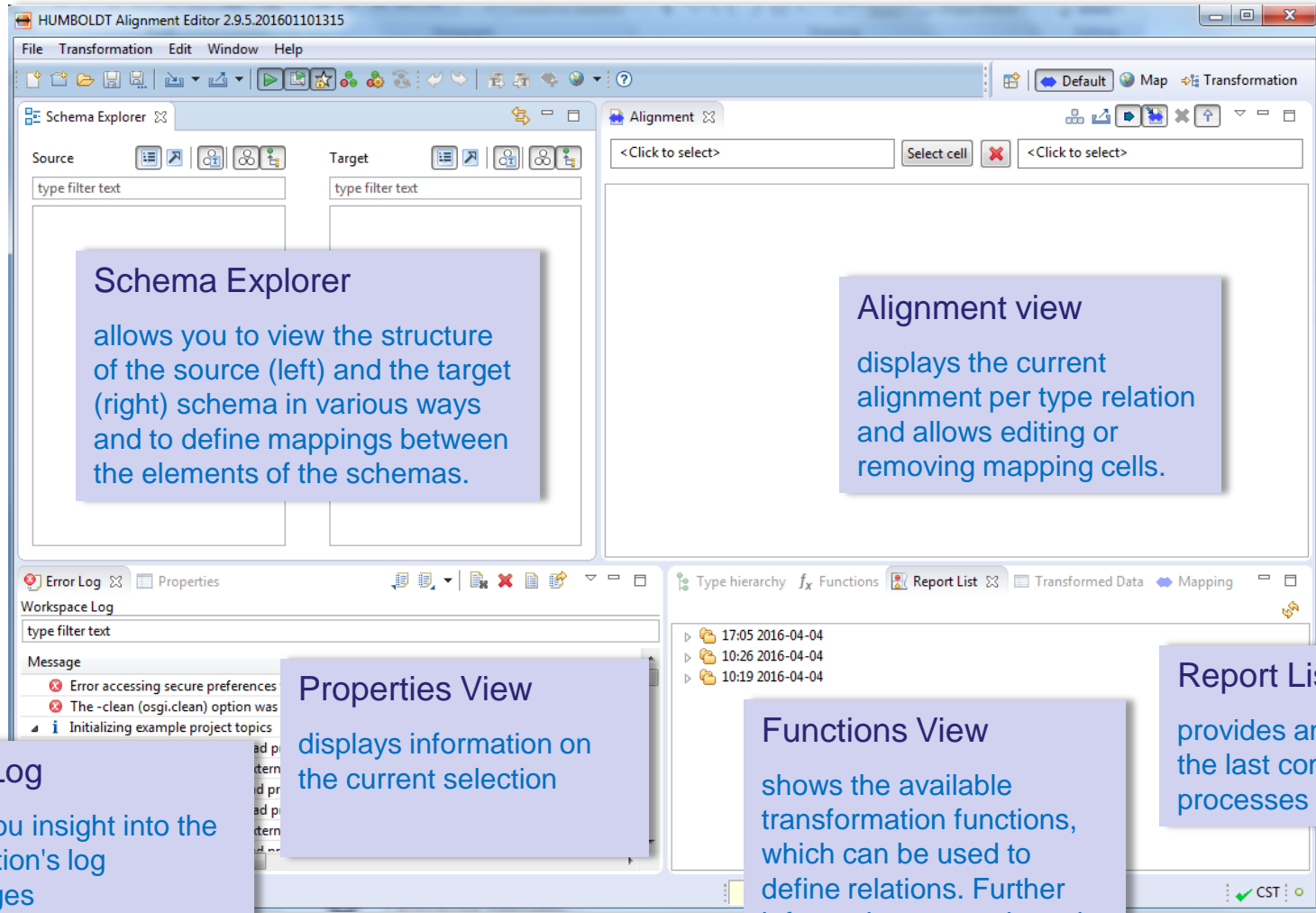


# Hale





# Hale Interface



**Schema Explorer**  
allows you to view the structure of the source (left) and the target (right) schema in various ways and to define mappings between the elements of the schemas.

**Alignment view**  
displays the current alignment per type relation and allows editing or removing mapping cells.

**Error Log**  
gives you insight into the application's log messages

**Properties View**  
displays information on the current selection

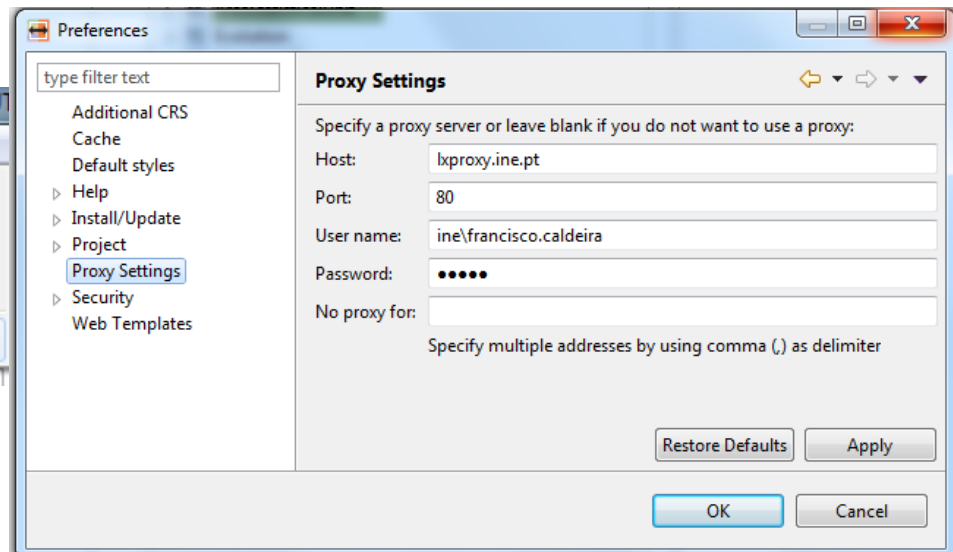
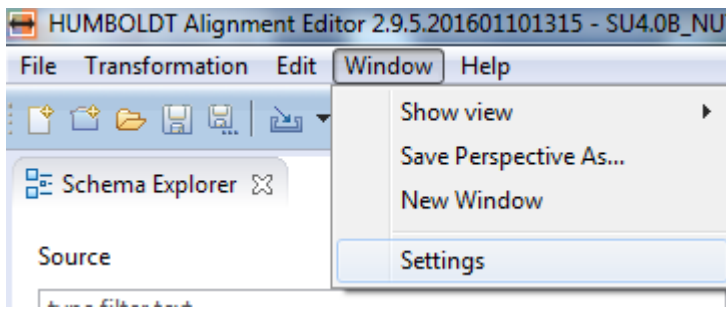
**Functions View**  
shows the available transformation functions, which can be used to define relations. Further information on a selected function will be displayed in the *Properties* view.

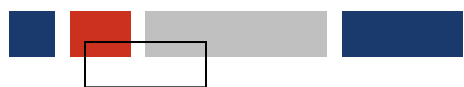
**Report List**  
provides an overview of the last completed processes



# Proxy Server

- A versão 2.9.4 de **Hale** parece não funcionar bem com proxy
- Fundamental para acesso a recursos na Web (**Codelists**, ...)
- Como definir *proxy*?





# Hale - Workflow

*Workflow*  
genérico para  
transformar  
CDG de  
acordo com  
os requisitos  
do target  
Schema

1.Importar Source/TargetSchemas

2.Importar dados

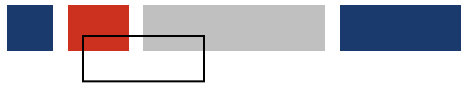
3.Definir mapping rules

4.Exportar dados transformados

5.Validar dados transformados







# Importar Source Schema

Workflow genérico para transformar CDG de acordo com os requisitos do target Schema

1. Importar Source/Target Schemas

## ▪ **Source Schema**

- Define a estrutura dos dados que desejamos transformar
- Várias fontes para dados, incluindo *online* (URL, WFS) , base de dados (PostgreSQL/PostGIS, SpatiaLite)

Shapefile (\*.shp)  
XML schema (\*.xsd, \*.xml)  
CSV file (\*.csv)  
MS OOXML Format Spreadsheet (XLSX) (\*.xlsx)  
SpatiaLite Database (\*.sqlite)  
HALE Schema Definition (\*.hsd, \*.haleschema)  
GZipped HALE Schema Definition (\*.hsd.gz, \*.haleschema.gz)  
Excel Spreadsheet (XLS) (\*.xls)



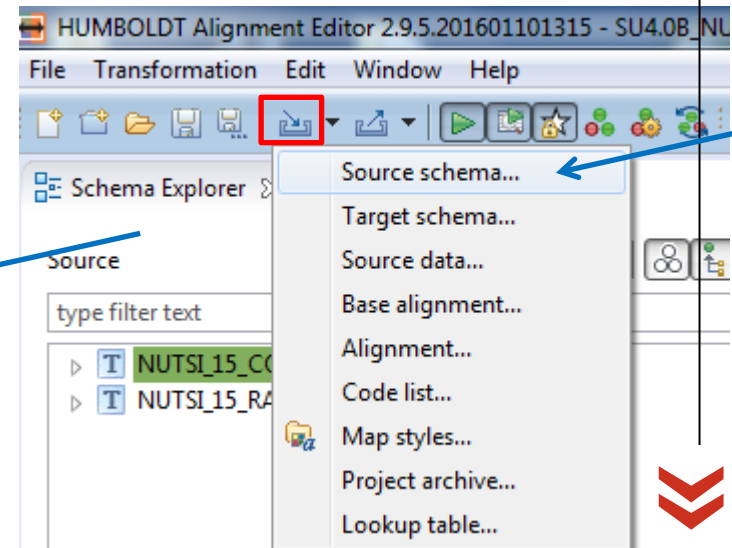
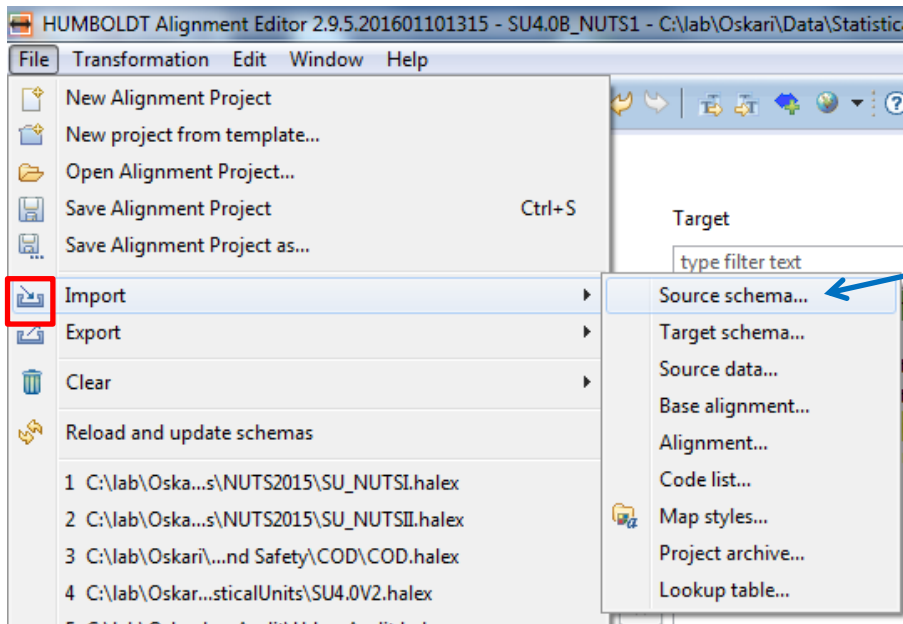
# Importar Source Schema

Workflow genérico para transformar CDG de acordo com os requisitos do target Schema

1. Importar Source/Target Schemas

## ■ **Source Schema**

### ■ Várias formas de invocar a funcionalidade

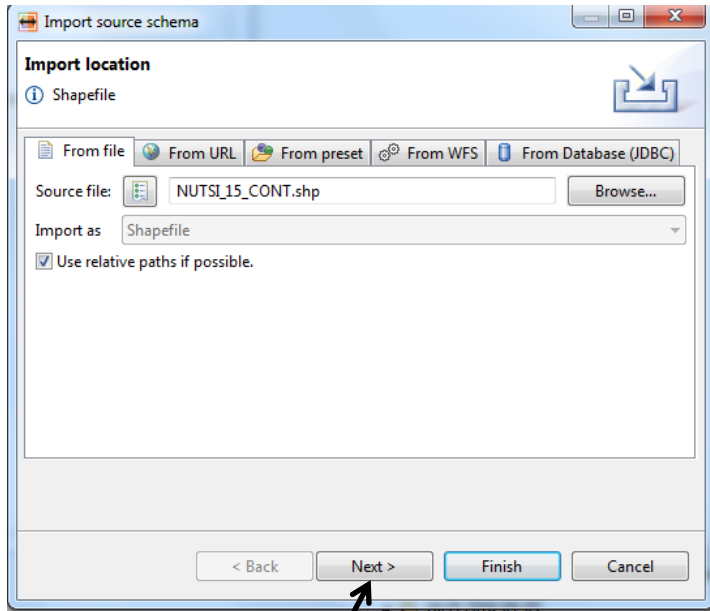


# Importar Source Schema

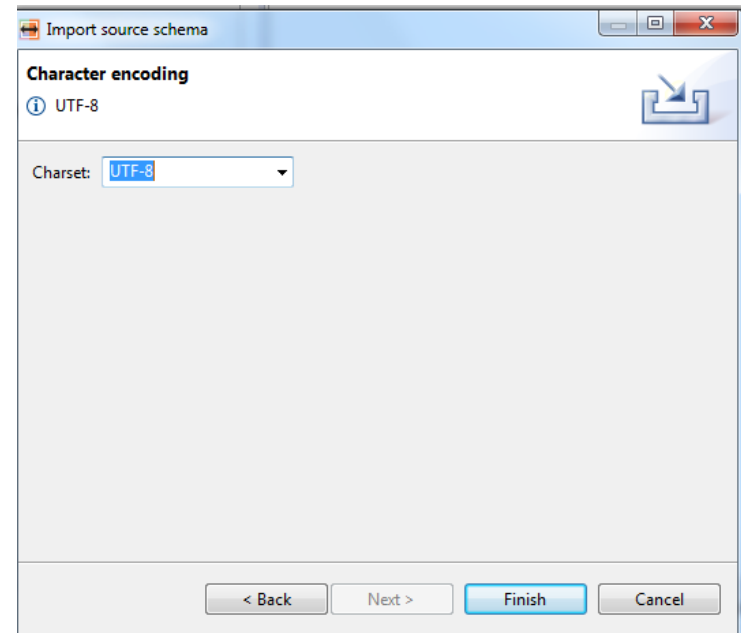
Workflow genérico para transformar ODG de acordo com os requisitos do target Schema

1. Importar Source/Target Schemas

## ■ importar Source Schema



## ■ Cuidado com os caracteres *PT*



# Importar Source Schema

Workflow genérico para transformar CDG de acordo com os requisitos do target Schema

1. Importar Source/Target Schemas

## ■ importar *Source Schema* - Simbologia

HUMBOLDT Alignment Editor 2.9.5.201601101315\*

File Transformation Edit Window Help

Schema Explorer

Source Target

type filter text type filter text

BGR11 (0..1)  
DTMN11 (0..1)  
filename  
FR11 (0..1)  
LUG11 (0..1)  
LUG11DESIG (0..1)  
N\_ALOJ (0..1)  
OBJECTID (0..1)  
SEC11 (0..1)  
SHAPE\_AREA (0..1)  
SHAPE\_LEN (0..1)  
SS11 (0..1)  
the\_geom (0..1)

String

Numérico

Geometria

### Properties

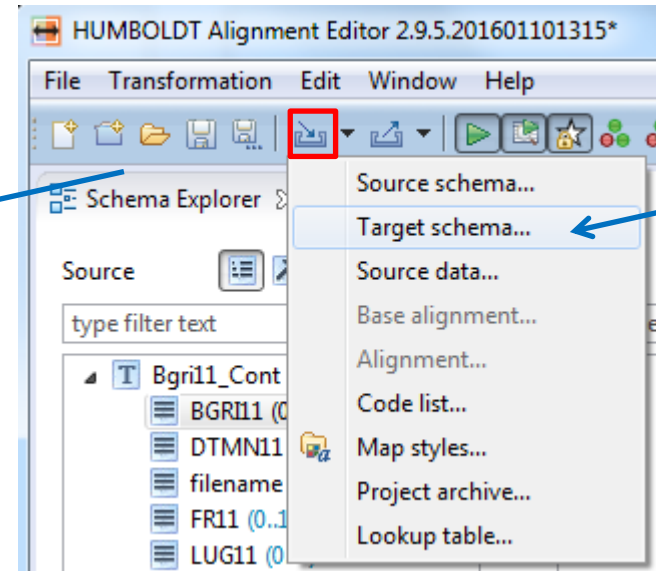
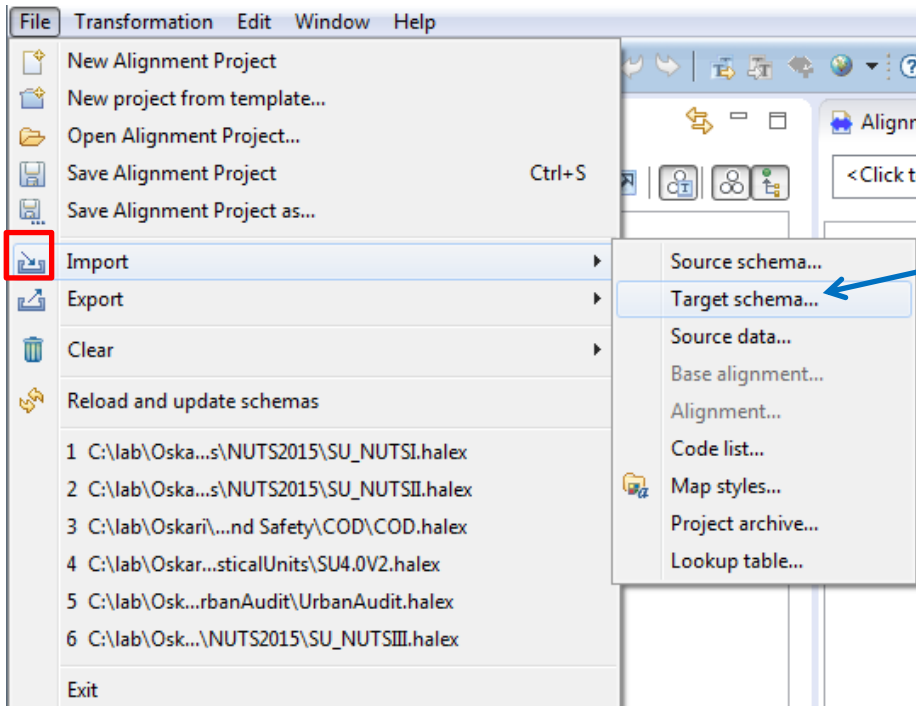
- String property
- Numeric property
- Geometry property
- Other (complex) property



# Importar target Schema

## Target Schema

- Define a estrutura para a qual desejamos transformar os dados
- Schemas definidos nas **Data Specifications Inspire**



# Importar Target Schema

## Target Schema

- Várias fontes, semelhante, ao **Source Schema** (File, Url, From Preset, WFS, BD)

The image shows a software interface for importing a target schema. The main window is titled "Import target schema" and has a section for "Import location" with options: "From file", "From URL", "From preset", and "From WFS". Below these options is a "Select preset:" field and an "Import as:" field. At the bottom are "Back" and "Next" buttons. A "Select a schema" dialog box is open in the foreground, displaying a list of schemas under the heading "Statistical":

- INSPIRE Statistical Units Base 3.0
- INSPIRE Statistical Units Base 4.0
- INSPIRE Statistical Units Grid 3.0
- INSPIRE Statistical Units Grid 4.0
- INSPIRE Statistical Units Vector 3.0
- INSPIRE Statistical Units Vector 4.0

The "INSPIRE Statistical Units Vector 4.0" item is highlighted. A yellow sticky note with the word "Tip!" is pinned to the dialog. A blue arrow points from the sticky note to the dialog, and another blue arrow points from the dialog to the "Import target schema" window. At the bottom of the dialog are "OK", "Cancel", and "None" buttons. The background window also shows a list of schemas, including "Social Governmental Services 3.0", "Social Governmental Services 4.0", "0", "0 (Bundled)", "0", "culture Facilities Model 3.0", "culture Facilities Model 4.0", "3.0", "3.0 (Bundled)", "4.0", and "Restriction and Regulation Zones". At the bottom of this window are "Cancel" and "None" buttons.



# Importar target Schema

Workflow genérico para transformar CDG de acordo com os requisitos do target Schema

1. Importar Source/Target Schemas

## ■ Target Schema

The screenshot displays the HUMBOLDT Alignment Editor interface. The 'Target' pane on the right lists the following schema elements:

- AreaStatisticalUnit
- Evolution
- GeometryDescriptor
- StatisticalTessellation
- VectorStatisticalUnit
- VectorStatisticalUnitGeometry

The 'Workspace Log' at the bottom shows the following messages:

- 11:55 2016-04-07 XML schema import 14:32.48
- 17:05 2016-04-04 Shapefile import 11:55.24
- 10:26 2016-04-04
- 10:19 2016-04-04

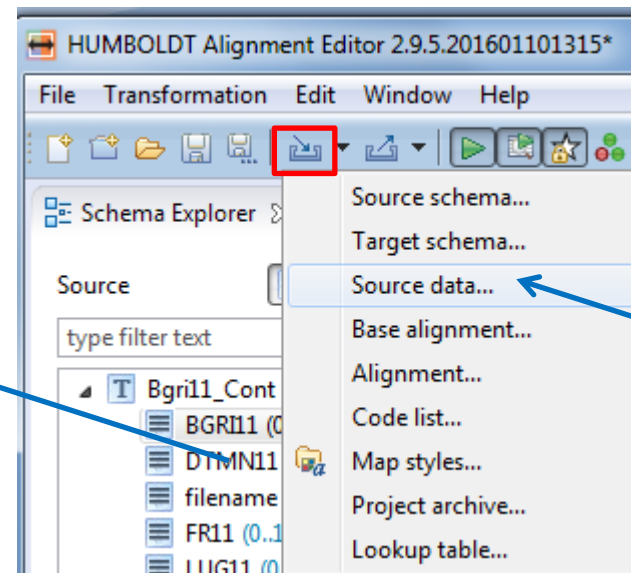
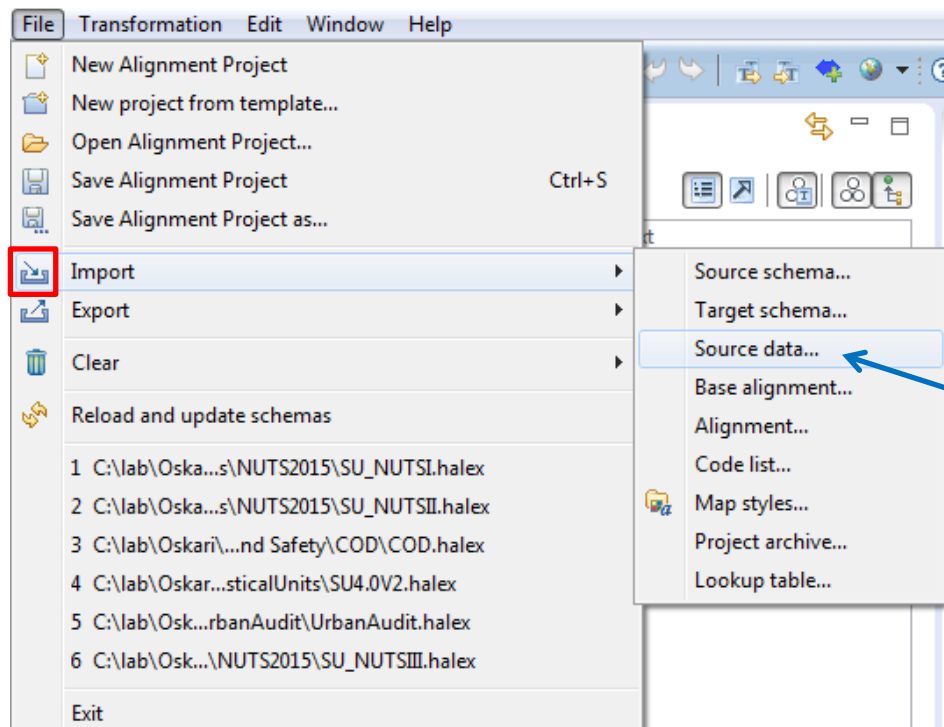
Blue arrows point from the 'VectorStatisticalUnitGeometry' entry in the Target pane to the 'XML schema import' entry in the Workspace Log.



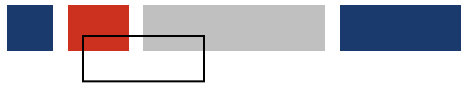
# Importar Source data

## ■ Importar *Source Data*

- Importar o **source Schema** não importa as *features* (polígonos, linhas ou pontos).





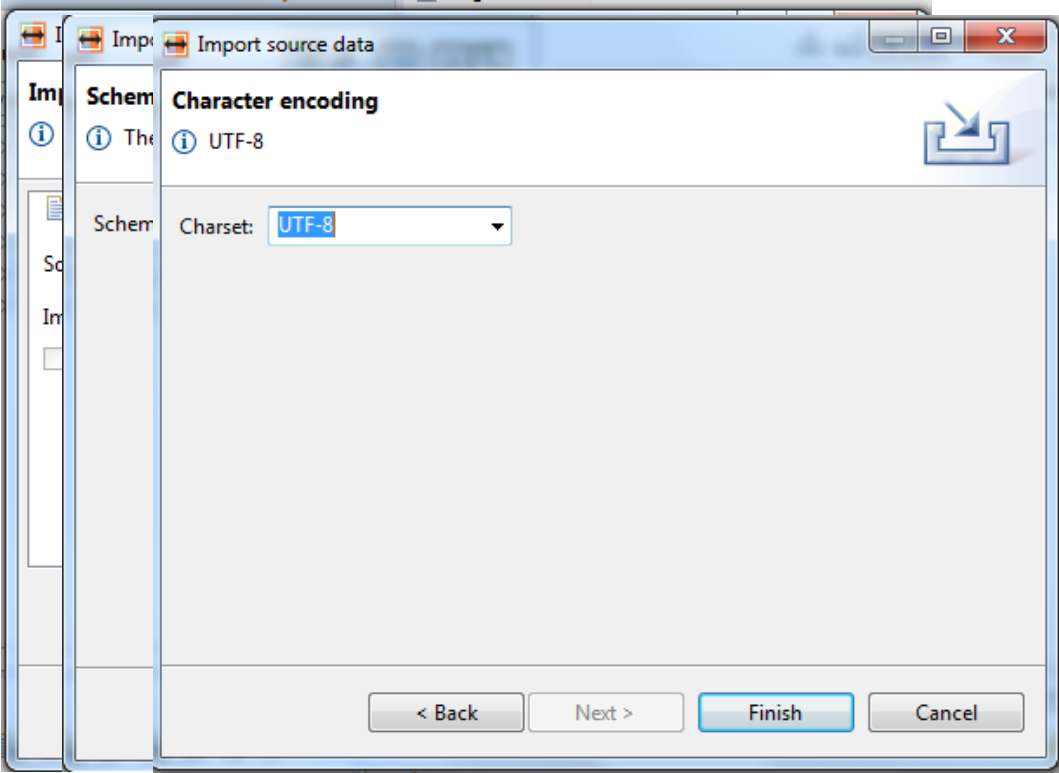


# Importar Source data

Workflow genérico para transformar CDG de acordo com os requisitos do target Schema

- 1. Importar dados
- 2. Importar dados
- 3. Importar dados
- 4. Importar dados

- **Importar Source Data**



- Cuidado com os caracteres **PT**



# Importar Source data

## ■ Importar Source Data



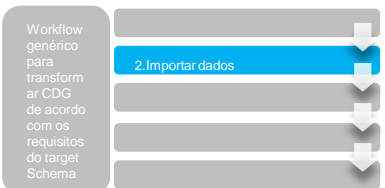
The screenshot shows the HUMBOLDT Alignment Editor interface. The 'Source' pane on the left lists various tables and columns, including 'AREA (0..1) x4470'. The 'Target' pane on the right lists a different set of tables, including 'AreaStatisticalUnit'. A blue arrow points from the 'AREA (0..1) x4470' entry in the Source pane to the 'AreaStatisticalUnit' entry in the Target pane. Below the panes, the 'Workspace Log' shows a message: 'i [14:57:48] Created cached schema representation'. At the bottom, a 'Type...' pane displays a task log with the following entries:

| Task                      | Time     |
|---------------------------|----------|
| 11:55 2016-04-07          |          |
| ✓ Load data into database | 15:11.58 |
| ✓ Shapefile import        | 15:11.58 |
| ✓ Shapefile import        | 14:57.47 |
| ✓ XML schema import       | 14:32.48 |
| ✓ Shapefile import        | 11:55.24 |
| 17:05 2016-04-04          |          |
| 10:26 2016-04-04          |          |
| 10:19 2016-04-04          |          |

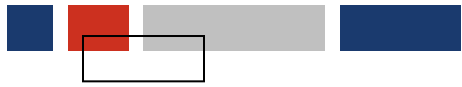


# Importar Source data

- Data perspective



| Bgr11_AC26 | 1                       | 2                      | 3                      |
|------------|-------------------------|------------------------|------------------------|
| Bgr11_AC26 | +                       | +                      | +                      |
| AREA       | 0.0                     | 0.0                    | 0.0                    |
| BGRI11     | 42051400101             | 42030400205            | 42030400403            |
| DTMN11     | 4205                    | 4203                   | 4203                   |
| filename   | Bgr11_AC26              | Bgr11_AC26             | Bgr11_AC26             |
| FR11       | 14                      | 04                     | 04                     |
| LEN        | 0.0                     | 0.0                    | 0.0                    |
| LUG11      | 027246                  | 027185                 | 027185                 |
| LUG11DESIG | São Brás                | Capelas                | Capelas                |
| N_ALOJ     | 16                      | 0                      | 78                     |
| OBJECTID   | 36174                   | 36175                  | 36176                  |
| SEC11      | 001                     | 002                    | 004                    |
| SHAPE_AREA | 310286.019904           | 7039.7060865           | 665119.270372          |
| SHAPE_LEN  | 3291.09965573           | 350.115823455          | 3838.44559463          |
| SS11       | 01                      | 05                     | 03                     |
| the_geom   | {CRS=ITRF93_UTM_Zone    | {CRS=ITRF93_UTM_Zone   | {CRS=ITRF93_UTM_Zone   |
| Metadata   | +                       | +                      | +                      |
| Identifier | 9b06e996-ee7c-467b-9ef1 | 124b5166-0ccb-401d-af9 | a56a3bd3-e2e8-444d-bb6 |



# Importar Source data

Workflow genérico para transformar ar CDG de acordo com os requisitos do target Schema

- 2.Importar dados

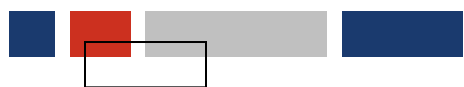
## Map perspective

The screenshot displays the HUMBOLDT Alignment Editor interface. The main map area is split into two views: 'Source data' on the left and 'Transformed data' on the right. The 'Source data' view shows a pink-shaded polygon representing the project area. The 'Transformed data' view shows the same area with a road alignment (ER 1-1A) overlaid. The interface includes a menu bar (File, Transformation, Edit, Map, Window, Help), a toolbar, and a status bar. A 'Transformation' panel on the right shows a list of operations with their completion times.

| Operation                 | Time     |
|---------------------------|----------|
| 15:57 2016-04-07          |          |
| ✓ Load data into database | 15:57:26 |
| ✓ Shapefile import        | 15:57:26 |
| ✓ Shapefile import        | 15:57:24 |
| ✓ XML schema import       | 15:57:23 |
| ✓ HALE XML project import | 15:57:14 |
| 11:55 2016-04-07          |          |
| 16:39 2016-04-05          |          |
| 17:05 2016-04-04          |          |

Source Data

| Property    | Value       |
|-------------|-------------|
| Bgri11_AC26 | 1           |
| Bgri11_AC26 | +           |
| AREA        | 0.0         |
| BGR11       | 42050300211 |
| DTMN11      | 4205        |
| filename    | Bgri11_AC26 |
| FR11        | 03          |



# Importar Source data



## ■ Importar *Source Data*

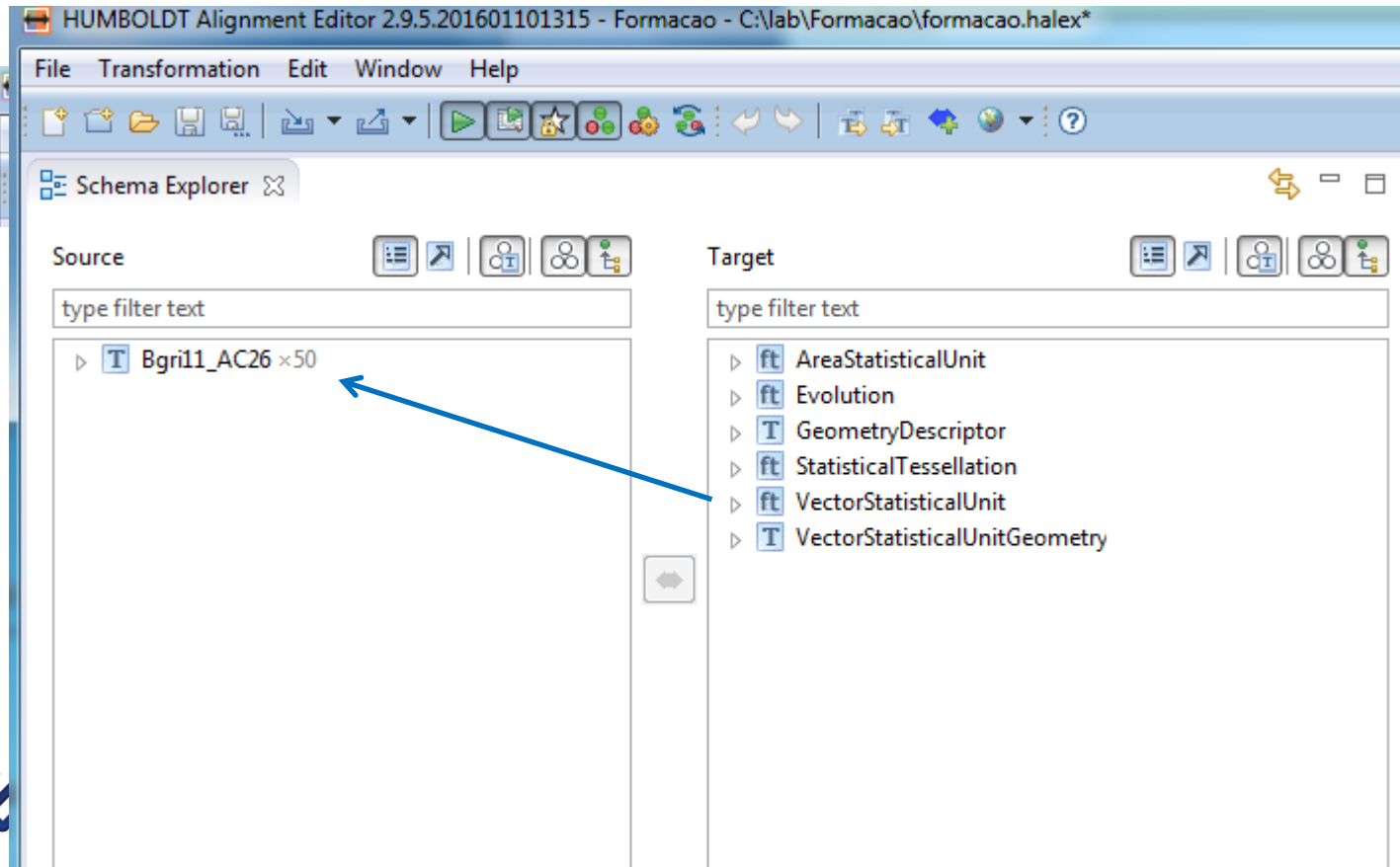
- Importar cdg's com um número elevado de *features* pode trazer problemas de performance porque o **HALE** transforma os dados e valida, por defeito, sempre que existe mapeamento de algum campo.

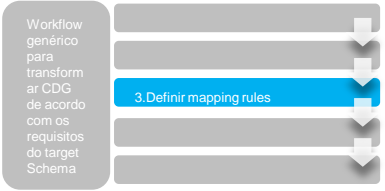


# Importar Source data

- É possível definir uma seleção de *features* para trabalhar com menos registos.

**Tip!**





# Definir Mapping rules

▪ No **Target Schema** é necessário identificar **tipo** relevante para o cdg que pretendemos harmonizar.

```
▶ ft AreaStatisticalUnit
▶ ft Evolution
▶ T GeometryDescriptor
▶ ft StatisticalTessellation
▶ ft VectorStatisticalUnit
▶ T VectorStatisticalUnitGeometry
```

▪ As data specifications muitas vezes indicam claramente o tipo

▪ The application schema on statistical units is composed of different packages:

- Base: The base package.
- Grid: Classes to represent statistical grids.
- Vector: Classes to represent statistical unit having vector geometries (point, line, surfaces).



# Definir Mapping rules

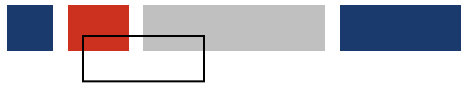
Workflow genérico para transformar ar ODG de acordo com os requisitos do target

- 1. Definir o schema de origem
- 2. Definir o schema de destino
- 3. Definir mapping rules**
- 4. Executar a transformação

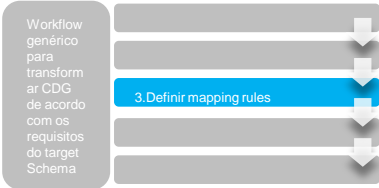
The screenshot shows the HUMBOLDT Alignment Editor interface. The 'Source' schema on the left includes a 'type filter text' field and a tree of properties like 'Bgri11\_AC26' with sub-properties 'AREA', 'BGR11', 'DTMN11', etc. The 'Target' schema on the right includes a 'type filter text' field and a tree of properties like 'AreaStatisticalUnit', 'Evolution', 'GeometryDescriptor', 'StatisticalTessellation', 'VectorStatisticalUnit', and 'VectorStatisticalUnitGeometry'. A mapping rule is defined in the 'Alignment' panel, showing a transformation from 'Bgri11\_AC26' to 'Retype' to 'VectorStatisticalUnit'. A blue callout box points to the 'VectorStatisticalUnit' in the target schema and the 'Instance validation' section of the report, which shows warnings for 'Cardinality' in the 'country' and 'geometry' properties.

Para cada instância na data source vai criar uma instância no tipo identificado do *target* Schema





# Definir Mapping rules



Target

type filter text

- ▶ **ft** AreaStatisticalUnit
- ▶ **ft** Evolution
- ▶ **T** GeometryDescriptor
- ▶ **ft** StatisticalTessellation
- ▶ **ft** VectorStatisticalUnit ×50
  - ▶ **ft** location (0..1)
  - ▶ **8** beginLifespanVersion
  - ▶ **ft** boundedBy (0..1)
  - ▶ **ft** country
  - ▶ **ft** description (0..1)
  - ▶ **ft** descriptionReference (0..1)
  - ▶ **8** endLifespanVersion (0..1)
  - ▶ **ft** evolutions (0..n)
  - ▶ **ft** geographicalName (0..n)
  - ▶ **ft** geometry (1..n)
  - ▶ **ft** id
  - ▶ **ft** identifier (0..1)
  - ▶ **ft** inspireId
  - ▶ **ft** metaDataProperty (0..n)
  - ▶ **ft** name (0..n)
  - ▶ **ft** referencePeriod
  - ▶ **ft** thematicId (0..n)
  - ▶ **ft** validityPeriod (0..1)

## ▪ Schema elements

### Properties

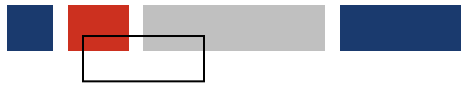
- String property
- Numeric property
- Geometry property
- Other (complex) property

### Groups

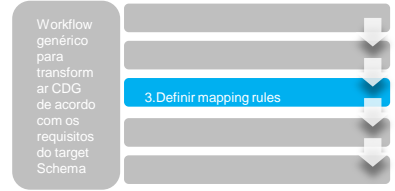
- Normal group containing a set of properties.
- Choice group, where only one of the specified properties is allowed as a child.
- A red asterisk marks properties that are mandatory, i.e. they occur exactly once and must have a value (and the value may not be null).  
*Please note that if the parent of such a property is a choice (choice icon) the choice takes precedence, i.e. only one of its children may be present in an object, but the child that is present may still not hold a null value if marked with a red asterisk.*
- A small brown error in the top left corner marks a property from a XML schema as being defined as a XML attribute.
- A property that is deemed to hold the main geometry of a type is marked with a small green triangle. Per type, you can set one property as the default geometry property. This property is then used when retrieving geometries for display in the map.

▪ Propriedades obrigatórias são facilmente identificadas no error log quando não estão preenchidas





# Definir Mapping rules



## ▪ Cardinalidade

- Para cada elemento existe informação sobre a sua cardinalidade
- O valor  $n$  ( $0..n$ ) representa um número ilimitado máximo de ocorrências
- Se não existe informação sobre a cardinalidade significa que a propriedade ocorre apenas uma vez
- Quando a cardinalidade é ( $0..1$ ) significa que podem ou não existir ocorrências, ou seja é um elemento opcional.

Target

type filter text

- ▶ ft AreaStatisticalUnit
- ▶ ft Evolution
- ▶ T GeometryDescriptor
- ▶ ft StatisticalTessellation
- ▶ ft VectorStatisticalUnit ×50
  - ▶ location (0..1)
  - ▶ beginLifespanVersion
  - ▶ boundedBy (0..1)
  - ▶ country
  - ▶ description (0..1)
  - ▶ descriptionReference (0..1)
  - ▶ endLifespanVersion (0..1)
  - ▶ evolutions (0..n) ←
  - ▶ geographicalName (0..n)
  - ▶ geometry (1..n)
  - ▶ id ←
  - ▶ identifier (0..1) ←
  - ▶ inspireId
  - ▶ metaDataProperty (0..n)
  - ▶ name (0..n)
  - ▶ referencePeriod
  - ▶ thematicId (0..n)
  - ▶ validityPeriod (0..1)



# Definir Mapping rules

Workflow genérico para transformar ar ODG de acordo com os requisitos do target Schema



HUMBOLDT Alignment Editor 2.9.5.201601101315 - Formacao - C:\lab\Formacao\formacao.halex\*

File Transformation Edit Window Help

Schema Explorer

Source

- Bgri1\_AC26
- AREA (0..1) ×50
- BGR11 (0..1) ×50
- DTMN11 (0..1) ×50
- filename ×50
- FR11 (0..1) ×50
- LEN (0..1) ×50
- LUG11 (0..1) ×50
- LUG11DESIG (0..1) ×50
- N\_ALOJ (0..1) ×50
- OBJECTID (0..1) ×50
- SEC11 (0..1) ×50
- SHAPE\_AREA (0..1) ×50
- SHAPE\_LEN (0..1) ×50
- SS11 (0..1) ×50
- the\_geom (0..1) ×50

Target

- endLifespanVersion (0..1)
- evolutions (0..n)
- geographicalName (0..n)
- geometry (1..n) ×50
  - VectorStatisticalUnitGeometry ×50
    - geometry ×50
      - AbstractGeometry (0..1) ×50
        - AbstractSolid
        - CompositeCurve
        - CompositeSolid
        - CompositeSurface
        - Curve
        - GeometricComplex
        - Grid
        - LineString
        - MultiCurve
        - MultiGeometry
        - MultiPoint
        - MultiSolid
        - MultiSurface ×50
        - OrientableCurve
        - OrientableSurface
        - Point
        - Polygon

Alignment

- Bgri1\_AC26 ×50
- Retype
- ft VectorStatisticalUnit ×50
- the\_geom
- Rename
- ...metry.MultiSurface

Error Log

Workspace Log

Message

- Instance transformation

Type hierarchy

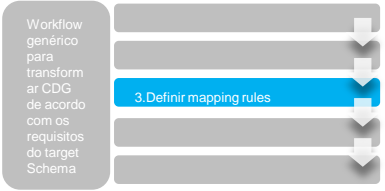
- 15:57 2016-04-07
  - Instance validation 11:40:57
  - Instance transformation 11:40:56
  - Instance validation 11:34:05
  - Instance transformation 11:34:04

Help < Back Next > Finish Cancel

Polygon

PolyhedralSurface





# Definir Mapping rules

## ▪ Geometry

**Tip!**

▪ Porque *Multisurface* e não *polygon*?

Se na existirem apenas harmonizem ignorados



Table

NUTSI\_15\_AC26

| FID | Shape * | OBJECTID | NUTS1_15 | NUTS1_15DE                 | SHAPE AREA    | SHAPE LEN     |
|-----|---------|----------|----------|----------------------------|---------------|---------------|
| 0   | Polygon | 1        | PT2      | Região Autónoma dos Açores | 2163885709,67 | 849427,366641 |

(1 out of 1 Selected)

NUTSI\_15\_AC26



# Definir Mapping rules

## Map View

- É possível ver a **transform data**, geometria e atributos

The screenshot shows the HUMBOLDT Alignment Editor interface. The main window is split into two map views: 'Source data' on the left and 'Transformed data' on the right. The 'Source data' map shows a road network with labels like 'ER 8-1A' and 'EN 1-1A'. The 'Transformed data' map shows the same area with a pink polygon highlighting a specific feature. Below the maps, there are two data tables. The 'Source Data' table shows attributes for 'Bgnr11\_AC26'. The 'Transformed Data' table shows attributes for 'VectorStatisticalUnit'.

| Source Data |             |
|-------------|-------------|
| Bgnr11_AC26 | 1           |
| Bgnr11_AC26 | +           |
| AREA        | 0.0         |
| BGR111      | 42051100110 |
| DTMN11      | 4205        |

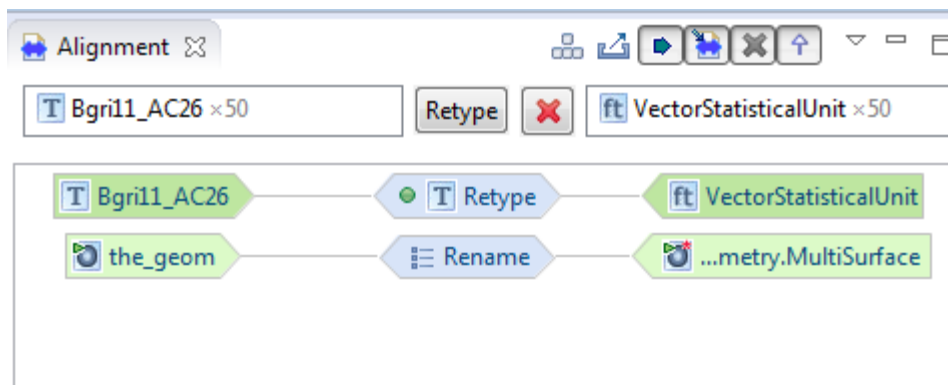
| Transformed Data      |          |
|-----------------------|----------|
| VectorStatisticalUnit | 1        |
| VectorStatisticalUnit |          |
| location              | no value |
| beginLifespanVersion  | no value |
| boundedBy             | no value |

# Definir Mapping rules

## Geoserver

### ▪ Função *rename*

- Permite copiar uma propriedade qualquer do **Source Schema** para o **target Schema**
- Foi utilizada na geometria, mas pode ser utilizada em qualquer propriedade



- ☆ Groovy script
- ☆ Groovy script (greedy)
- ☰ Regex Analysis
- ☰ Rename**
- ☰ Classification
- abc Formatted string
- ➡ Assign (Bound)
- Augmentations
- id Generate sequential ID
- ➡ Assign
- ➡ Generate Unique Id





# Definir Mapping rules

Workflow genérico para transformar o CDG de acordo com os requisitos do target Schema

- 1. Definir o CDG
- 2. Definir o target Schema
- 3. Definir mapping rules**
- 4. Definir o output

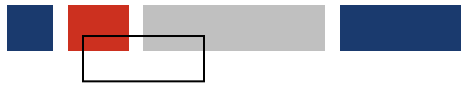
## Geoserver

### ▪ Função Assign

- Permite atribuir um valor a uma propriedade do **target Schema**, que não exista no **Source Schema**

The screenshot shows a context menu for a property named 'id' in a 'geometryDescriptor'. The menu items are: Groovy script, Groovy script (greedy), Mathematical Expression, Date extraction, Regex Analysis, Rename, Classification, Formatted string, Assign (Bound), Augmentations, **id** Generate sequential ID, **Assign**, and Generate Unique Id. A blue arrow points to the 'Assign' option.





# Definir Mapping rules

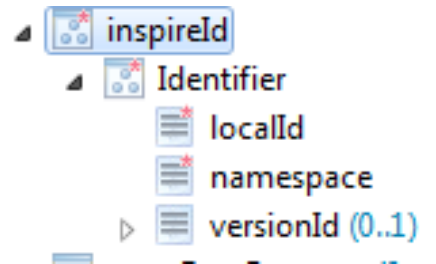
Workflow genérico para transformar o ODG de acordo com os requisitos do target Schema

- 1. Definir o ODG
- 2. Definir o target Schema
- 3. Definir mapping rules**
- 4. Executar o workflow

## Geoserver

### ▪ Função *Assign*

- Exemplo da função *Assign* com campo *InspireId*
- *InspireId* é um campo complexo constituído por 3 campos *LocalId*, *NameSpace* e *VersionId*, em que os dois primeiros são obrigatórios.





# Definir Mapping rules

## ▪ Função Assign

Geoserver

- O atributo **NameSpace** não existe na **Source data** deve ter o valor <http://id.igeo.pt/so/SU/VectorStatisticalUnit>

## ▪ On a

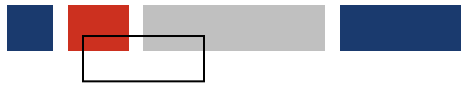
- Source
- type filter text
- Bgri11
- AI
- BC
- D
- fil
- FF
- LE
- LL
- LL
- N
- OI
- SE
- SH
- SH
- SS
- th

alld}/{ver

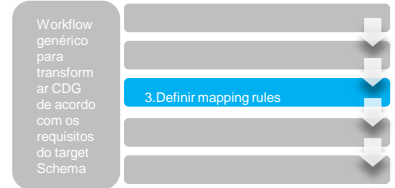
Inspire}

geográfico





# Definir Mapping rules



## ▪ InspireId

### ▪ Os restantes campos:

▪ **LocalId rename** de um campo com código único ou designação do recurso original + “\_” + código do objecto geográfico

▪-- Definition --

▪ A local identifier, assigned by the data provider. The local identifier is unique within the namespace, that is no other spatial object carries the same unique identifier.

▪-- Description --

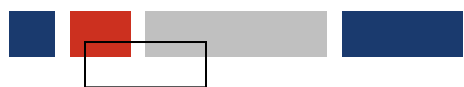
▪ NOTE It is the responsibility of the data provider to guarantee uniqueness of the local identifier within the namespace.

▪ **VersionId assign** de um valor da versão

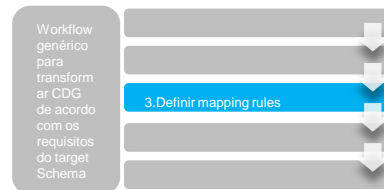
▪-- Definition --

▪ The identifier of the particular version of the spatial object, with a maximum length of **25 characters**. If the specification of a spatial object type with an external object identifier includes life-cycle information, the version identifier is used to distinguish between the different versions of a spatial object. Within the set of all versions of a spatial object, the version identifier is unique.





# Definir Mapping rules



## ▪ **Codelists**



- São listas de códigos e seus valores, conforme definido nas regras de implementação **INSPIRE** relativas à interoperabilidade dos conjuntos e serviços de dados espaciais (Regulamento (UE) no 1089/2010 da Comissão).



# Definir Mapping rules

## ■ **Codelists**

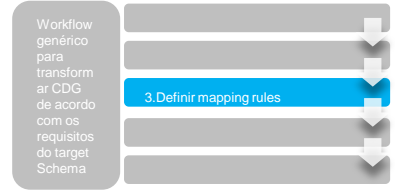


■ No **target Schema** esta estrutura de elementos está associada a **Codelist**

■ O campo **Href** irá receber os valores



# Definir Mapping rules



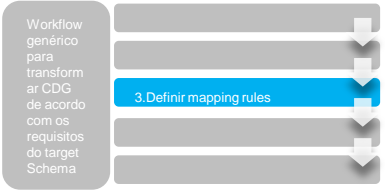
## ■ Como importar **Codelists**

Clicar

| Task                     | Time     |
|--------------------------|----------|
| INSPIRE code list import | 16:35.56 |
| Instance validation      | 16:16.38 |
| Instance transformation  | 16:16.38 |
| Load data into database  | 16:16.29 |
| Shapefile import         | 16:16.29 |
| Shapefile import         | 16:16.29 |
| XML schema import        | 16:16.29 |

Clicar      Clicar

# Definir Mapping rules



**Alignment**

Source: T Bgri11\_AC26 ×50, the\_geom

Target: ft VectorStatisticalUnit ×50

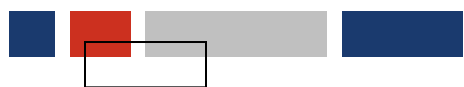
Mapping Rules:

- T Bgri11\_AC26 → Retype → ft VectorStatisticalUnit
- the\_geom → Rename → ...metry.MultiSurface
- the\_geom → Assign → ...geometryType.href
- the\_geom → Assign → ...entifier.namespace

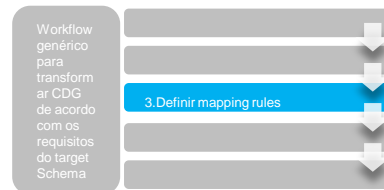
A blue arrow points to the bottom 'Assign' rule.

clicar  
clicar





## Definir Mapping rules



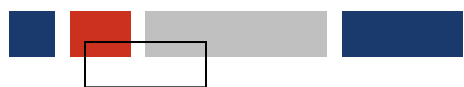
### ▪ ***Codelists Externas***

▪ Existem codelist mantidas por organizações externas ao Inspire

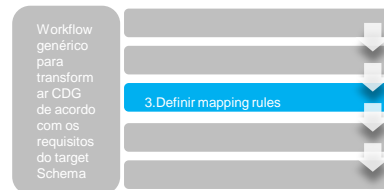
▪ Causas de morte Eurostat

▪ Patologia World Health Organization WHO





## Definir Mapping rules



### ▪ ***Codelist ainda sem valores***

▪ Existem algumas *codelist* sem códigos definidos, por vezes completamente vazias

▪ *INSPIRE Registry*:

▪ Padrão em estudo

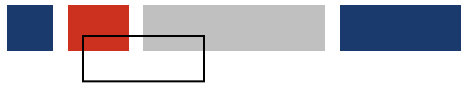
▪ 1. <http://registo.igeo.pt/codelist/{NomeListaCodigos}Value>

▪ 2. <http://registo.igeo.pt/codelist/{NomeListaCodigos}Value/{ValorCodigo}>

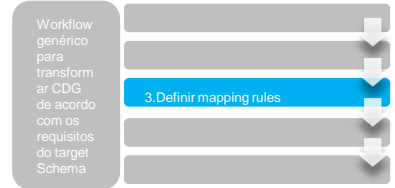
▪ <http://registo.igeo.pt/codelist/AggregationUnitValue/NUTSIII>







# Definir Mapping rules

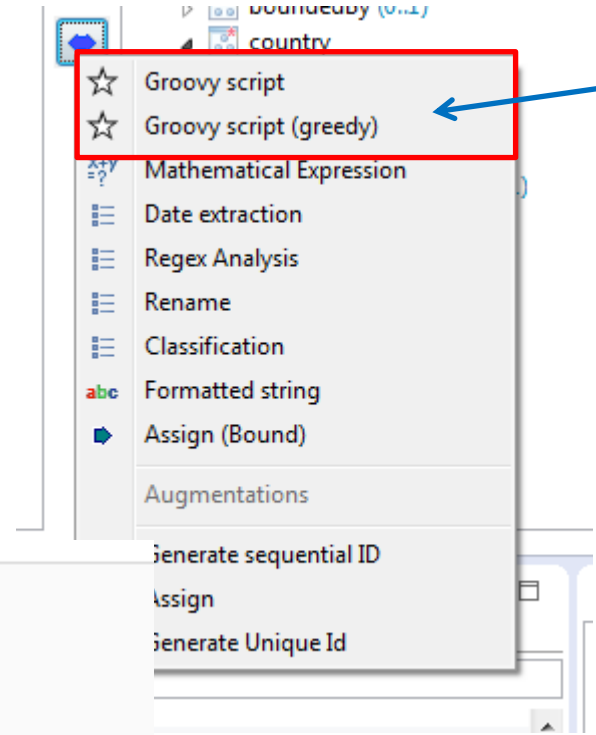


## Geoserver

### Outras funções

#### Groovy Script

- Permite o cálculo através de código de um valor de uma propriedade no *target schema*

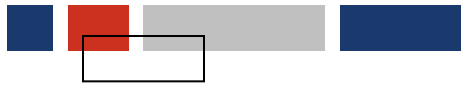


```
// retrieve all values of "result" in the merged object
def list = _source.p.result.values()

// concatenate the values, separated by comma
def str = list.join(',')

// assign the string to the property "resultString" in the target object
_target {
  resultString(str)
}
```





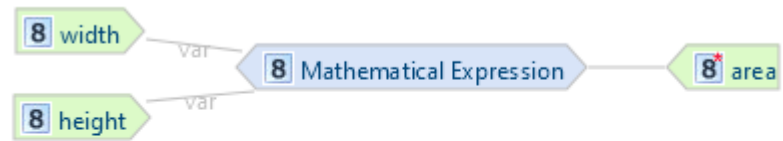
# Definir Mapping rules

## Geoserver

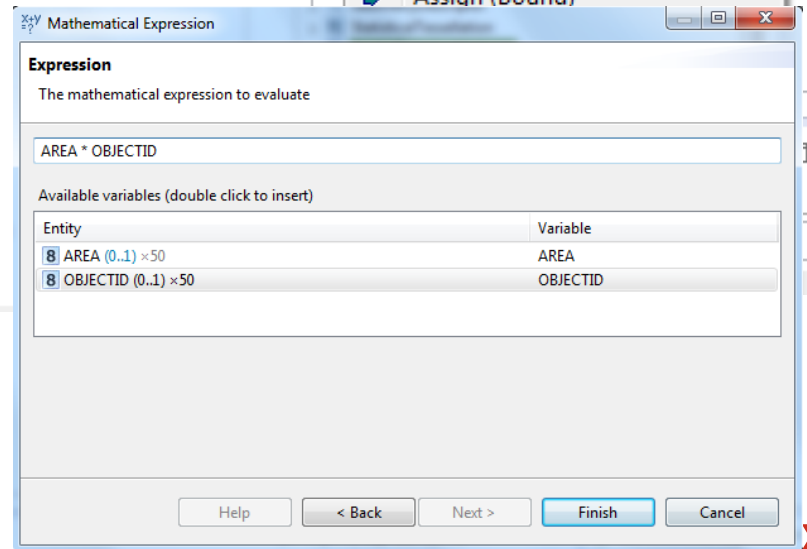
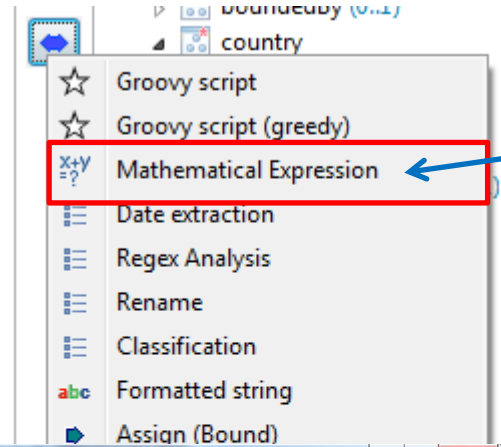
### Outras funções

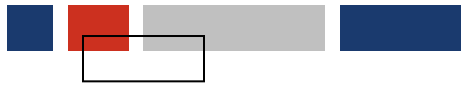
#### Mathematical Expression

- Define um valor através de uma expressão matemática
- Pode utilizar variáveis

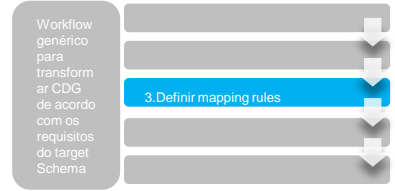


| Parameter  | Value          |
|------------|----------------|
| expression | width * height |





# Definir Mapping rules

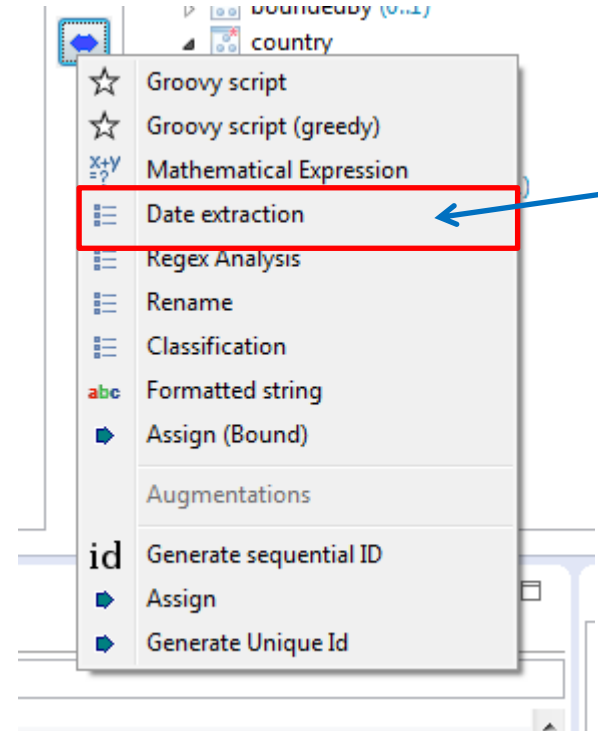


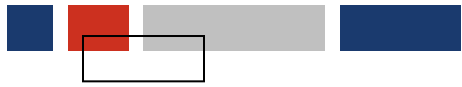
## Geoserver

### ■ Outras funções

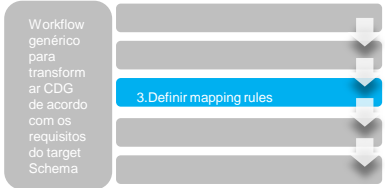
#### ■ *Date Extraction*

- Extrai uma data de uma string
- **yyyy-MM-dd HH:mm:ss**





# Definir Mapping rules

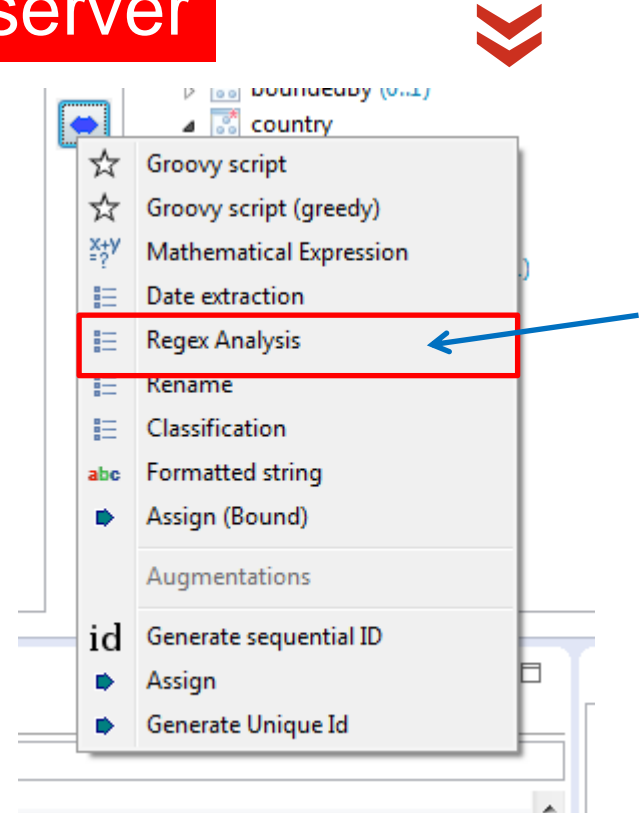


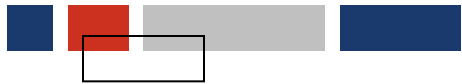
## Geoserver

- Outras funções

- **Regex Analysis**

- Aplica uma *regular expression* a uma *string*





# Definir Mapping rules

Workflow genérico para transformar CDG de acordo com os requisitos do target Schema

- 1. Definir regras de mapeamento
- 2. Definir regras de transformação
- 3. Definir mapping rules**
- 4. Definir regras de validação
- 5. Definir regras de logging

Let us assume that the source property represents a date of the format:

**20081209**

and that the target property needs a date in the format :

**YYYY-MM-DD HH:MM:SS**

We can use a regular expression like:

**([0-9]{4})([0-9]{2})([0-9]{2})**

and an output format:

**{1}-{2}-{3} 00:00:00**

to achieve our goal.

In fact the result of the above example will result in:

**2008-12-09 00:00:00**

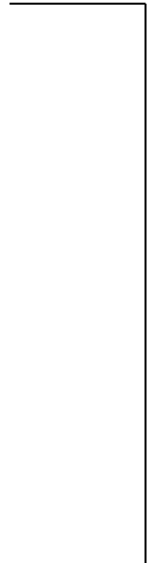
### *Explained*

The regular expression used defines 3 groups, which are separated by the round brackets:

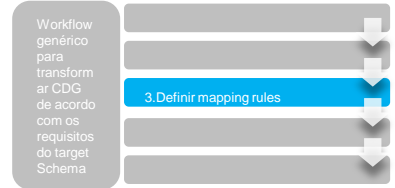
| Regex Group       | Explained                         | Catched part of <b>20081209</b> |
|-------------------|-----------------------------------|---------------------------------|
| <b>([0-9]{4})</b> | catches 4 numbers between 0 and 9 | 2008                            |
| <b>([0-9]{2})</b> | catches 2 numbers between 0 and 9 | 12                              |
| <b>([0-9]{2})</b> | catches 2 numbers between 0 and 9 | 09                              |

The output can then be formatted concatenating groups with any string.

The format **{1}-{2}-{3} 00:00:00** simply defines that the first three groups will be concatenated by a dash and that a default time will be added at the end of the resulting string.

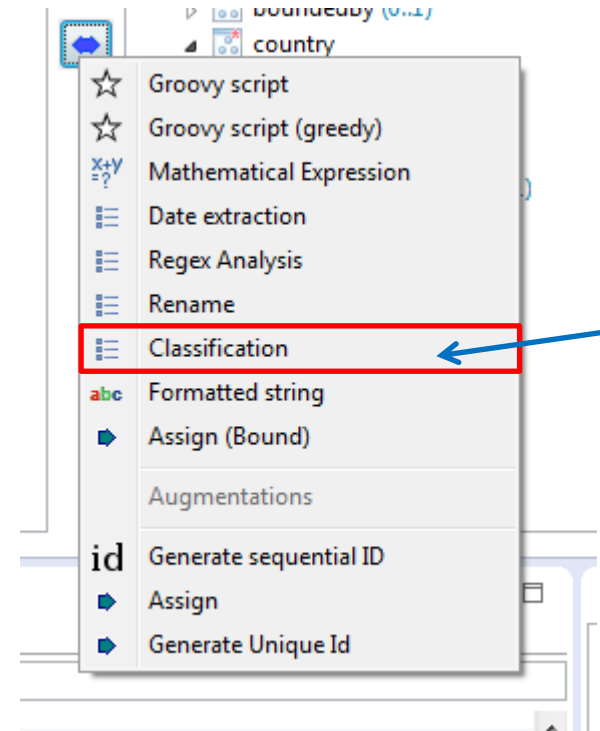
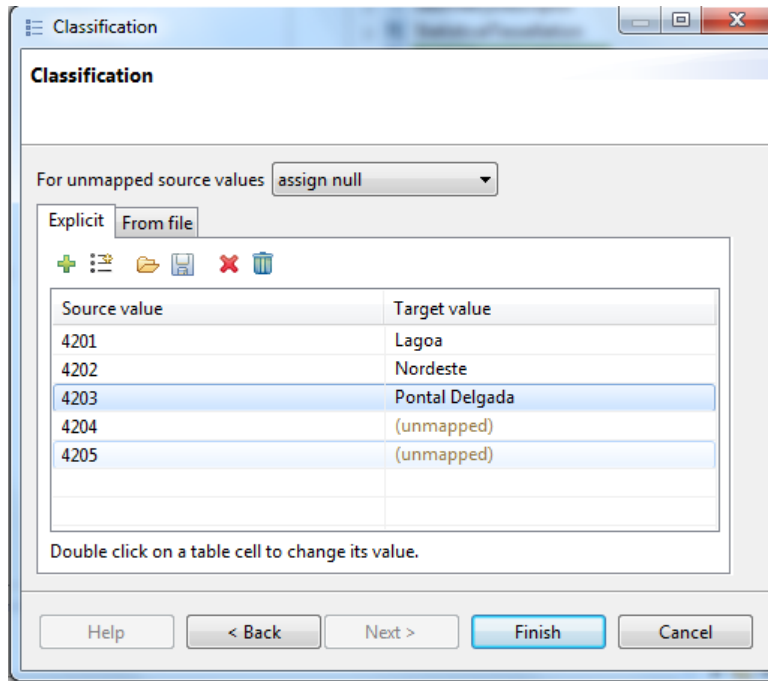


# Definir Mapping rules



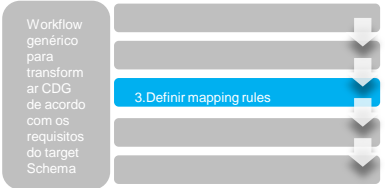
Geoserver

- Outras funções
  - Classification
    - Reclassifica valores





# Definir Mapping rules



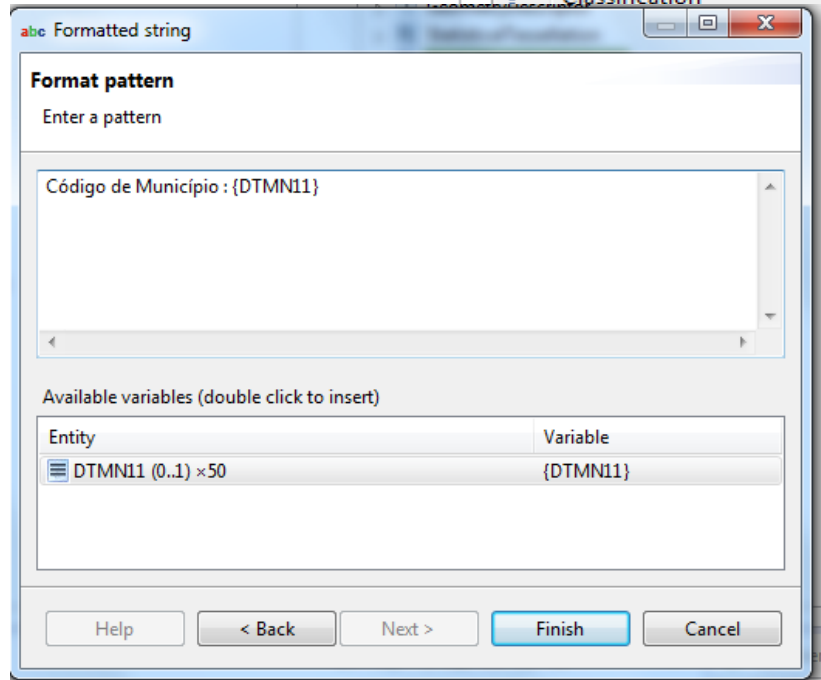
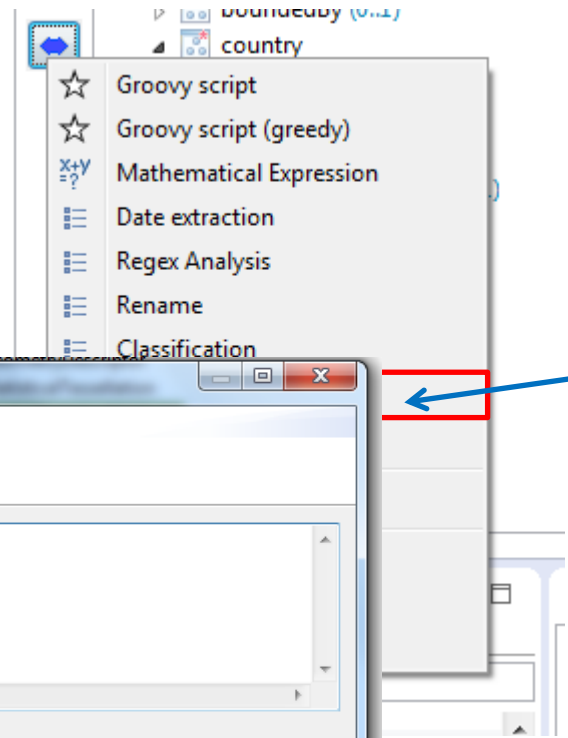
▪ Outras funções

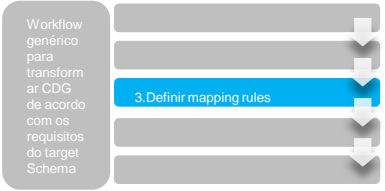
Geoserver



## ▪ *Formatted String*

- Semelhante ao Mathematical Expression mas com Strings
- Cria uma string através de um padrão e variáveis





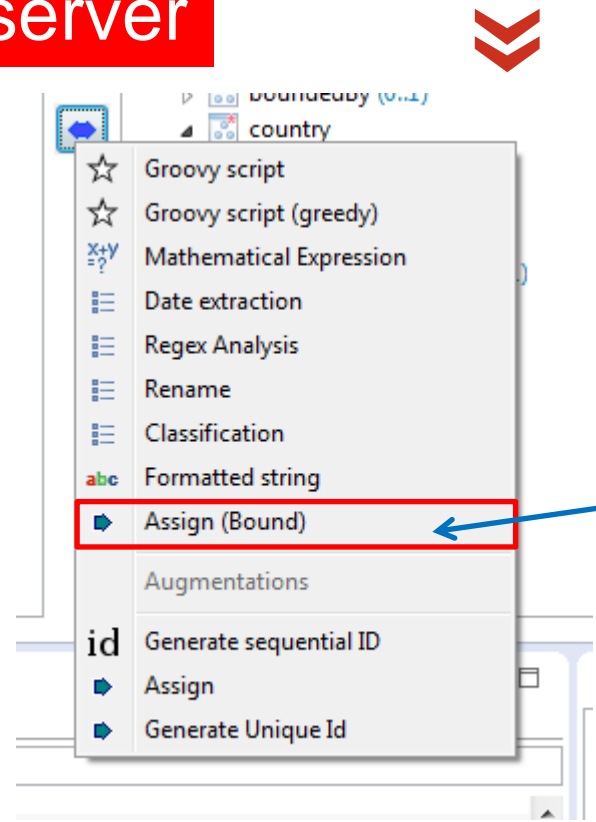
# Definir Mapping rules

▪ Outras funções

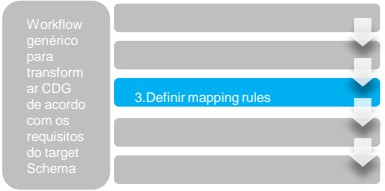
Geoserver

## ▪ *Assign Bound*

▪ Atribui um valor a uma propriedade (*target*) se uma propriedade na fonte estiver presente.







# Definir Mapping rules

▪ Outras funções

**Geoserver**

▪ **Generate Sequential ID** e **Generate UniqueID**

▪ Cria um identificador sequencial numérico



id Generate sequential ID

**Generate sequential ID**

Please configure the identifier generation

Sequence: Per target instance type

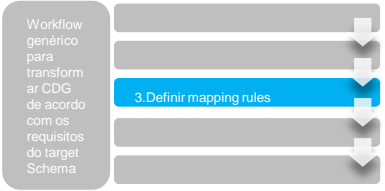
Prefix: Id

Suffix: INE

Example: Id1INE

Buttons: Help, < Back, Next >, Finish, Cancel



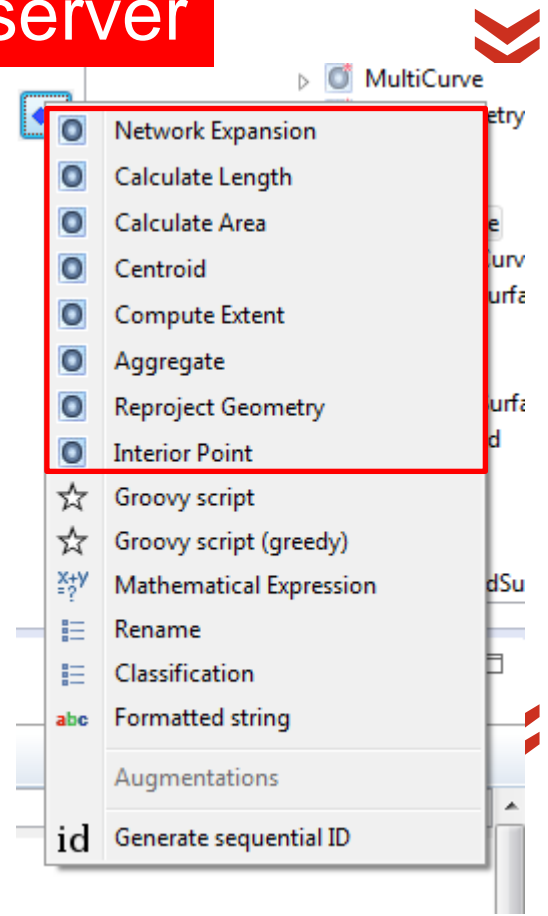


# Definir Mapping rules

## Funções de Geometria

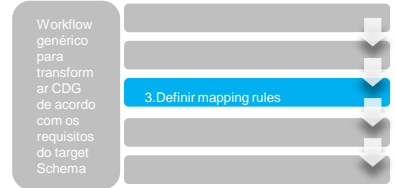
Geoserver

- Ficam ativas se clicar na geometria do **source schema**
- Funções de geometria simples na maioria.
- **Network expansion** permite converter para um diferente tipo de geometria (linhas -> Pontos)
- **Aggregate** - Junta geometrias





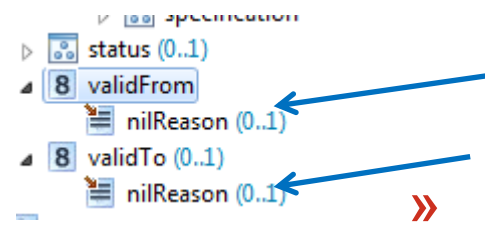
# Definir Mapping rules



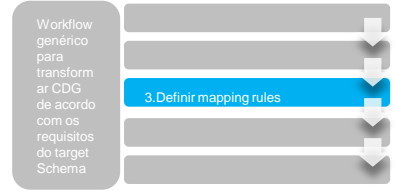
▪ Algumas propriedades no **target Schema**, especialmente as obrigatórias podem receber o valor de **void**

▪ As propriedades **voidable**, podem ter uma subpropriedade **nilReason**

Para caracterizar a **nilReason** podemos utilizar uma **codelist (void reason value)** está no folder **Others**.



# Definir Mapping rules



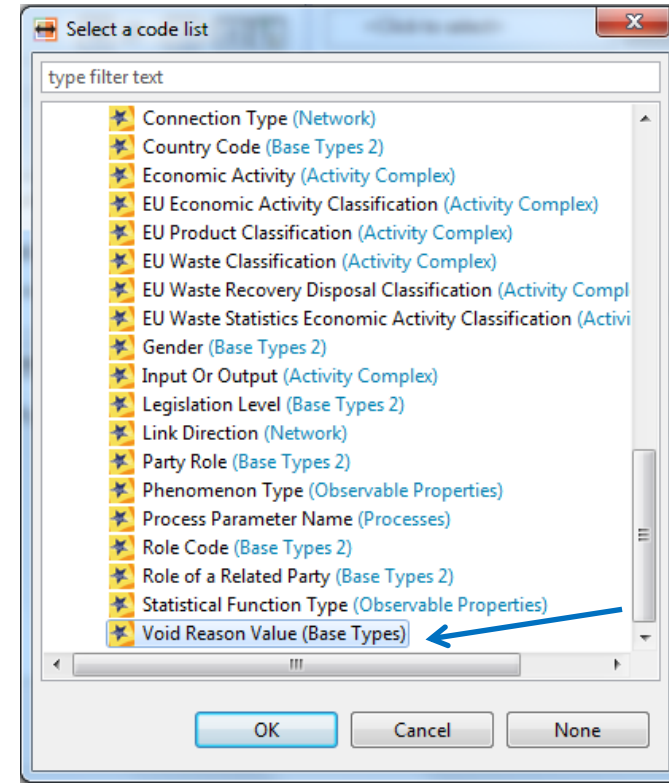
- ***Void reason value***

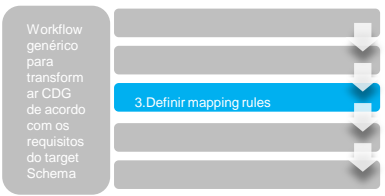
- Pode conter 3 valores:

- ***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.

- 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'.





# Definir Mapping rules

- ***Void reason value***
- ***Unpopulated***

The characteristic is not part of the dataset maintained by the data provider. However, the characteristic may exist in the real world.

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'.





- ***Withheld***

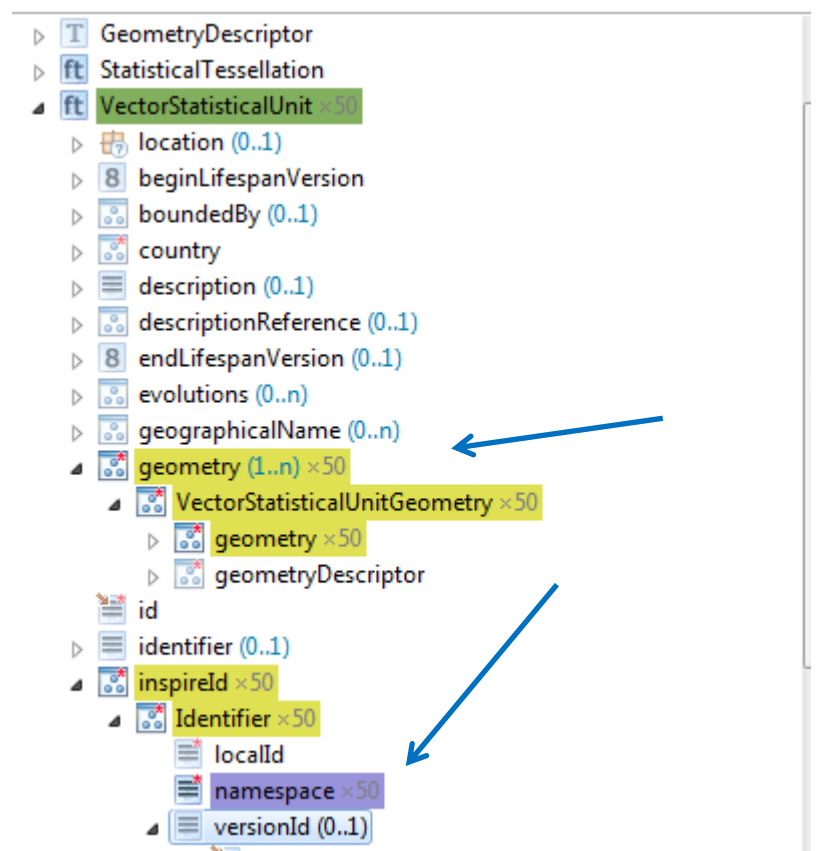
The characteristic may exist, but is confidential and not divulged by the data provider.

# Definir Mapping rules

## ▪ Sistema de cores dos elementos

▪ Para identificar os elementos do **schema** já mapeados o **Hale** utiliza um sistema de cores que nos permitem ter uma noção do trabalho já realizado

-  city Not mapped
-  street Mapped explicitly with a relation
-  address Mapped implicitly due to the mapping of a sub-property
-  type Value assignment independent of the source schema

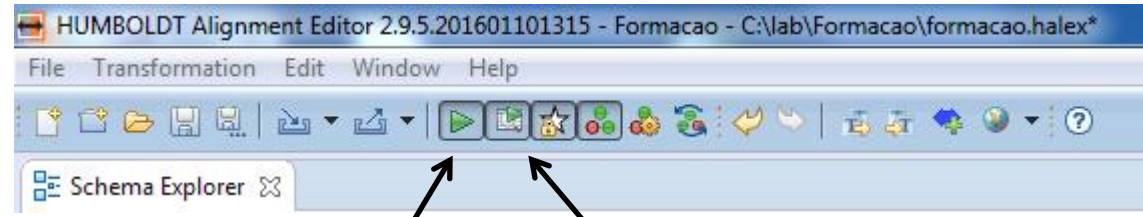


The screenshot shows a hierarchical schema tree. The root node is 'GeometryDescriptor'. Under it are 'StatisticalTessellation' and 'VectorStatisticalUnit ×50'. 'VectorStatisticalUnit ×50' is expanded to show its children: 'location (0..1)', 'beginLifespanVersion', 'boundedBy (0..1)', 'country', 'description (0..1)', 'descriptionReference (0..1)', 'endLifespanVersion (0..1)', 'evolutions (0..n)', 'geographicalName (0..n)', 'geometry (1..n) ×50', 'id', 'identifier (0..1)', 'inspireId ×50', and 'versionId (0..1)'. The 'geometry (1..n) ×50' node is expanded to show 'VectorStatisticalUnitGeometry ×50', which contains 'geometry ×50' and 'geometryDescriptor'. The 'inspireId ×50' node is expanded to show 'Identifier ×50', which contains 'localId', 'namespace ×50', and 'versionId (0..1)'. Blue arrows point to the 'geographicalName (0..n)' node and the 'namespace ×50' node. The 'VectorStatisticalUnit ×50' node and its children are highlighted in green, yellow, and purple, corresponding to the color-coding system described in the adjacent list.

# Definir Mapping rules

## ▪ Sistema de validação

- Verifica se existem inconsistências no mapeamento nomeadamente atributos obrigatórios no **target schema** não mapeado, ou restrições ignoradas.



### **Live transformation button**

Permite que as instancias carregadas no projeto sejam transformadas sempre que haja um mapeamento.

### **Validate transformed instances automatically**

Executa uma validação das instâncias carregadas sempre que haja um mapeamento.

# Definir Mapping rules

Workflow genérico para transformar o ODG de acordo com os requisitos do target Schema

- 1. Definir o ODG
- 2. Definir o target Schema
- 3. Definir mapping rules**
- 4. Executar o processo

## ▪ Sistema de validação, com erros



The screenshot shows a software interface with two main panels. The left panel, titled 'Instance validation', displays a tree structure of warnings. The right panel, titled 'Mapping', shows a list of tasks with their completion status and timestamps.

**Instance validation Warnings:**

- VectorStatisticalUnit (200 warnings)
  - country (50 warnings)
    - Cardinality (showing 5 of 50)
      - Not enough values for the property present: 0 < 1
      - Not enough values for the property present: 0 < 1
      - Not enough values for the property present: 0 < 1
      - Not enough values for the property present: 0 < 1
      - Not enough values for the property present: 0 < 1
  - id (50 warnings)
    - Cardinality (showing 5 of 50)
      - Not enough values for the property present: 0 < 1
      - Not enough values for the property present: 0 < 1
      - Not enough values for the property present: 0 < 1
      - Not enough values for the property present: 0 < 1

**Mapping Log:**

| Task                     | Time     |
|--------------------------|----------|
| Instance validation      | 13:57.28 |
| Instance transformation  | 13:57.26 |
| Instance validation      | 13:57.17 |
| Instance transformation  | 13:57.15 |
| Instance validation      | 11:02.03 |
| Instance transformation  | 11:02.01 |
| Instance validation      | 11:01.02 |
| Instance transformation  | 11:01.00 |
| Instance validation      | 11:00.47 |
| Instance transformation  | 11:00.46 |
| Instance validation      | 11:00.35 |
| Instance transformation  | 11:00.32 |
| Instance validation      | 09:39.59 |
| Instance transformation  | 09:39.57 |
| INSPIRE code list import | 09:39.51 |
| Load data into database  | 09:39.40 |
| Shapefile import         | 09:39.40 |

Total e lista de erros

Instance validation

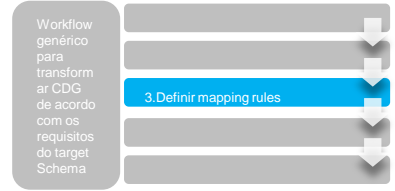
Símbolo Warning

Significa que existem erros





# Definir Mapping rules



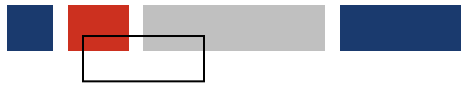
## ▪ Sistema de validação, sem erros

The screenshot displays a software interface with two main panels. The left panel, titled 'Instance validation', shows a 'Report' section with the following details: Success: true, Summary: Finished successfully, Time: Mon Apr 11 14:51:04 BST 2016, and Duration: 31 milliseconds. The right panel, titled 'Type hierarchy', shows a list of tasks for the date 14:47 2016-04-11. The first task, 'Instance validation', is highlighted with a blue box and a green checkmark. Other tasks include 'Instance transformation', 'HALE XML project export', and multiple 'INSPIRE code list import' tasks, all marked with green checkmarks. A red double arrow icon is visible at the top right and bottom right of the interface.

Lista de erros deixa de existir

*Instance validation*

Símbolo = visto (*checked*)



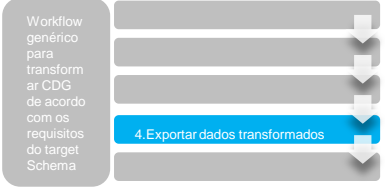
## Exportar dados transformados



### ▪ Exportar dados quando:

- não existem erros reportados no ***Instance validation***
- não existe mais nenhum mapeamento para realizar
- não existir ***instance sampling***

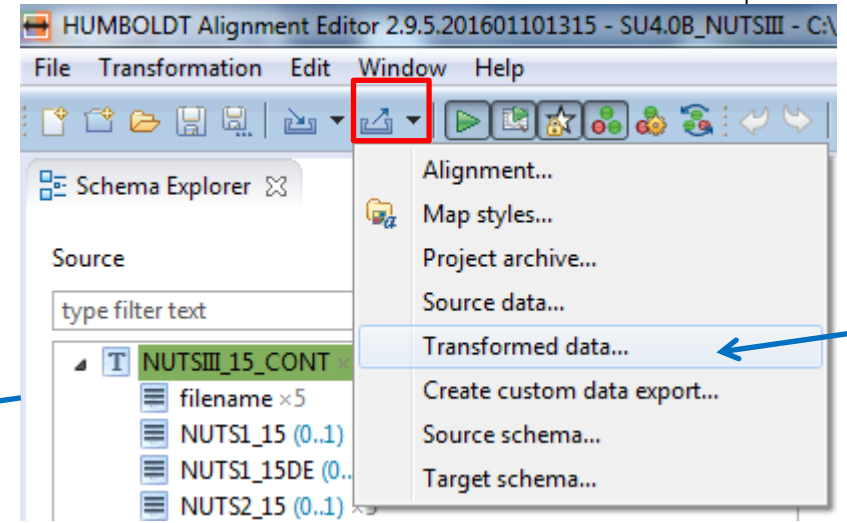
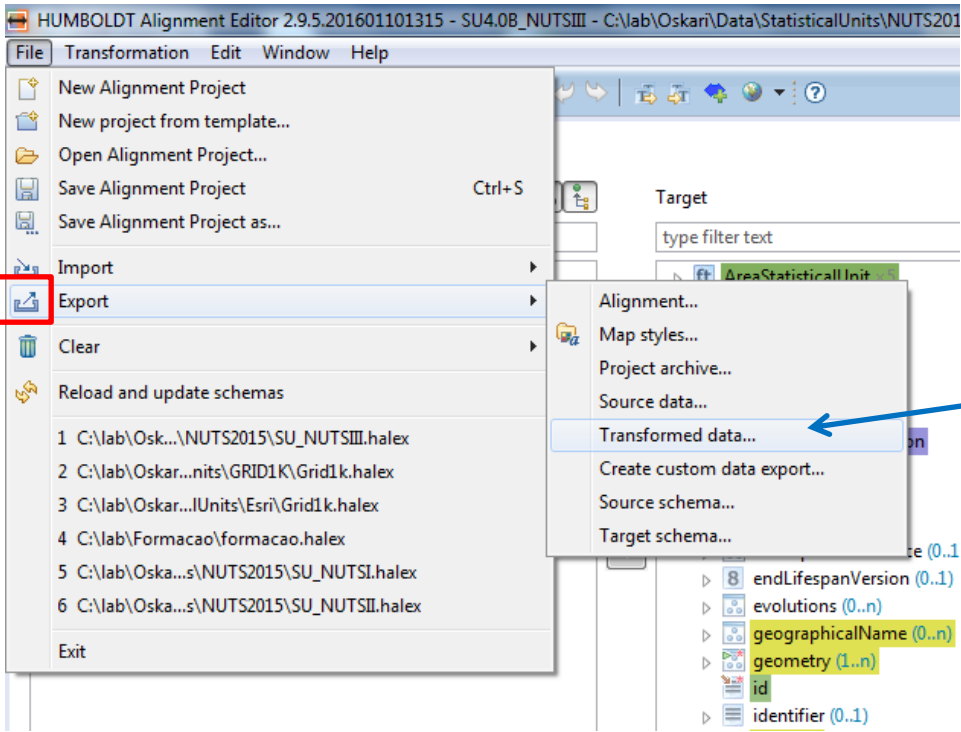




# Exportar dados transformados

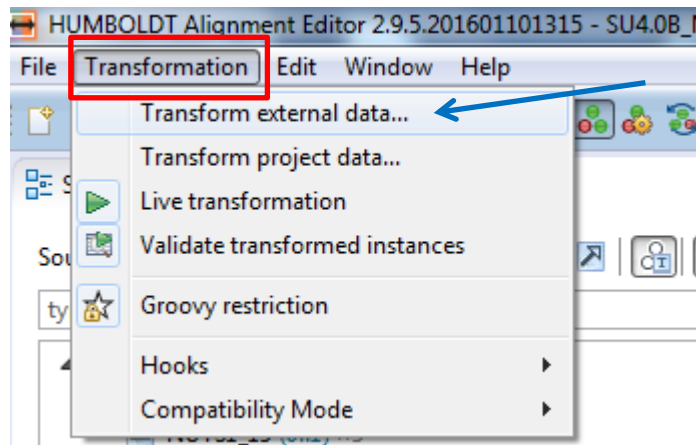
## Existem duas formas para transformar dados

### 1. Transformar as instâncias carregadas no projeto



## Existem duas formas para transformar dados

### 2. Transformar dados externos



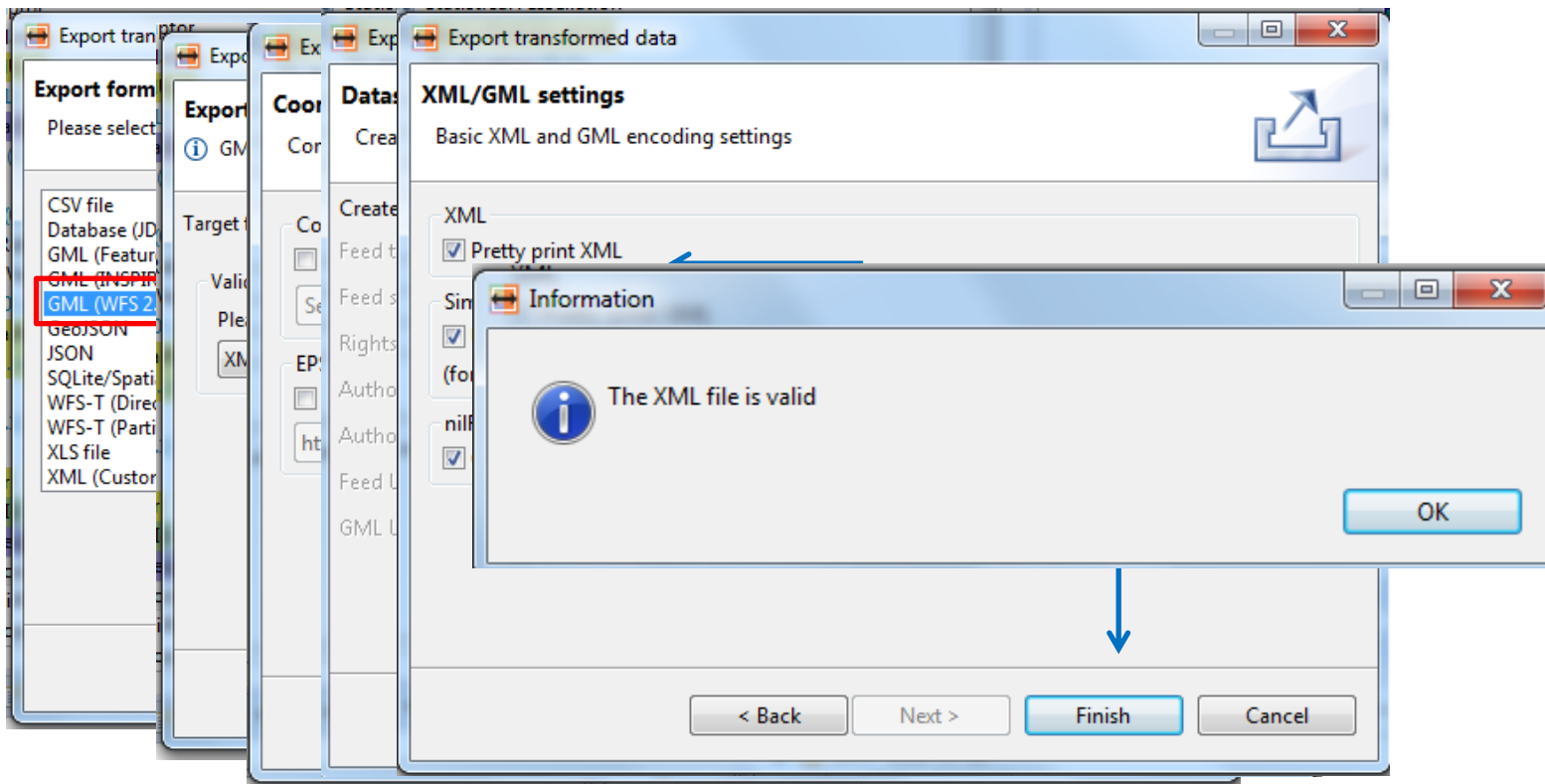
▪ Não se recomenda este método para dados com muitas instâncias: no caso da BGRI (~200 000 registos) não chega a terminar.

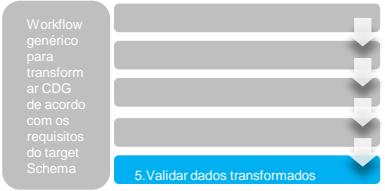
▪ Utiliza-se para partilhar o projeto Hale com dados de Portugal continental e R.A.'s

# Exportar dados transformados



## ▪ Tipo de formato e outras opções





# Validação dados transformados

- **eENVplus Validation Service**
- [http://cloud.epsilon-italia.it/eenvplus\\_new/](http://cloud.epsilon-italia.it/eenvplus_new/)
- Só 4 temas validam com **Schematron**

OGC® Making location count.

eENVplus Validation Service

TEAM Engine v4

Results for session s0007

Test Suite: GML 3.2 (ISO 19136:2007) Conformance Test Suite

Test tns:Main (View Details): Passed

Summary of results

|               |        |          |            |         |         |        |                    |
|---------------|--------|----------|------------|---------|---------|--------|--------------------|
| Best Practice | Passed | Continue | Not Tested | Warning | Skipped | Failed | Failed (Inherited) |
| 0             | 1      | 0        | 0          | 0       | 0       | 0      | 0                  |

See the [detailed test report](#).

Execute this session again | Delete this session | Download log Files | Create execution log report file

[Sessions list](#)

If you have any questions or suggestions, feel free to contact the [eENVplus team](#).

TEAM Engine 4.0.5



# Validação dados transformados

Workflow genérico para transformar CDG de acordo com os requisitos do target Schema

5. Validar dados transformados

■ **Oxygen**

■ **Software com custos**

Project: sample.xpr

Outline: wfs:FeatureCollection [http://inspire.ec.europa.eu/...]

```
1 <?xml version="1.0" ?>
2 <wfs:FeatureCollection xmlns:au="http://inspire.ec.europa.eu/schemas/au/4.0"
3   xmlns:gts="http://www.isotc211.org/2005/gts"
4   xmlns:gco="http://www.isotc211.org/2005/gco"
5   xmlns:ad="http://inspire.ec.europa.eu/schemas/ad/4.0"
6   xmlns:hfp="http://www.w3.org/2001/XMLSchema-hasFacetAndProperty"
7   xmlns:gn="http://inspire.ec.europa.eu/schemas/gn/4.0"
8   xmlns:bu-base="http://inspire.ec.europa.eu/schemas/bu-base/4.0"
9   xmlns:base="http://inspire.ec.europa.eu/schemas/base/3.3"
10  xmlns:cp="http://inspire.ec.europa.eu/schemas/cp/4.0"
11  xmlns:su-core="http://inspire.ec.europa.eu/schemas/su-core/4.0"
12  xmlns:xlink="http://www.w3.org/1999/xlink"
13  xmlns:net="http://inspire.ec.europa.eu/schemas/net/4.0"
14  xmlns:tn="http://inspire.ec.europa.eu/schemas/tn/4.0"
15  xmlns:ns1="http://www.w3.org/2001/XMLSchema-instance"
16  xmlns:base2="http://inspire.ec.europa.eu/schemas/base2/3.3"
17  xmlns:gml="http://www.opengis.net/gml/3.2"
18  xmlns:su-vector="http://inspire.ec.europa.eu/schemas/su-vector/4.0/StatisticalUnitVector.xsd"
19  xmlns:gss="http://www.isotc211.org/2005/gss"
20  xmlns:sc="http://www.interact.org/2005/sc"
21  xmlns:gsr="http://www.isotc211.org/2005/gsr"
22  xmlns:gmd="http://www.isotc211.org/2005/gmd"
23  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
24  xmlns:wfs="http://www.opengis.net/wfs/2.0"
25  xsi:schemaLocation="http://inspire.ec.europa.eu/schemas/su-vector/4.0/StatisticalUnitVector.xsd
26  http://www.opengis.net/wfs/2.0 http://schemas.opengis.net/wfs/2.0/wfs.xsd"
27  numberMatched="5" numberReturned="5" timeStamp="2016-04-11T16:10:09.578+01:00">
28   <wfs:member>
29     <su-vector:AreaStatisticalUnit gml:id="Id1">
30       <gml:name>NUTSIII 2015</gml:name>
31     </su-vector:AreaStatisticalUnit>
32   </wfs:member>
33 </wfs:FeatureCollection>
```

Attributes: wfs:FeatureCollection [http://www.opengis.net/wfs/2.0]

| Attribute      | Value   |
|----------------|---|
| numberMatched  | 5   |
| numberReturned | 5   |
| timeStamp      | 2016-04-11T16:10:09.578+01:00                   |
| xmlns:ad       | http://inspire.ec.europa.eu/schemas/ad/4.0      |
| xmlns:au       | http://inspire.ec.europa.eu/schemas/au/4.0      |
| xmlns:base     | http://inspire.ec.europa.eu/schemas/base/3.3    |
| xmlns:base2    | http://inspire.ec.europa.eu/schemas/base2/3.3   |
| xmlns:bu-base  | http://inspire.ec.europa.eu/schemas/bu-base/4.0 |
| xmlns:cp       | http://inspire.ec.europa.eu/schemas/cp/4.0      |

Validate with dialog:

URL: file:/C:/lab/Oskari/Data/SLD/StatisticalUnitVector.xsd

Schema type: XML Schema

Public ID:

Schematron phase:

OK Cancel

Error: E [Xerces] cvc-elt.1.a: Cannot find the declaration of element 'wfs:FeatureCollection'.