

# Handleiding Gebruik HALE voor INSPIRE dataharmonisatie

## Inleiding

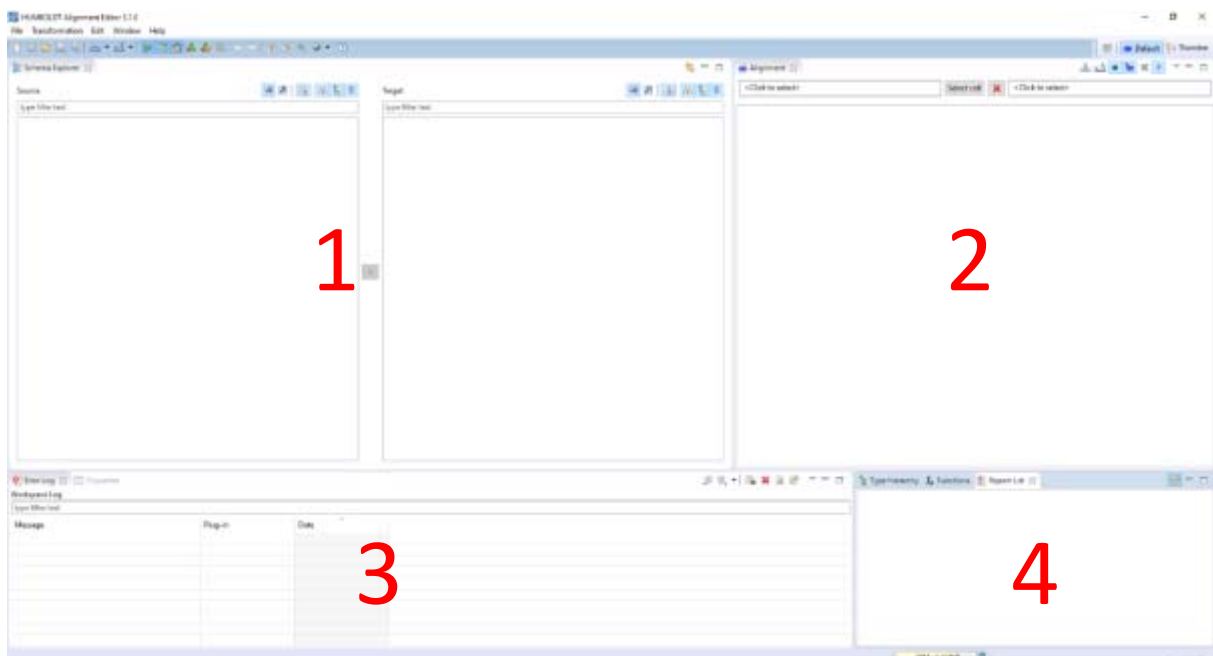
HALE Studio is een Open Source oplossing die datatransformaties uitvoert om van een dataset in en bepaald schema/formaat naar een dataset in een ander schema/formaat te komen.

Deze handleiding beschrijft in het kort de belangrijkste stappen en aandachtspunten om een datatransformatie in het kader van de INSPIRE data harmonisatie uit te voeren.

De huidige versie van HALE is versie 3.3.0 van april 2017.

## HALE opstarten

Start HALE op. Het default opstartscherm ziet er als volgt uit:



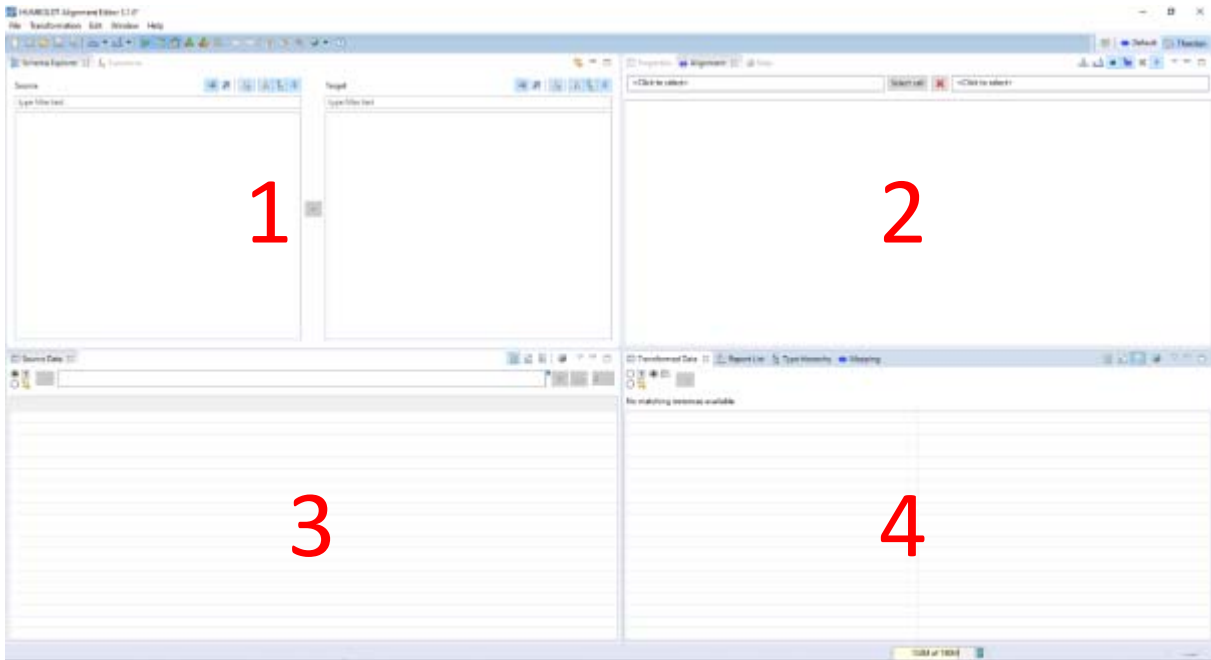
Dit opstartscherm bestaat uit vier vensters:

1. Venster Schema Explorer: Hierin staat het schema van het transformeren bestand (input) en het schema van het getransformeerde bestand (output). Door middel van de pijl in het midden kan de transformatie worden uitgevoerd.
2. Venster Alignment: Hierin wordt de transformatie grafisch afgebeeld
3. Venster Error Log/: Hierin wordt de logfile van de sessie afgebeeld. In het tabblad Properties kun je de eigenschappen van attributen van de input en output en foutmeldingen in detail bekijken.
4. Venster Report List: Hierin worden de resultaten van de acties in de lopende sessie afgebeeld. In het tabblad Type Hierarchy kun je de hiërarchie van de objecttypen bekijken. In het tabblad Functions toont de beschikbare transformatie functies.

De bovenste balk bevat de iconen voor de transformatie. De menubalk bevat de dropdown menu's hiervoor.

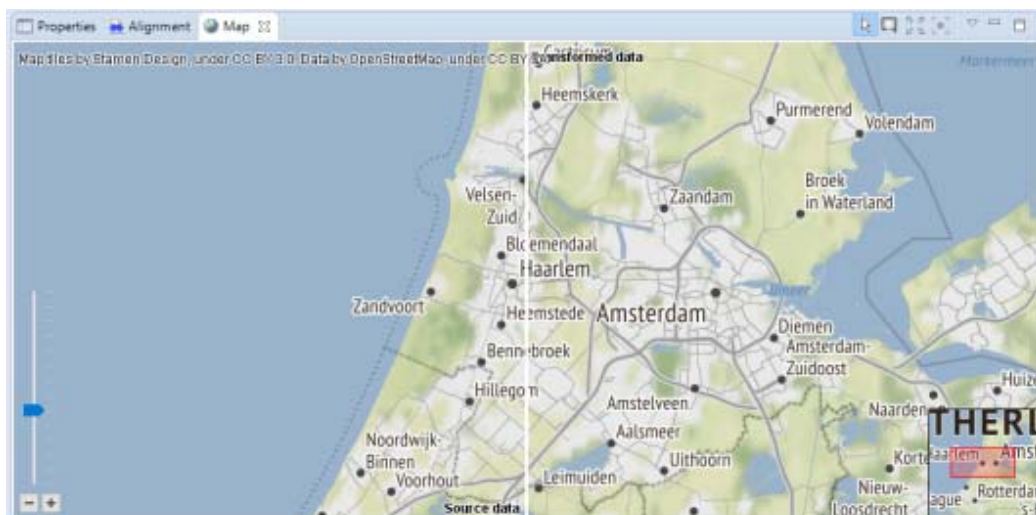
Het opstartscherm is zelf aan te passen door in de menubalk op Window – Show View te drukken. Hiermee kun je zelf vensters aanmaken en slepen naar de juiste plek of als tabblad toevoegen aan een venster. Er zijn een aantal voor gedefinieerde scherm lay-outs, perspectives. Deze vind je in de rechterbovenhoek.

Het heeft de voorkeur het scherm aan te passen met het perspective Thorsten. Door hierop te drukken verandert het scherm.



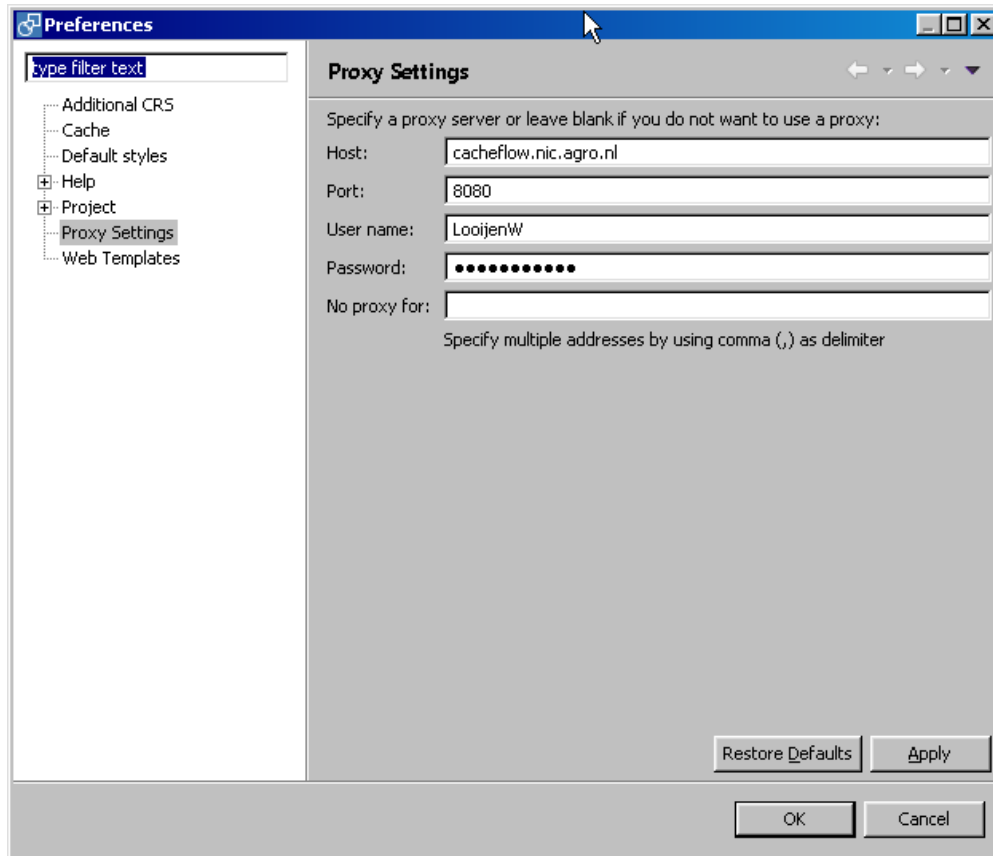
Dit scherm bevat nog steeds vier vensters maar biedt nu de mogelijkheid om de attributen van de brondata en de getransformeerde data te bekijken en de mogelijkheid om ook grafisch naar de transformatie te kijken.

1. Venster Schema Explorer: hieraan is het tabblad Functions toegevoegd.
2. Venster Alignment: hieraan is het tabblad Properties en het tabblad Map toegevoegd.



a. Het tabblad Map laat een basiskaart zien in twee schermen (links en rechts) en een overzichtsvenster. Links staat de brondata en rechts staat de getransformeerde data. Door met de linkermuisknop het beeld te schuiven loopt de brondata over in de getransformeerde data.

Zie je geen achtergrondkaart dan zijn de proxy settings niet goed ingesteld. Ga naar Window – Settings en vul de volgende gegevens in:



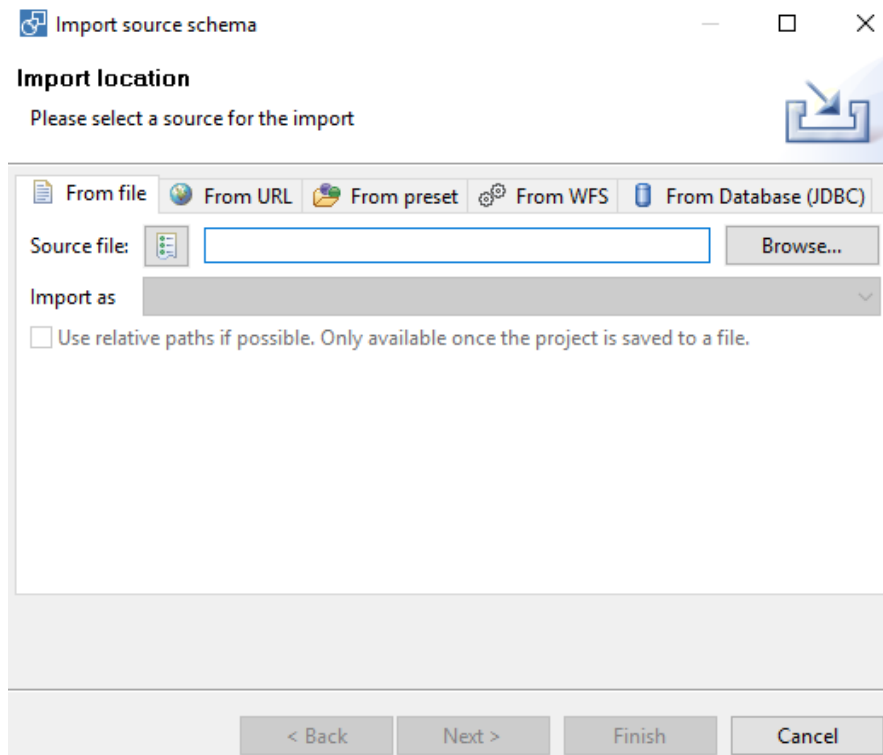
3. Venster Source Data: Hierin staan maximaal 5 objecten uit de brondata met alle attribuutwaarden

4. Venster Transformed Data: Hierin staan maximaal 5 objecten uit de getransformeerde data met alle attribuutwaarden. Het tabblad Report List en Type Hierarchy zijn hieraan toegevoegd. Het tabblad Mapping laat de detailtransformatie zien.

### Data transformatie opstarten

Voor het opstarten van de data transformatie is het volgende van belang. Er moet een bronschema en een doelschema geladen worden. Daarnaast moet de brondata ingelezen worden. Het bronschema is in feite niets anders dan de opbouw van de brondata. Het doelschema is niets anders dan de opbouw van de getransformeerde data.

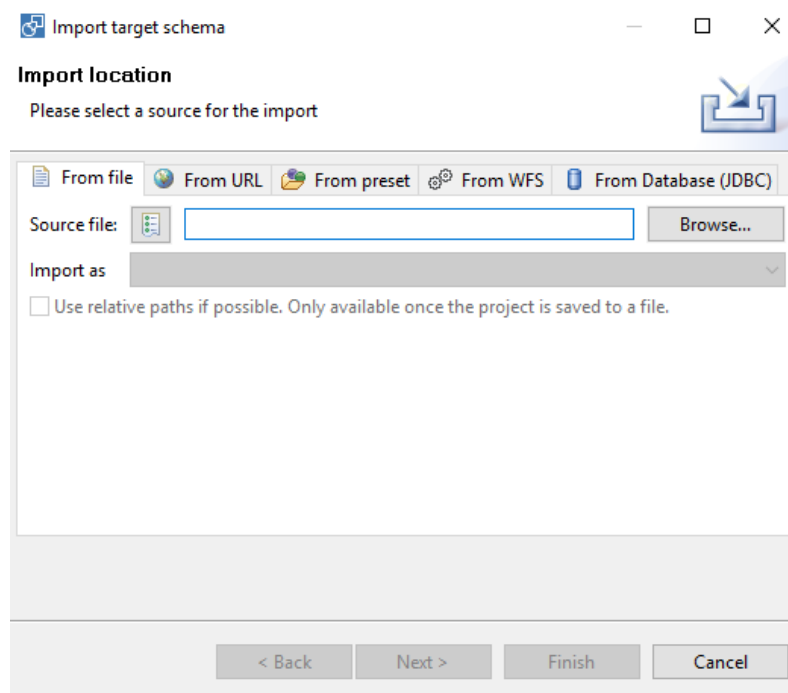
Stap 1: Ga naar de menubalk File – Import – Source schema:



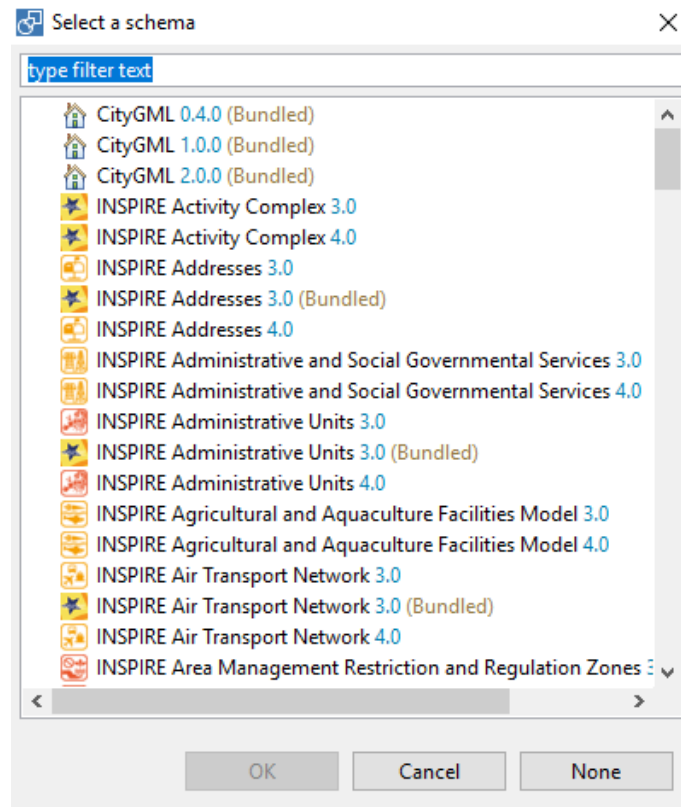
Je kunt het bronschema op verschillende manieren importeren. Bij de INSPIRE data harmonisatie gaat het om de transformatie van as-is bestanden die voornamelijk als shapefile beschikbaar zijn. Om het bronschema te importeren importeer je deze shapefile.

In het venster Schema Explorer zie je nu het geïmporteerde bronschema. Door op de > te drukken voor de naam van het bestand worden de attributen zichtbaar.

Stap 2: Ga naar de menubalk File – Import – Target schema:



Ook het doelschema kun je op verschillende manieren importeren. HALE heeft de INSPIRE dataspecificaties al vertaald naar doelschema's en deze zijn te importeren via het tabblad From preset. Door hierop te drukken en met de muis in het lege vak Source file te gaan staan en op de rechtermuisknop te drukken kun je een INSPIRE doelschema selecteren:



Kies het benodigde schema (altijd versie 4.0 Simple) welke past bij het thema van de brondata.

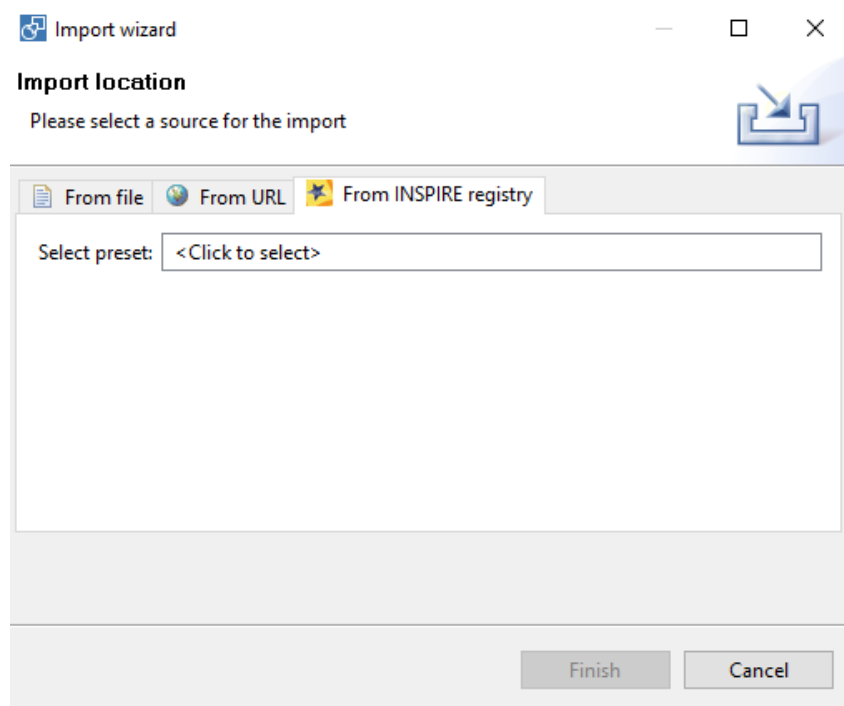
In het venster Schema Explorer zie je nu het geïmporteerde doelschema. Door op de > te drukken voor de naam van het bestand worden de attributen zichtbaar.

Stap 3: Ga naar de menubalk File – Import – Source data

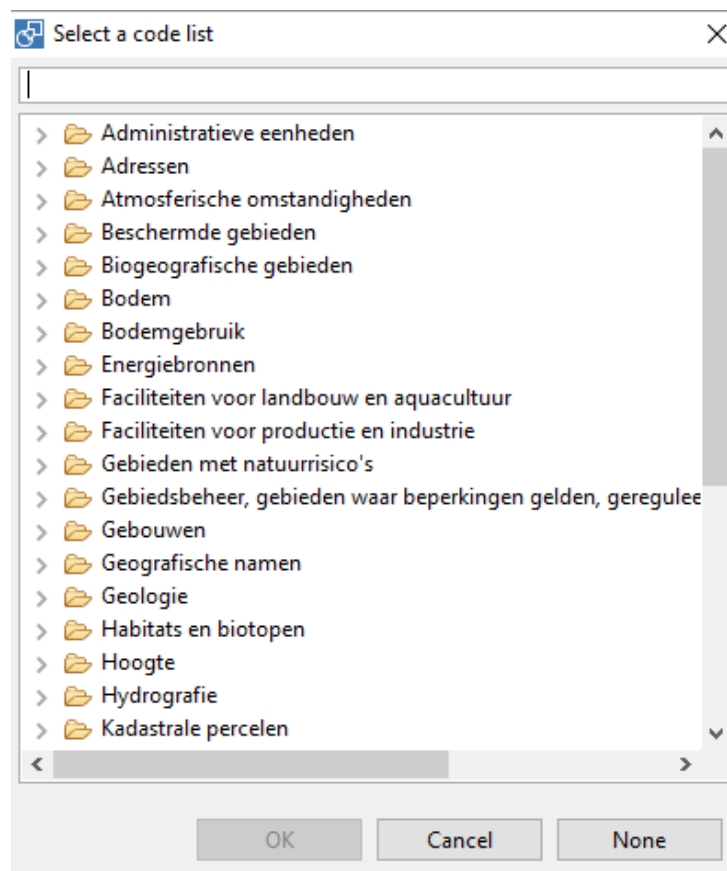
Je kunt de brondata op verschillende manieren importeren. Bij de INSPIRE data harmonisatie gaat het om de transformatie van as-is bestanden die voornamelijk als shapefile beschikbaar zijn. Om de brondata te importeren importeer je deze shapefile.

Het is handig om van tevoren na te gaan welke projectie de bron data heeft. In geval van een shapefile als bron data hernoem dan de .prj file en lees dan de bron data in. Je wordt dan gevraagd om de projectie op te geven (in het geval van RD NEW is dat EPSG 28992). Dit heeft als voordeel dat de uiteindelijke transformatie naar de ETRS89 projectie (EPSG: 3035) kan worden uitgevoerd, zonder foutmeldingen.

Stap 4: Ga naar de menubalk File – Import – Code list:







dda INSPIRE registry. Door hierop te drukken en met de muis in het lege vak Source file te gaan staan en op de rechtermuisknop te drukken kun je een INSPIRE codelijst selecteren:





Kies de benodigde codelijst welke past bij het thema van de brondata.

In de Schema Explorer zie je naast tekst verschillende andere symbolen. Deze geven de verschillende classificaties weer van de objecttypen:

-  Een type gedefinieerd in het schema.
-  Een abstract type , hiervoor is mogelijk geen data representatie
-  Zogeheten feature types, die geïmporteerd zijn vanuit een shapefile en die in GML applicatie schema's anders zijn dan de abstracte feature type
-  Een abstract feature type





Er zijn ook groepen van types gedefinieerd:

-  Normale groep met types
-  Keuze groep waar alleen één van de types een waarde kan bevatten

De verschijningsvorm van de types zijn:


-  Tekst
-  Numeriek
-  Geometrie
-  Ander (complex)

Bovenop het classificatie icoon staat aanvullende informatie:


-  Een rood sterretje geeft aan dat dit schema element verplicht is, ze komen één keer voor en ze moeten een waarde hebben, deze waarde mag niet 0 zijn. Let op: als de "parent" van een type een keuzegroep is () dan kan er maar één "child" aanwezig zijn, deze mag nog steeds niet leeg zijn of de waarde 0 bevatten indien het rode sterretje aanwezig is.
-  Een bruin pijltje geeft aan dat dit een eigenschap van een XML schema is die gedefinieerd als een attribuut.
-  Een groen driehoekje geeft aan dat dit object de belangrijkste geometrie bevat, die getoond wordt in het Map tabblad.

Bronschema context: Als de naam van een schema element tussen haakjes staat betekent dit dat dit element een context definieert op basis van een origineel schema element. Er zijn verschillende mogelijkheden


Conditie context type: Er is een conditie geplaatst op het type op basis van de bron data (hier: alleen de items van het type shirt)

 (item) type = 'shirt'


Conditie context eigenschap: Er is een conditie geplaatst op het eigenschap van het type (hier alleen de waardes die Bd\* bevatten)

 (bulkDensityOrSolidVolume) (0..1) value like 'Bd%'


Index context: Alleen de eigenschap van de nul index is gebruikt (hier de eerste waarde van de eigenschap name ).

\* (name) (1..n) [0]

Object context: geeft een additionele verschijningsvorm aan van een eigenschap

 (attribute) (0..n)


Naast de eigenschap van een type staat informatie over de kardinaliteit van die eigenschap. Deze bestaat uit twee waardes tussen haakjes. De eerste waarde is het minimal aantal keren dat de eigenschap voorkomt. De tweede waarde geeft het maximum aantal keren aan.

 description (0..1)

De waarde  $n$  geeft aan dat het maximal aantal keren oneindig is.

Zijn er geen waarden aangeven dan betekent dat dat de eigenschap maar één keer voorkomt.

Wanneer data is geladen dan geeft het getal na het vermenigvuldigingsteken aan hoeveel objecten er aanwezig voor die eigenschap van het object (hier 180).

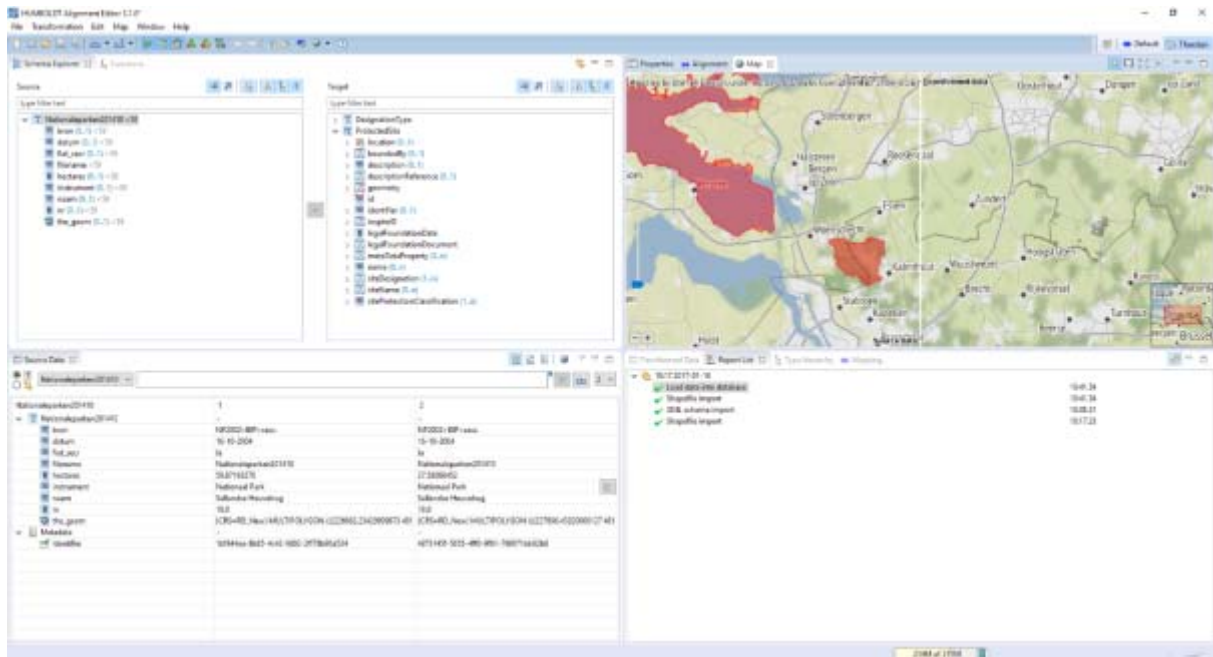
 description (0..1) ×180

Als er een tweede waarde is tussen haakjes dan geeft dit aan hoeveel “parents” er zijn voor deze eigenschap (hier komen er 173 waardes voor in 165 objecten, sommige objecten hebben dus meerdere waardes voor de eigenschap identifier).

 identifier (0..n) ×173 (165)

In het venster Schema Explorer zie je nu de aantallen in de brondata achter de attributen staan. In het linkerdeel van het Map tabblad wordt de data afgebeeld. In het tabblad Report List staan de uitgevoerde acties. Het groene vinkje geeft aan dat de acties succesvol zijn uitgevoerd.





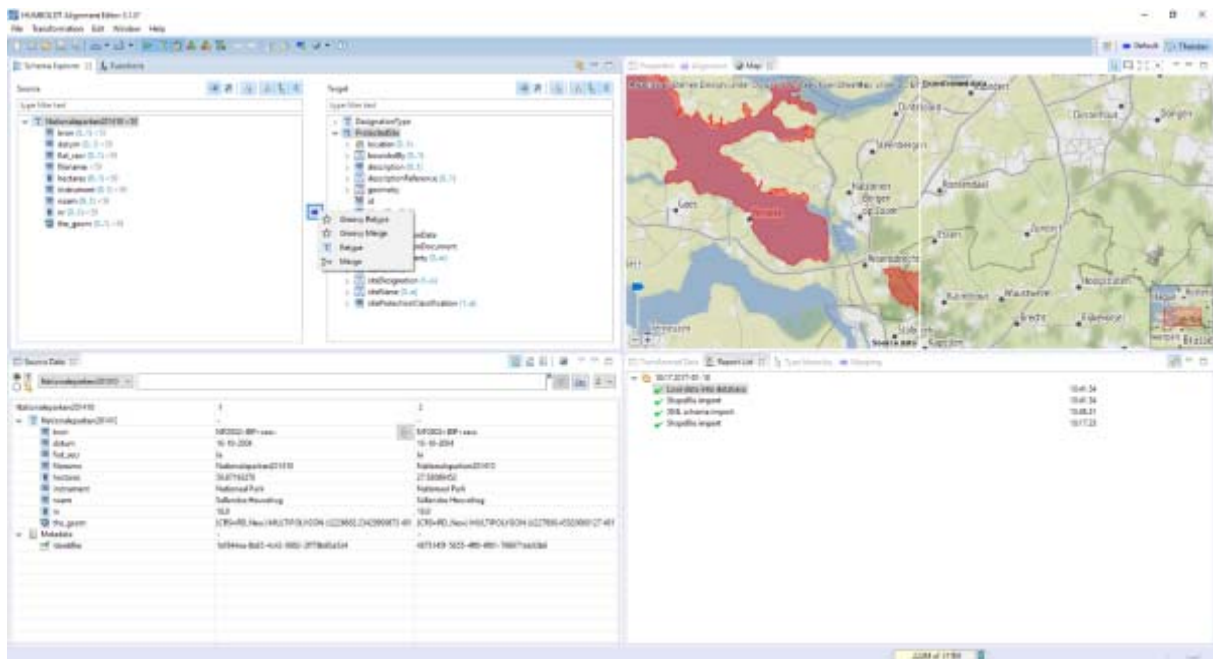
Er kan nu begonnen worden met de daadwerkelijke data transformatie.

### Data transformatie



Om te starten met de data transformatie zijn er een paar stappen nodig die de relatie leggen tussen brondata en doelschema. Deze dien je op voorhand uit te voeren:

Stap 1: Zorg dat Live transformation geselecteerd is door dit in de menubalk onder Transformation te checken. De data transformatie wordt nu live uitgevoerd. Dit is niet aan te raden bij extreem grote dataset (meer dan 5000 rijen).

Stap 2: Leg de relaties tussen bronschema/data en doelschema door in de Schema Explorer het hoofdtype in het bronschema en het hoofdfeature type in het doelschema te selecteren en met de linkermuisknop op de blauwe pijl te drukken. Hierdoor komen de verschillende functies beschikbaar die gebruikt kunnen worden voor de data transformatie van deze objecten. De beschikbare functies zijn afhankelijk van de eigenschappen en verschijningsvormen van de objecten (zowel bron als doel). In dit geval selecteren we Retype. Door deze stap uit te voeren, alle te volgen schermen kun je doorklikken zie je dat de feature type in het doelschema nu hetzelfde aantal objecten bevat als de brondata in het bronschema.

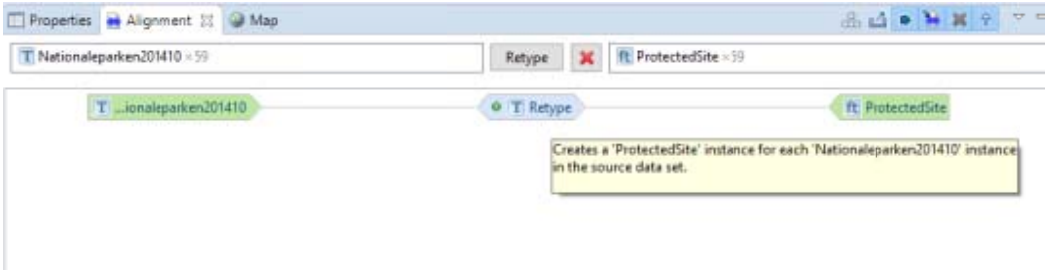


We zien nu dat in de Report List de volgende meldingen staan:

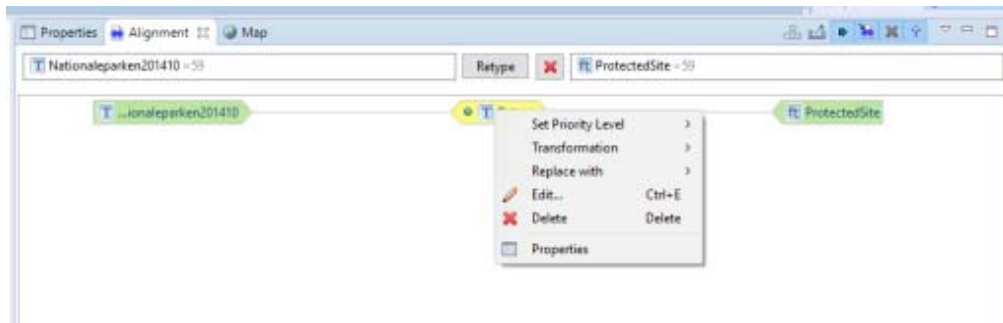
-  Instance validation
-  Instance transformation

Het groene vinkje voor Instance transformation geeft aan dat de Retype goed is gebeurd. Het gevarendriehoekje voor de Instance validation geeft aan dat de validatie van het getransformeerde bestand, die live gebeurt, nog fouten bevat. Om deze fouten te kunnen zien, selecteer de regel met het gevarendriehoekje en open het Properties tabblad en ga naar de regel met Warnings en zie de lijst met foutmeldingen. Omdat door deze eerste stap nog niet alle verplichte (de eigenschappen met de rode sterretjes) zijn ingevuld worden deze foutmeldingen gegenereerd. In deze stap kun je deze foutmeldingen negeren, ze hoeven pas gebruikt te worden als alle verbindingen tussen bron en doelschema gelegd zijn. Zodra alle verbindingen zijn gelegd op de juiste manier zal er ook een groen vinkje voor de Instance validation komen.

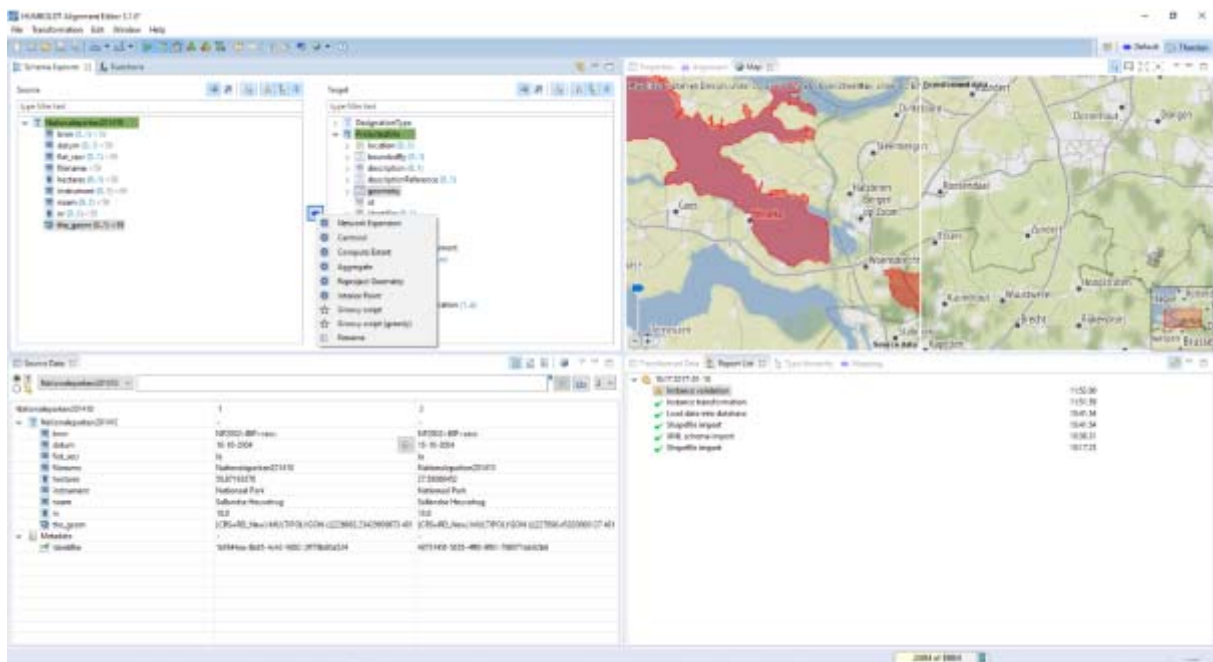
Onder het tabblad alignment kun je de grafische representatie van de datatransformatie zien.



Door met de rechtermuisknop op Retype te klikken kun je deze aanpassen of verwijderen of vervangen door een andere functie of de prioriteit aanpassen.



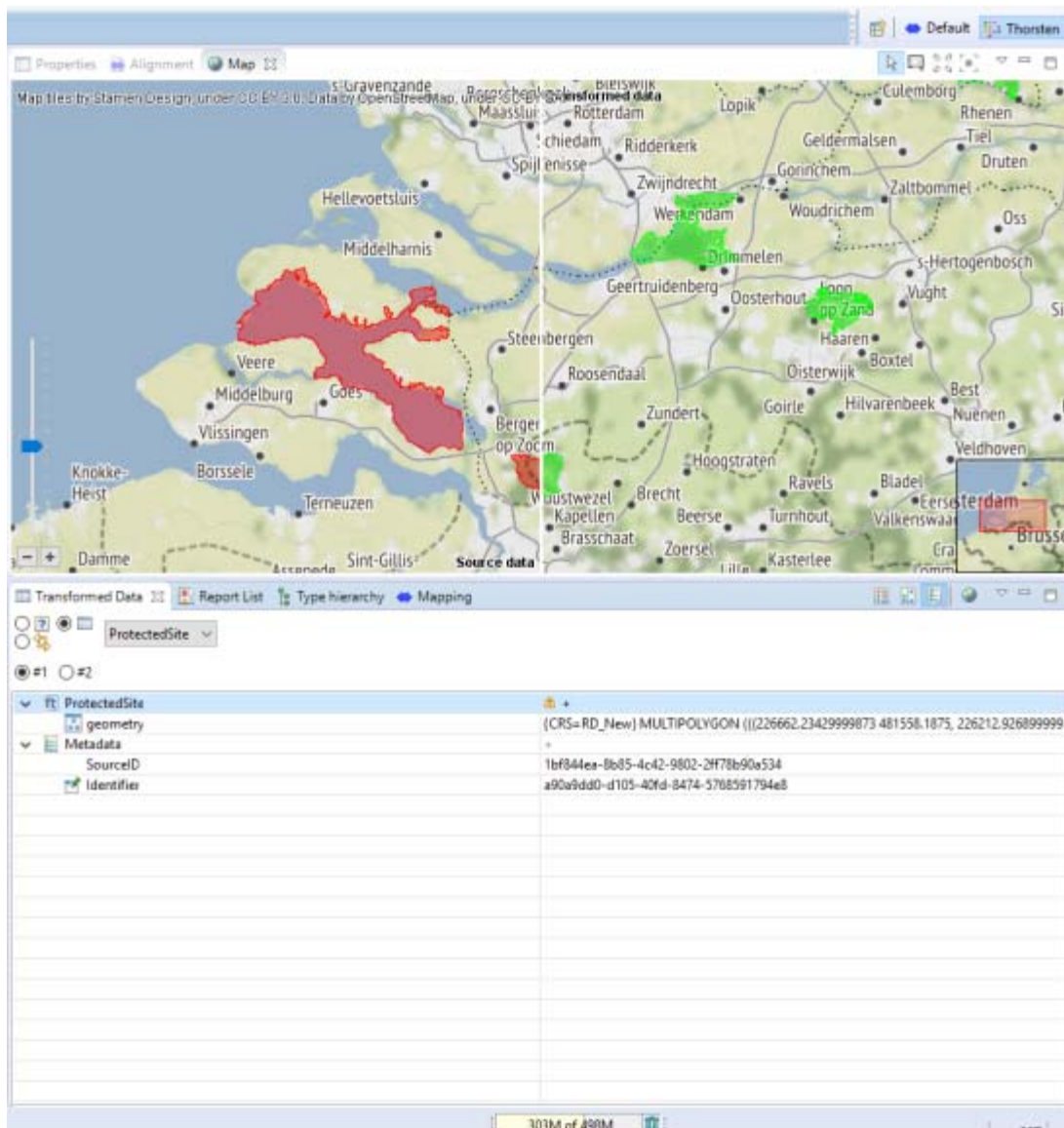
Stap 2: Transformeer de geometrie door in de Schema Explorer in het bronschema attriboot the\_geom te selecteren en in het doelschema geometry, dan naar de blauwe pijl.



Als je geen geometrie ziet in de transformed data map (het rechter scherm) dan moet je met de rechtermuis knop in het doelschema geometry en selecteer set default geometry.

Er zijn nu andere functies beschikbaar. Omdat we geometrie in dit geval één op één overnemen gebruiken we functie Rename. In het geval je geometrische operaties wilt uitvoeren kun je één van de andere functie gebruiken.

In het tabblad Map zie je nu dat in het rechterdeel de getransformeerde geometrie (groen) zichtbaar is geworden. In het tabblad Transformed Data zie je nu de attribootgegevens van de getransformeerde data.



Stap 3 en volgend: Transformeer de attributen uit de brondata naar de gewenste objecten in het doelschema. Gebruik daarbij de kennis van de data en de inhoudelijke kennis van inhoud van de data.

Los eventuele gevarendriehoekjes bij de Instance validation en Instance transformation op, net zolang er bij beiden groene vinkjes staan.

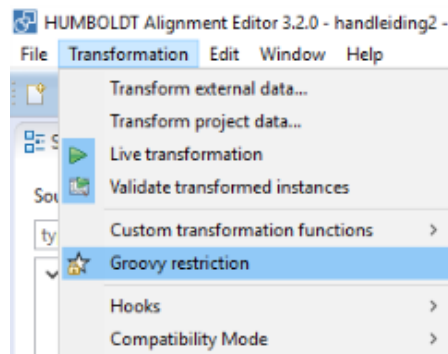
### Aandachtspunten harmonisatie

Als het datum veld in de bron data van het type date is dan moet de stap onder punt 2 worden uitgevoerd. Als het datum veld in de bron data van het type string is dan moeten de volgende stappen uitgevoerd worden:

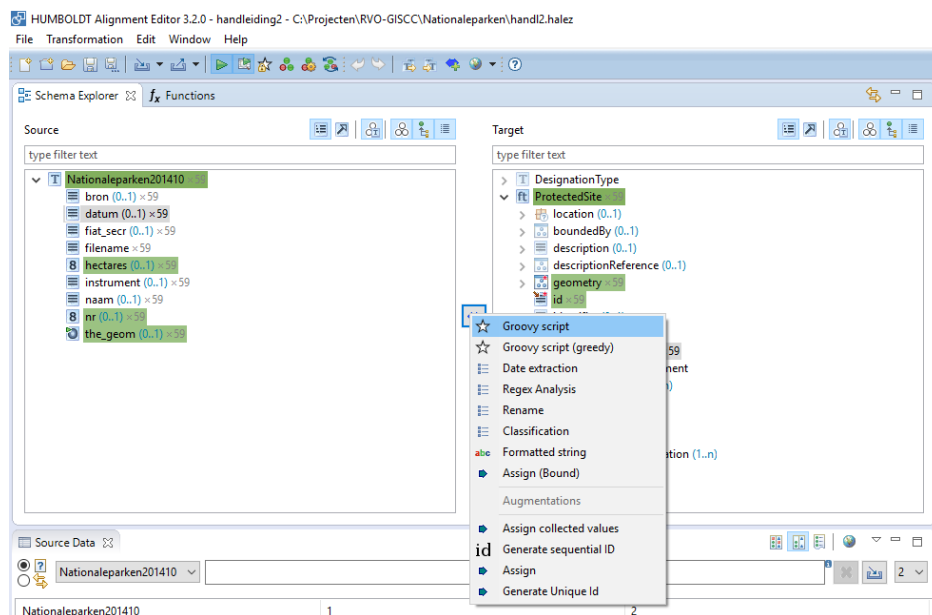
1. Als er datum velden moeten worden getransformeerd dan moet dat (voorlopig; er is een issue gedefinieerd voor nieuwe releases) gebeuren via een zogeheten Groovy script omdat de systeemtijd afwijkt van de Oracle/Java tijd en de resulterende gml dan een foute datum aangeeft.

Dit doe je op de volgende manier:

- Ga naar Transformation en selecteer Groovy restriction. Dit zorgt ervoor dat het Groovy script altijd wordt uitgevoerd (dus wees er zeker van dat het script goed is).

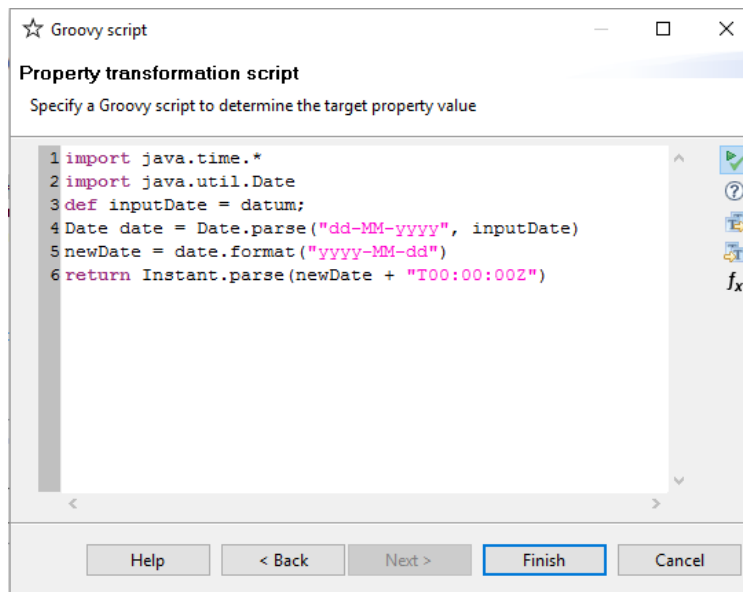


- Selecteer de bron datum variabele en de doel datum variabele die je wilt transformeren en druk op blauwe pijl selecteer daar Groovy script:

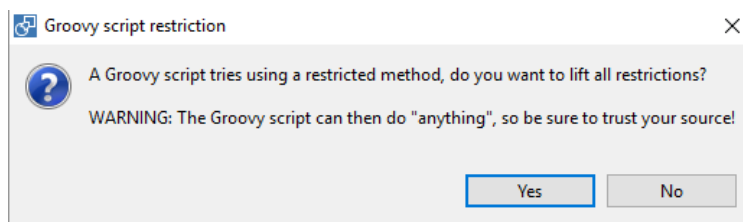


- Je komt dan in een scherm waarin je het Groovy script kunt opstellen. Er wordt een ingebouwde validator gebruikt. Is het script niet goed dan blijft de Finish knop uitgedrukt en kan het script niet gedraaid worden. Het script is al volgt wanneer de datum variabele in de vorm dd-MM-yyyy als String is geschreven.:

```
import java.time.*  
  
import java.util.Date  
  
def inputDate = datum;  
  
Date date = Date.parse ("dd-MM-yyyy", inputDate)  
  
newDate = date.format ("yyyy-MM-dd")  
  
return Instant.parse(newDate + "T00:00:00Z")
```



- In dit voorbeeld is datum de naam van het attribuut uit de bron data die de datum aangeeft. In dit geval is deze datum opgebouwd als dd-MM-yyyy. Is dit in andere bron data anders dan moet dat uiteraard aangepast worden. Omdat de schematransformatie een ISO datum verwacht die conform yyy-MM-dd is opgebouwd moet je deze herschrijven.
- Deze workaround zorgt ervoor dat de transformatie een foutmelding geeft, maar deze kan genegeerd worden. **Tip: is wel doe eerst alle transformaties en check of deze goed zijn en doe dan pas de datum transformatie.**

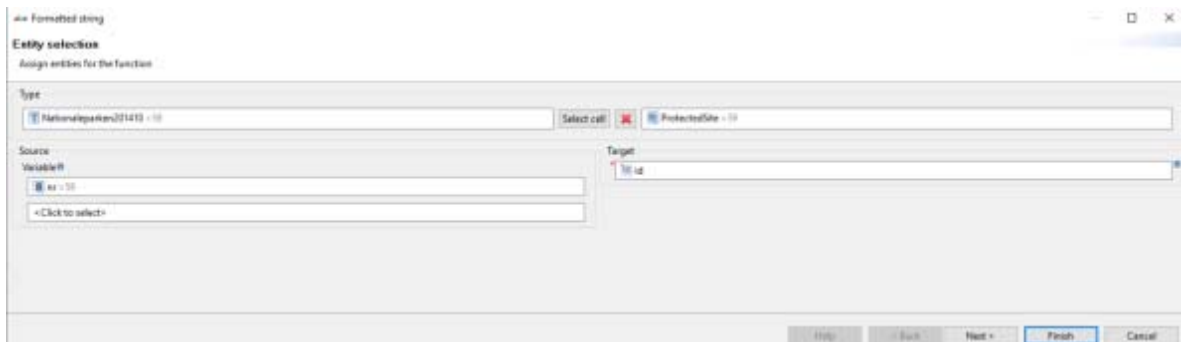
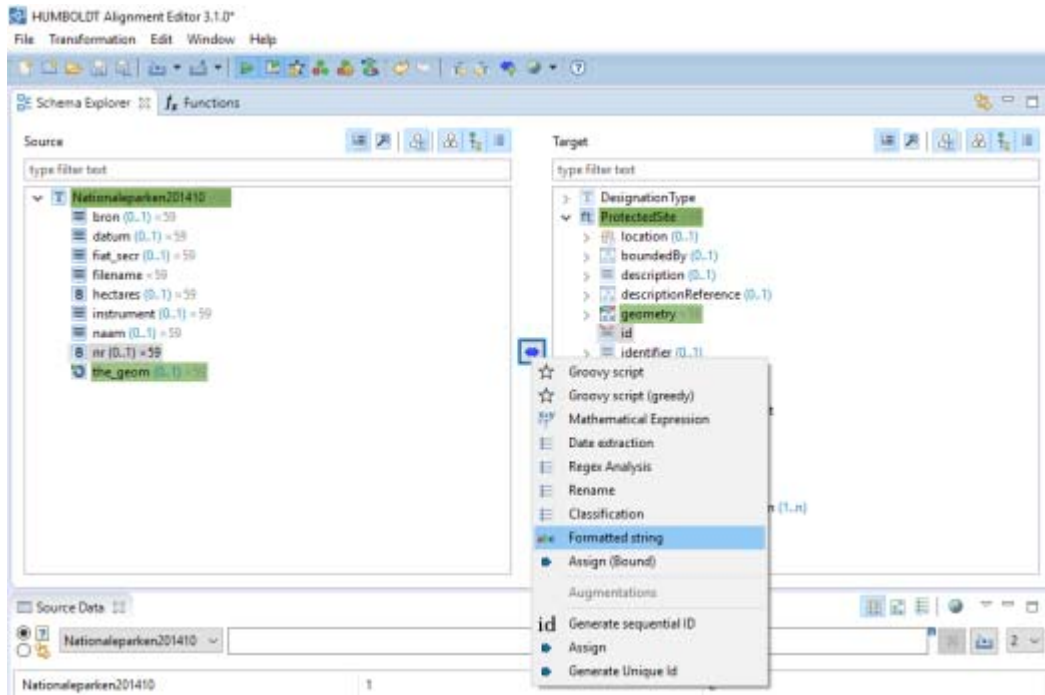


- Als het project opgeslagen is en opnieuw wordt geopend dan krijg je bovenstaande melding. Druk op Yes.

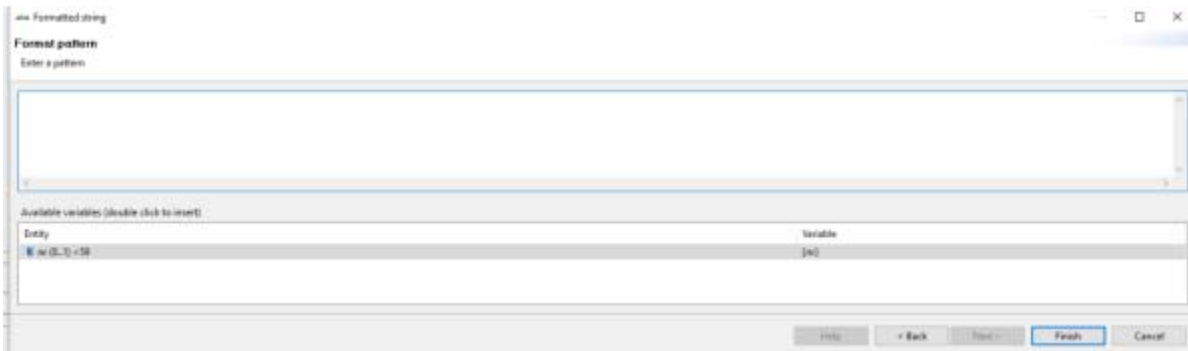
2. Het datum veld uit de bron data moet via een Rename omgezet worden naar het datum veld in de doel data.

3. Het verplichte id in het doelschema dient vooraf te worden gegaan door een alfanumeriek karakter (bv underscore). Dit karakter zit vaak niet in het id van de brondata. Dit wordt opgelost door de functie Formatted String te gebruiken (hier verbinden we het id in de brondata/schema met id in het doelschema):

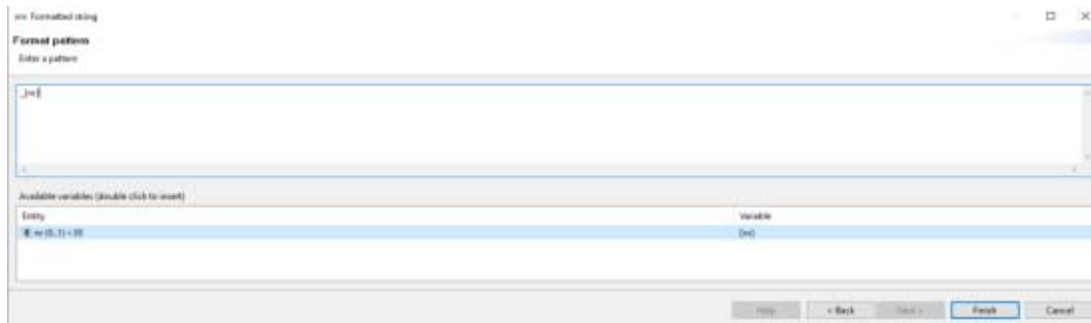




Klik Next



In het veld Enter a pattern type een underscore en selecteer de benodigde variabele (hier nr)



Klik Finish en het id in het doelschema is nu voorzien van een underscore.

4. Het InspireID in het doelschema bevat als onderliggend element een Identifier, die op zijn beurt bestaat uit een localId, namespace en versionID. De eerste twee zijn verplicht en moeten dus ingevuld worden.

Het localId kan hetzelfde zijn als het nr uit de brondata/schema, nu zonder alfanumeriek karakter, dit kan met een Rename.

5. De naam van het object moet worden aangegeven als siteName/GeographicalName. Hiervoor moeten een aantal attributen worden ingevuld:

Het attribuut language moet nld zijn

Voor het attribuut nameStatus moet een codelijst worden geïmporteerd, namelijk de codelijst Naamstatus onder Geografische namen.

6. De namespace moet gehaald worden uit het officiële namespace register van Geonovum op <http://inspirelab.geonovum.nl/namespaces/listnamespaces.php>. De namespace voor EZ zijn hier al aangemaakt. Om de namespace te vullen selecteer je in het doelschema namespace, druk op de blauwe pijl en gebruik de functie Assign (let op niet Assign (bound)). In het tweede opvolgende scherm dient de namespace met de hand ingevuld te worden.

Schermafdruck namespaces TBD



## Documentatie genereren

Alle voorgaande stappen worden door HALE vastgelegd. Deze rapportage dient als naslagwerk. HALE genereert een HTML+SVG document waarin de genomen stappen zijn gedocumenteerd. Het genereren van de rapportage gebeurt door naar:

File – Export – Alignment te gaan en daar de optie Mapping Documentation (HTML+SVG) te kiezen, kies een uitvoerlocatie en de rapportage wordt gegenereerd.

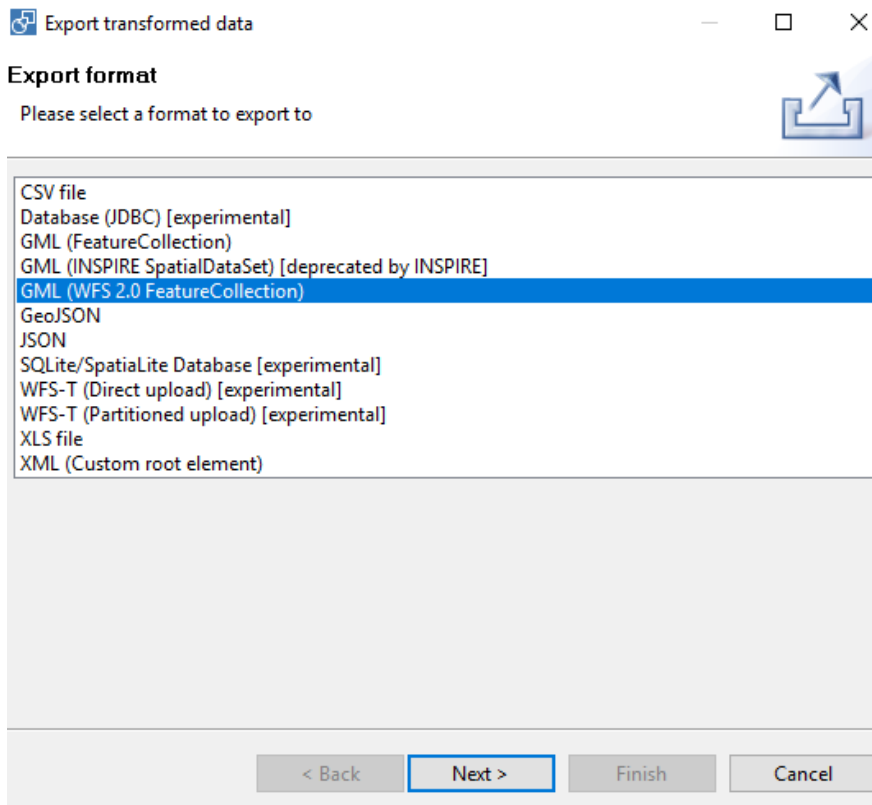
De html pagina is niet te openen met de huidige versie van Internet Explorer maar wel met Firefox en Chrome. In de html pagina is door te klikken op de rode en blauwe i knopjes meer informatie te krijgen.

Dat ziet er als volgt uit:

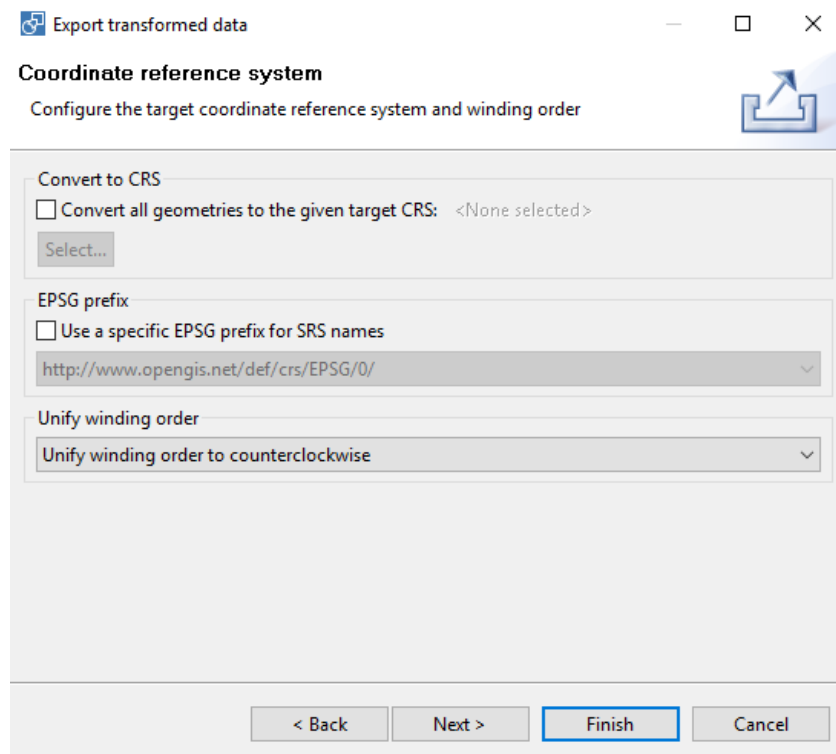


## INSPIRE GML bestand aanmaken

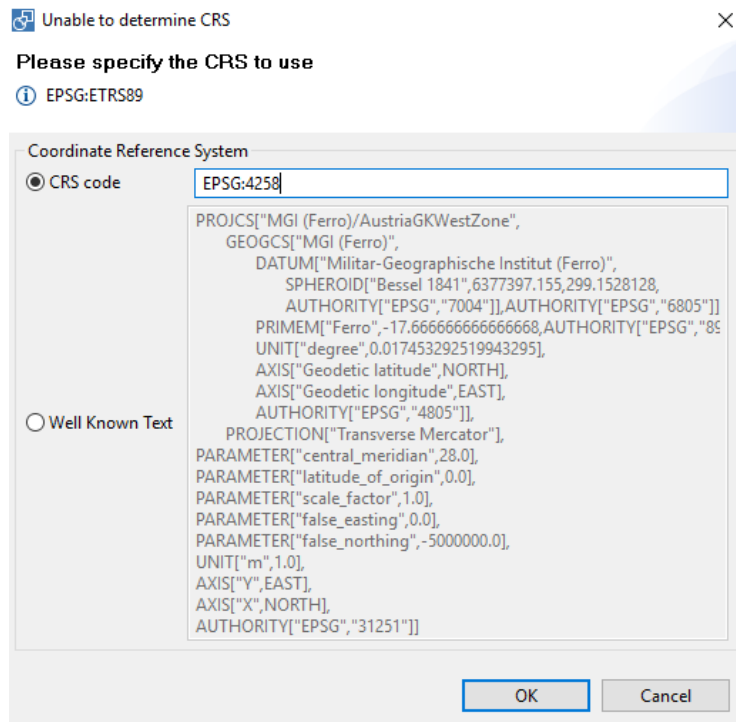
Het maken van een INSPIRE GML bestand gebeurt door in de menubalk naar File – Export – Transformed data te gaan en daar de optie GML (WFS2.0 Feature Collection) te selecteren



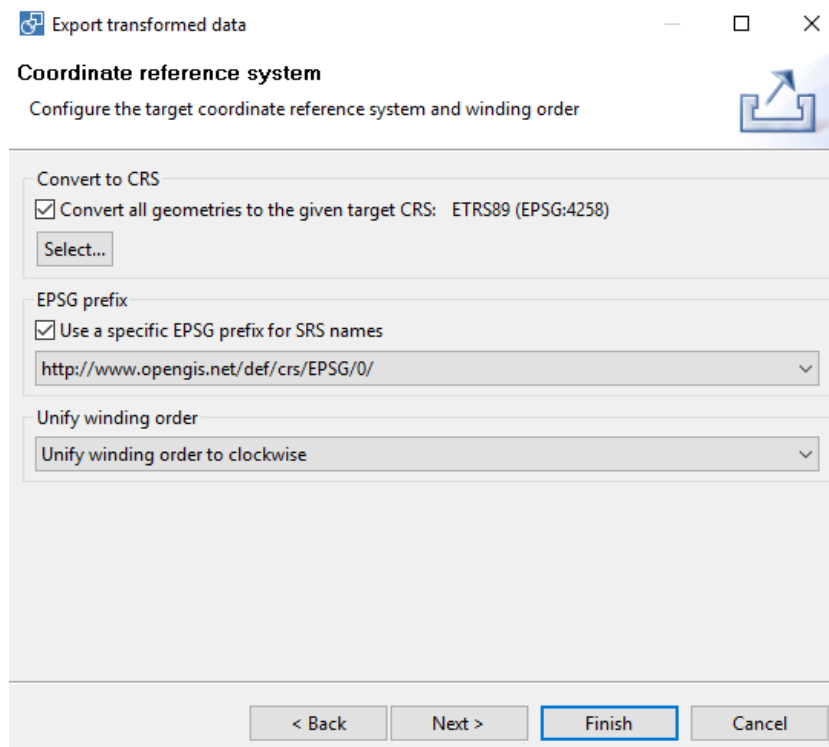
Geef in het daarop volgende scherm de uitvoer locatie en selecteer bij Validator de XML validator (Java XML API) en druk Next.



In dit scherm kun je de getransformeerde data converteren naar het gewenste CRS (Coördinatenstelsel). In INSPIRE verband is afgesproken dat alle data in ETRS89 formaat (EPSG:4258) worden aangeleverd. Deze kan hier ingevuld worden.

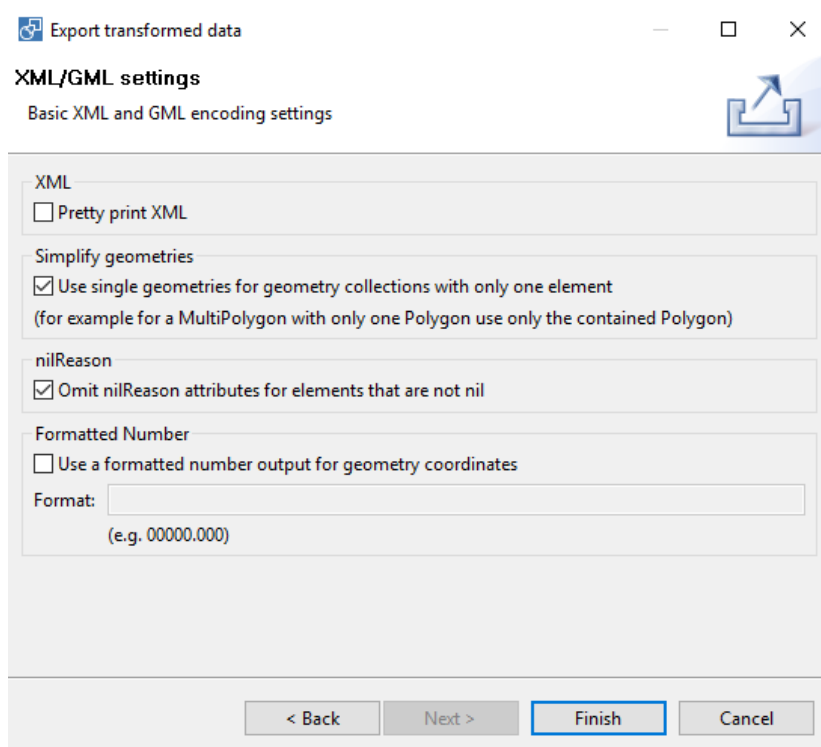


Druk op OK en je komt terug in het eerdere scherm.



De prefix moet aangevinkt zijn en de unify winding order moet altijd op clockwise staan.

Je kunt daarnaast, in het daaropvolgende scherm, nog een ATOM feed aanmaken maar dat wordt vooralsnog niet gedaan.

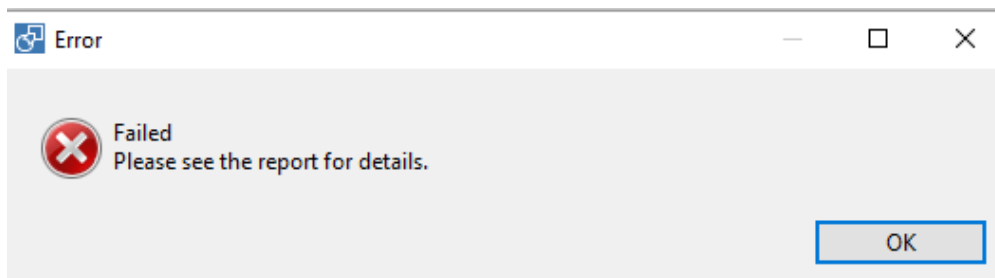


In het daaropvolgende scherm het vinkje zetten bij Pretty print XML zodat de resulterende GML makkelijk leesbaar wordt. De rest van de vinkjes laten staan.


De GML is klaar.

#### [Foutmeldingen bij export naar GML](#)

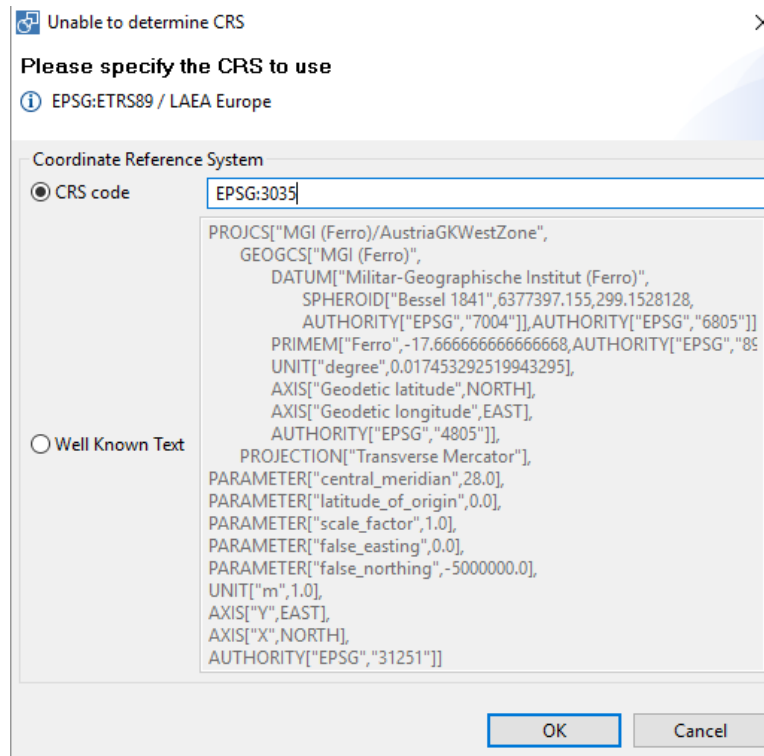
Als je bij de export naar GML de volgende foutmelding krijgt:



Dan kun je het rapport bekijken in de property lijst. Een van de foutmeldingen zou kunnen zijn:

 **Could not convert geometry to target CRS**

Deze foutmelding wordt veroorzaakt door de .prj file in de shapefile. Je kunt deze foutmelding oplossen door de .prj file te verwijderen of te renamen (voor ArcGIS heb je bijvoorbeeld deze .prj file wel nodig). Onthoud wel welke projectie de brondata heeft. De source data te verwijderen via File – Clean – Source data en daarna de bron data zonder de .prj file opnieuw te importeren. Omdat HALE geen projectie zit wordt je gevraagd deze handmatig in te vullen:



Is de brondata projectie WGS84 dan vul bij CRS code EPSG:4326 in, is de brondata projectie dan vul EPSG: 28992. Druk op OK en de brondata wordt geïmporteerd en de transformatie wordt direct uitgevoerd. Hierna kun je export naar GML weer uitvoeren.

### Inspectie van de GML

De GML kun je nu inlezen in QGIS zodat je de data kunt checken.

Omdat de GML complex kan zijn is het verstandig om de QGIS Complex GML plugin te gebruiken om de attributwaarden per object te kunnen lezen.

### Opslaan van het HALE project

Het HALE project kan worden opgeslagen op verschillende manieren afhankelijk van het gebruik. De transformatie kan namelijk op nieuwe datasets met dezelfde structuur gebruikt worden. De drie manieren om een HALE project op te slaan zijn:

1. Als **HALE project** (.hale) en bevat de project configuratie en de transformatie mapping in één bestand.
2. Als **HALE XML project** (.halex) en slaat het project op als een XML bestand met de transformatie mapping als een separaat bestand. Dit is de beste keus als je met versie controle werkt.
3. Als **HALE project archive** (.halez) en bevat het project, transformatie mapping en (bron)data als een ZIP archief. Deze manier sluit het beste aan als je gezamenlijk aan een project werkt.

Om het project op te slaan ga je naar de menubalk File – Save Alignment Project as... of naar File – Save Alignment Project, vul de gewenste uitvoerlocatie en de eventuele aanvullende gegevens in en het project wordt opgeslagen.