



HALE Studio

Michał Mackiewicz
www.gis-support.pl



MINISTERSTWO
ŚRODOWISKA



Sfinansowano ze środków
Narodowego Funduszu
Ochrony Środowiska
i Gospodarki Wodnej

1. Omówienie zasad przejścia ze struktury źródłowej do struktury docelowej, w szczególności w kontekście harmonizacji danych przestrzennych na potrzeby inicjatywy INSPIRE
2. Konfiguracja środowiska narzędzia i jego omówienie, w tym omówienie możliwości readerów i writerów, ograniczenia, porównanie do innych narzędzi typu ETL
3. Przygotowanie zbioru, przeprowadzenie harmonizacji próbki danych
4. Mapowanie danych z kilku źródeł jednocześnie - przygotowanie danych i struktury do takiej operacji.
5. Wykorzystanie struktury i danych zapisanych w geobazie (*.mdb)
6. Masowe przetwarzanie formatów danych, struktur danych oraz układu odniesienia
7. Generowanie plików XML/GML
8. Transformacje schematów danych do innych formatów
9. Tworzenie sparametryzowanych szablonów hale, które będą mapować i-ty zestaw danych do zawartego w szablonie schematu.
10. Publikacja danych
11. Rozszerzenia Hale Studio
12. Tworzenie projektów mapowania z parametrami, za pomocą Hale CLI

W trakcie szkolenia będą wykorzystane realne dane pochodzące z rejestrów publicznych:

Geoserwis GDOŚ

Mapa Podziału Hydrograficznego Polski

Digitální báze vodohospodářských dat (Republika Czeska)

Omówienie zasad przejścia ze struktury źródłowej do struktury docelowej, w szczególności w kontekście harmonizacji danych przestrzennych na potrzeby inicjatywy INSPIRE

Zapewnienie dostępu do danych przestrzennych w reprezentacjach, które umożliwiają łączenie tych danych w sposób spójny z innymi zharmonizowanymi danymi, korzystając z usług sieciowych oraz stosując wspólne specyfikacje produktów danych.

(INSPIRE Generic Conceptual Model)

Dyrektywa 2007/2/WE ustanawiająca infrastrukturę informacji przestrzennej we Wspólnocie Europejskiej (INSPIRE)

Rozporządzenie Komisji (UE) nr 1089/2010 z dnia 23 listopada 2010 r. w sprawie wykonania dyrektywy 2007/2/WE Parlamentu Europejskiego i Rady w zakresie interoperacyjności zbiorów i usług danych przestrzennych (Dz.U. L 323 z 8.12.2010, s. 11–102).

Rozporządzenie Komisji (UE) nr 102/2011 z dnia 4 lutego 2011 r. zmieniające rozporządzenie (UE) nr 1089/2010 w sprawie wykonania dyrektywy 2007/2/WE Parlamentu Europejskiego i Rady w zakresie interoperacyjności zbiorów i usług danych przestrzennych (Dz.U. L 31 z 5.2.2011, s. 13–34)

Rozporządzenie Komisji (UE) nr 1312/2014 z dnia 10 grudnia 2014 r. zmieniające rozporządzenie (UE) nr 1089/2010 w sprawie wykonania dyrektywy 2007/2/WE Parlamentu Europejskiego i Rady w zakresie interoperacyjności usług danych przestrzennych (Dz.U. L 354 z 11.12.2014, s. 8–16)

"Państwa członkowskie są zobowiązane do udostępniania, w ramach tej infrastruktury, zbiorów danych, które dotyczą jednego lub kilku załączników do dyrektywy 2007/2/WE, oraz odpowiednich usług danych przestrzennych zgodnie z rozwiązaniami technicznymi służącymi interoperacyjności, oraz, gdy jest to wykonalne, harmonizacji zbiorów i usług danych przestrzennych." (art. 7 Dyrektywy)

Pogrubionym pismem oznaczono pojęcia najważniejsze z punktu widzenia HALE.

Wszystkie definicje podano w ich dokładnym brzmieniu zgodnie z Rozporządzeniem Komisji (UE) nr 1089/2010 z dnia 23 listopada 2010 r. w sprawie wykonania dyrektywy 2007/2/WE Parlamentu Europejskiego i Rady w zakresie interoperacyjności zbiorów i usług danych przestrzennych

„typ abstrakcyjny” (abstract type) oznacza typ, w przypadku którego nie można tworzyć instancji, ale który może posiadać atrybuty i role asocjacji;

„rola asocjacji” (association role) oznacza wartość lub obiekt, z którymi typ ma związek (art. 8 ust. 2 lit. b) dyrektywy 2007/2/WE);

„**atrybut**” (**attribute**) oznacza cechę typu (art. 8 ust. 2 lit. c) dyrektywy 2007/2/WE);

„proponowany typ” (candidate type) oznacza typ, który jest już stosowany w ramach specyfikacji tematu danych przestrzennych wymienionego w załączniku I do dyrektywy 2007/2/WE, ale który zostanie w pełni określony w temacie danych przestrzennych wymienionym w załączniku II lub III do dyrektywy 2007/2/WE, do którego należy pod względem tematycznym;

„**lista kodowa**” (**code list**) oznacza otwarte wyliczenie, które może być rozszerzane;

„**typ danych**” (**data type**) oznacza deskryptor zbioru wartości, które wymagają opisanie, zgodnie z ISO 19103;

„**wyliczenie**” (**enumeration**) oznacza typ danych, w którym instancje tworzą stałą listę wartości wymienionych literalnie. Atrybuty wyliczenia mogą przyjmować jedynie wartości z tej listy;

„**zewnętrzny identyfikator obiektu**” (**external object identifier**) oznacza jednoznaczny identyfikator obiektu publikowany przez odpowiedzialny organ, możliwy do zastosowania przez aplikacje zewnętrzne celem odniesienia do obiektu przestrzennego;

„**identyfikator**” (**identifier**) oznacza niezależną językowo sekwencję znaków umożliwiającą identyfikowanie w sposób jednoznaczny i trwały tego, z czym jest powiązana, zgodnie z EN ISO 19135;

„**tworzyć instancję**” (**instantiate**) oznacza tworzyć obiekt zgodny z definicją, atrybutami, rolami asocjacji i ograniczeniami przyporządkowanymi typowi, dla którego instancja jest tworzona;

„**warstwa**” (**layer**) oznacza podstawową jednostkę informacji geograficznej, która może być uzyskiwana jako mapa z serwera zgodnie z EN ISO 19128;

„**informacje na temat cyklu życia**” (**life-cycle information**) oznaczają zbiór własności obiektu przestrzennego opisujących czasowe cechy wersji obiektu przestrzennego lub zmiany między wersjami;

„element metadanych” (metadata element) oznacza dyskretną jednostkę metadanych, zgodnie z EN ISO 19115;

„pakiet” (package) oznacza mechanizm ogólnego stosowania na potrzeby organizowania elementów w grupy

„rejestr” (register) oznacza zbiór plików zawierający identyfikatory przypisane do rejestrowanych elementów, wraz z opisami powiązanych elementów, zgodnie z EN ISO 19135;

„**typ obiektu przestrzennego**” (spatial object type) oznacza kategorię obiektów przestrzennych;

„styl” (style) oznacza przyporządkowanie sparametryzowanych symboli stosowanych przy sporządzaniu map do typów obiektów przestrzennych, ich własności i ograniczeń;

„podtyp” (sub-type of) oznacza związek między typem bardziej szczegółowym a typem bardziej ogólnym, gdzie typ bardziej szczegółowy jest w pełni zgodny z typem bardziej ogólnym oraz zawiera dodatkowe informacje (na podstawie ISO 19103);

„**typ**” (**type**) oznacza typ obiektu przestrzennego lub typ danych;

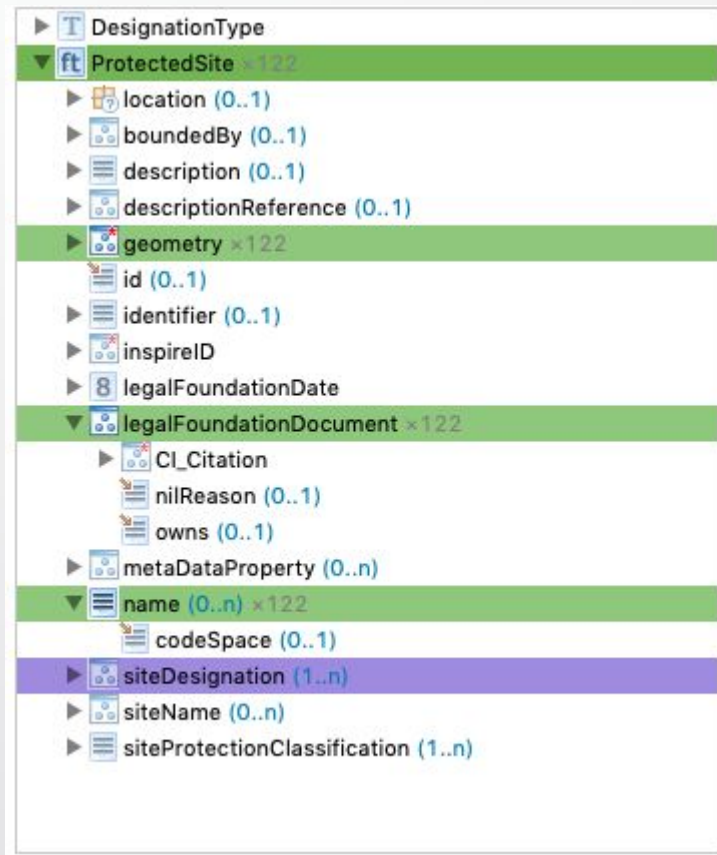
„**voidable**” (**zastępowalne przez void**) oznacza, że atrybutowi lub roli asocjacji może zostać przypisana wartość „void” (nieokreślona), jeżeli zbiory danych przestrzennych utrzymywane przez państwa członkowskie nie zawierają odpowiedniej wartości, bądź jeżeli nie jest możliwe wyprowadzenie odpowiedniej wartości z wartości istniejących przy rozsądnych kosztach. Jeżeli atrybut lub rola asocjacji nie są zastępowalne przez void, komórkę tabeli określającą „voidability” (zastępowalność przez void) pozostawia się pustą.

- Oprogramowanie utworzone w ramach projektu HUMBOLDT - pierwotna nazwa **HUMBOLDT Alignment Editor** - w celu ułatwienia przekształcania danych do schematów INSPIRE
- Napisane w języku Java, możliwe do uruchomienia we wszystkich głównych systemach operacyjnych (Mac, Linux, Windows)
- Aktualnie rozwijane przez firmę WeTransform (<http://wettransform.to>)
- Kod jest dostępny na licencji open source - GNU Lesser General Public License (LGPL) v3.0
- Dostępne jest wsparcie komercyjne od WeTransform
- Pobieranie wersji instalacyjnej - <https://www.wettransform.to/downloads/>, kodu źródłowego - <https://github.com/halestudio/hale>
- Dokumentacja <http://help.halestudio.org/latest/index.jsp>
- Szkolenie jest przygotowane dla wersji 3.5 - wydane 11 stycznia 2019

Schemat (**Schema**) jest opisem struktury danych. Wyróżniane są 2 typy schematów:

- Logical Schema - schemat konkretnego zbioru danych (pliku, dokumentu XML, bazy danych)
- Conceptual Schema - model koncepcyjny, opisuje dane bez wskazania konkretnej implementacji (np. model UML)

Do przeprowadzenia transformacji (harmonizacji) wymagane jest posiadanie Logical Schema.



W procesie transformacji w HALE Studio schemat może być:

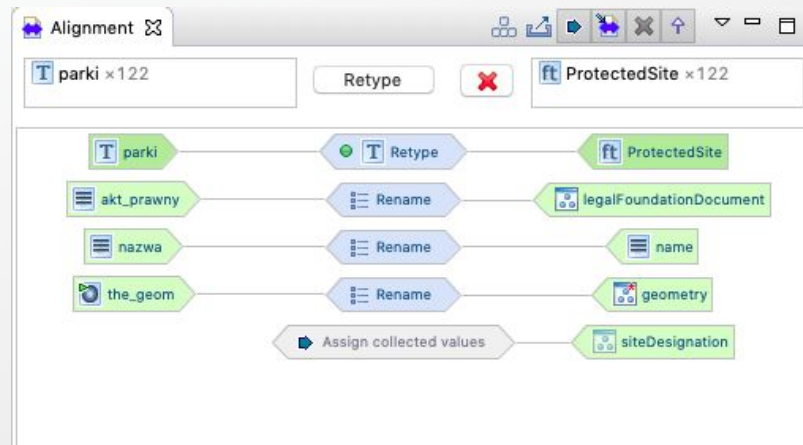
- zaimportowany z pliku definicji (np. XSD)
- zaimportowany z gotowych szablonów - **presets** - HALE posiada wbudowane gotowe definicje dla schematów aplikacyjnych INSPIRE
- odtworzony na podstawie istniejących danych (np. pliku SHP, bazy danych)

Dla projektu HALE musi zostać zdefiniowany schemat dla danych wejściowych (Source Schema) i wynikowych (Target Schema) by można było dokonać transformacji.

Parametry transformacji są określane jako **Alignment** i definiują zestaw przekształceń pomiędzy schematem źródłowym a wynikowym.

Pojedynczy **Alignment** składa się ze zbioru obiektów typu **Mapping cell**, które definiują przejście pomiędzy typami lub atrybutami wejściowymi i wynikowymi.

Zestaw parametrów transformacji oraz wskaźników do danych wejściowych i wynikowych nazywa się **Alignment project** i może być zapisany na dysku w formie pliku - analogicznie do projektów QGIS.



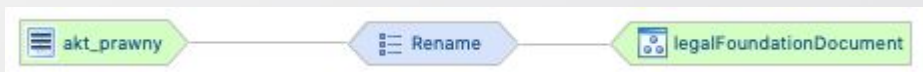
Mapping cell składa się z bytu wejściowego, bytu wyjściowego oraz funkcji przekształcenia.

Przykłady:

Funkcja **Retype** przenosi typ wejściowy "parki" na typ wyjściowy "ProtectedSite"
(funkcja **Retype** przeznaczona jest do mapowania **typów danych**)



Funkcja **Rename** przenosi wartość atrybutu "akt_prawny" na atrybut "legalFoundationDocument"
(funkcja **Rename** jest przeznaczona do mapowania **atrybutów**)



Proces harmonizacji danych w HALE składa się z następujących etapów:

- Import schematu danych źródłowych
- Import schematu danych wynikowych
- Import próbki danych źródłowych
- Identyfikacja typów danych wynikowych, które mogą być utworzone z danych źródłowych
- Stworzenie relacji pomiędzy typami wejściowymi a wynikowymi
- Identyfikacja i konfiguracja relacji pomiędzy wejściowymi a wynikowymi atrybutami
- Weryfikacja poprawności transformacji
- Eksport danych wynikowych

Konfiguracja środowiska narzędzia i jego omówienie, w tym omówienie możliwości readerów i writerów, ograniczenia, porównanie do innych narzędzi typu ETL)

HALE Studio jest programem napisanym w Javie i jako taki wymaga wirtualnej maszyny Java do działania.

Dystrybucja dla systemu Windows zawiera już wbudowaną maszynę wirtualną, więc nie ma potrzeby instalacji. W systemach MacOS i Linux wymagane jest Java JDK (nie JRE) w wersji min. 8.

Szczegóły instalacji są podane w zeszycie ćwiczeń.

Extract, Transform and Load - proces pozyskania danych dla baz danych, składający się z:

- pozyskania danych ze źródeł zewnętrznych,
- przekształcenia danych,
- załadowania danych do bazy docelowej.

ETL jest szerszym pojęciem od harmonizacji, wykorzystywany jest np. do migracji danych pomiędzy systemami, przy aktualizacji, do zasilania systemów klasy Business Intelligence.

Inne oprogramowanie typu ETL, które może być użyte do danych przestrzennych to na przykład:

- GDAL / OGR (open source)
- GeoKettle (open source)
- FME (komercyjny)

HALE Studio obsługuje następujące formaty wejściowe **schematów**:

- XSD
- własny format Hale Schema Definition (w wariantach JSON i XML)
- CSV
- XLS i XLSX
- MDB
- SQLite
- ESRI Shapefile
- WFS DescribeFeatureType
- PostGIS
- Microsoft SQL Server

HALE Studio obsługuje następujące formaty wejściowe **danych**:

- XML i GML - także skompresowane programem gzip (to co innego, niż najpopularniejszy ZIP)
- CSV
- XLS i XLSX
- ESRI Shapefile
- MDB
- SQLite
- WFS GetFeature
- PostGIS
- Microsoft SQL Server

HALE Studio obsługuje następujące formaty wynikowe **schematów**:

- deegree
- Hale Schema Definition (XML)
- Hale Schema Definition (JSON)

HALE nie umożliwia obsługi schematów koncepcyjnych, jak UML. Taką funkcjonalność miało narzędzie HUMBOLDT Conceptual Schema Transformer - niestety usługa sieciowa zakończyła działalność i narzędzie już nie jest dostępne.

HALE Studio obsługuje następujące formaty wynikowe **danych**:

- CSV
- PostGIS
- Microsoft SQL Server
- GML
- GeoJSON
- SQLite
- XLS
- XML użytkownika
- WFS-T

HALE vs inne ETL

Cecha	HALE Studio	GeoKettle	FME	OGR
Wbudowane schematy INSPIRE	TAK	NIE	TAK	NIE
Możliwość odczytu i zapisu z i do WFS	TAK	NIE	TAK	TAK
Dostępna usługa chmurowa	TAK - do przechowywania projektów i publikacji danych (hale connect)	NIE	TAK - do przeprowadzania transformacji w chmurze (FME Cloud)	NIE
Dostępne formaty danych	10 formatów	45 formatów	335 formatów	45 formatów
Dostępne funkcje geoprocessingu	6 funkcji	27 funkcji	80 funkcji	4 funkcje
Cena	0	0	2000 €	0
Główne zastosowanie	Harmonizacja danych INSPIRE	Migracja danych przestrzennych	Migracja i przetwarzanie danych przestrzennych	Konwersja formatów danych przestrzennych

Przygotowanie zbioru, przeprowadzenie harmonizacji próbki danych

Ćwiczenie "Harmonizacja próbki danych" będzie polegało na wykonaniu harmonizacji dla danych o obszarach Natura 2000 pochodzących z Geoserwisu GDOŚ.

Metodyka harmonizacji została oparta o publikację:

Krawczyk A., Garguła A., 2018, Harmonization of Polish Natura2000 data sets with the protected sites data schema of inspire directive in the environment of Humboldt Alignment Editor (HALE), Geoinformatica Polonica 17:7-15
DOI: 10.4467/21995923GP.18.001.9158

Interfejs HALE

The screenshot displays the HALE software interface, which is used for aligning and transforming spatial data. The interface is divided into several main sections:

- Source Panel:** Labeled "schemat wejściowy" (input schema).
- Target Panel:** Labeled "schemat wynikowy" (output schema).
- Alignment Panel:** Labeled "mapowanie" (mapping).
- Workspace Log:** A table at the bottom left showing the sequence of operations performed during the session.
- Report List:** A table at the bottom right showing the results of various validation and transformation tasks.

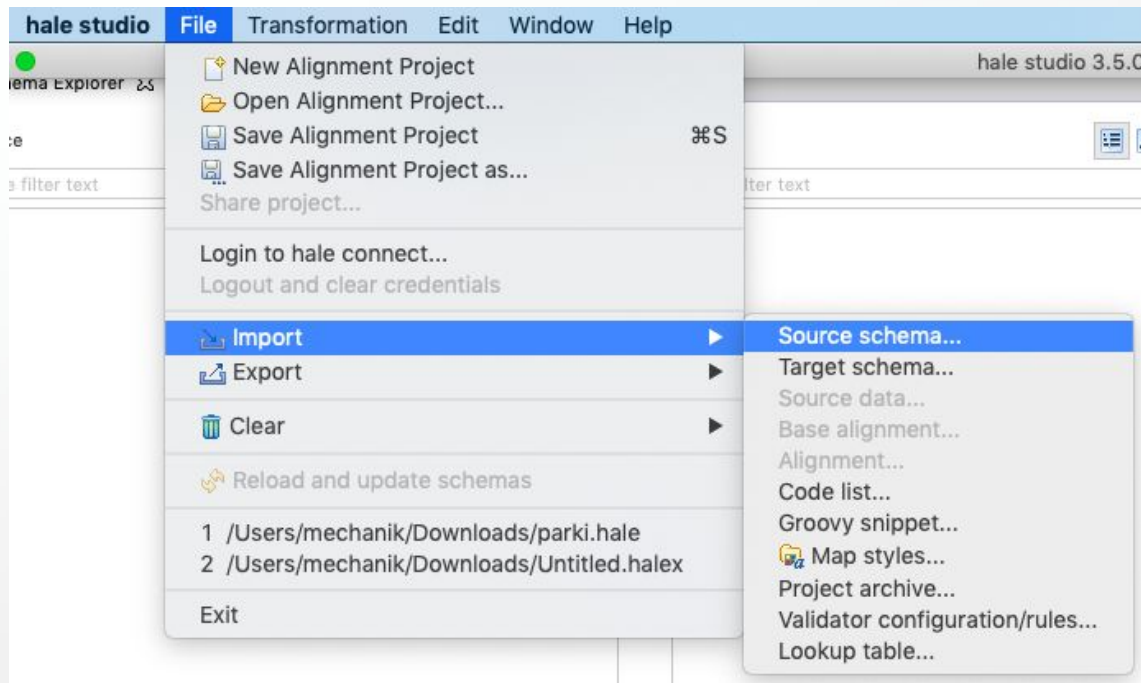
Workspace Log Table:

Message	Plug-in	Date
Instance transformation	eu.esdihumboldt.hale.ui	8/29/19, 4:35 PM
[16:35:57] Instance transformation - Finished successfully	eu.esdihumboldt.hale.co...	8/29/19, 4:35 PM
Styled Layer Descriptor import	eu.esdihumboldt.hale.co...	8/29/19, 4:35 PM
[16:35:48] Finished task 'Load styles from SLD'.	eu.esdihumboldt.hale.co...	8/29/19, 4:35 PM
[16:35:48] Starting task 'Load styles from SLD'...	eu.esdihumboldt.hale.co...	8/29/19, 4:35 PM
hale alignment import	eu.esdihumboldt.hale.co...	8/29/19, 4:35 PM
[16:35:47] Finished task 'Load hale alignment'.	eu.esdihumboldt.hale.co...	8/29/19, 4:35 PM
[16:35:47] Starting task 'Load hale alignment'...	eu.esdihumboldt.hale.co...	8/29/19, 4:35 PM
Shapefile import	eu.esdihumboldt.hale.ui	8/29/19, 4:35 PM
[16:35:46] Created cached schema representation	eu.esdihumboldt.hale.co...	8/29/19, 4:35 PM

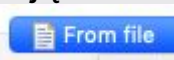
Report List Table:

Task	Time
Instance validation	16:35:58
Instance transformation	16:35:56
Load data into database	16:35:53
Shapefile import	16:35:53
Shapefile import	16:35:43
XML schema import	16:35:31
hale project import	16:35:29
13:42 2019-08-27	
09:49 2019-08-27	
13:52 2019-08-26	

Import -> Source schema




Schemat wejściowy może być zadany poprzez definicję schematu XSD lub istniejący zbiór danych (plik lub baza danych). W przypadku obszarów Natura 2000 rozpowszechnianych w formie plików SHP, właściwą opcją będzie import z pliku -






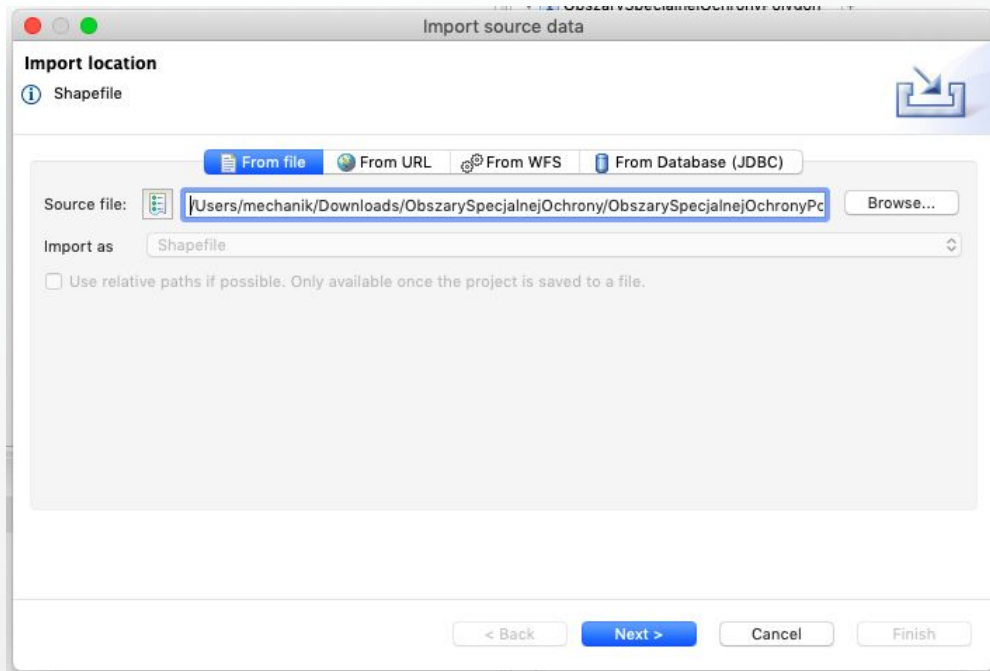
Import SHP jako schematu powoduje załadowanie wyłącznie schematu- same dane nie zostaną wczytane.

Symbol  oznacza atrybut tekstowy

 oznacza atrybut liczbowy

 oznacza geometrię

Import danych



Import danych musi być wykonany po wczytaniu schematu wejściowego. Domyślnie importowana jest próbka 500 pierwszych obiektów. Wielkość próbki można ustawić w oknie Window -> Preferences -> Project -> Source data.


Dostępne są 2 widoki dla danych, widok tabeli

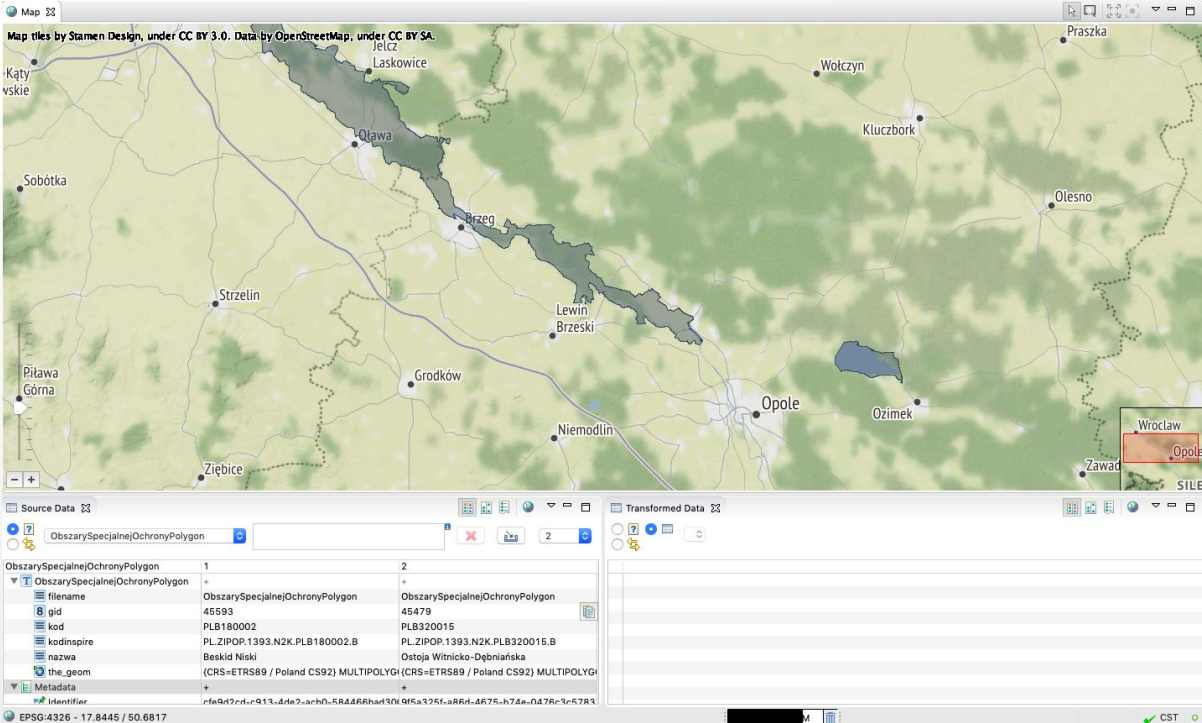
The screenshot displays a data inspection application with four main panels:

- Alignment:** Contains two input fields labeled "<Click to select>" and buttons for "Select cell" and a red "X" icon.
- Properties:** Displays the message "Properties are not available."
- Source Data:** Shows a table view for the dataset "ObszarySpecjalnejOchronyPolygon". The table has two columns, 1 and 2, and several rows of data. A small icon in the top right corner of this panel is highlighted by a line pointing to a separate icon in the top right of the slide.
- Transformed Data:** Currently empty.

The data in the 'Source Data' table is as follows:

	1	2
ObszarySpecjalnejOchronyPolygon	+	+
ObszarySpecjalnejOchronyPolygon	+	+
filename	ObszarySpecjalnejOchronyPolygon	ObszarySpecjalnejOchronyPolygon
gid	45593	45479
kod	PLB180002	PLB320015
kodinspire	PL.ZIPOP.1393.N2K.PLB180002.B	PL.ZIPOP.1393.N2K.PLB320015.B
nazwa	Beskid Niski	Ostoja Witnicko-Dębniańska
the_geom	{CRS=ETRS89 / Poland CS92} MULTIPOLYGON	{CRS=ETRS89 / Poland CS92} MULTIPOLYGON
Metadata	+	+
Identifier	cfe9d2cd-c913-4de2-acb0-584466bad301	9f5a325f-a86d-4675-b74e-0476c3c5783

oraz widok mapy  górnym rogu ekranu.

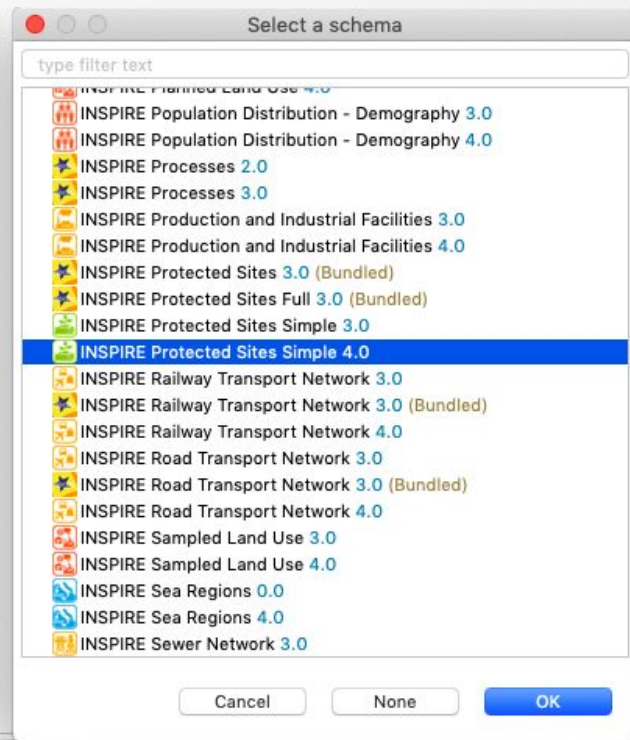
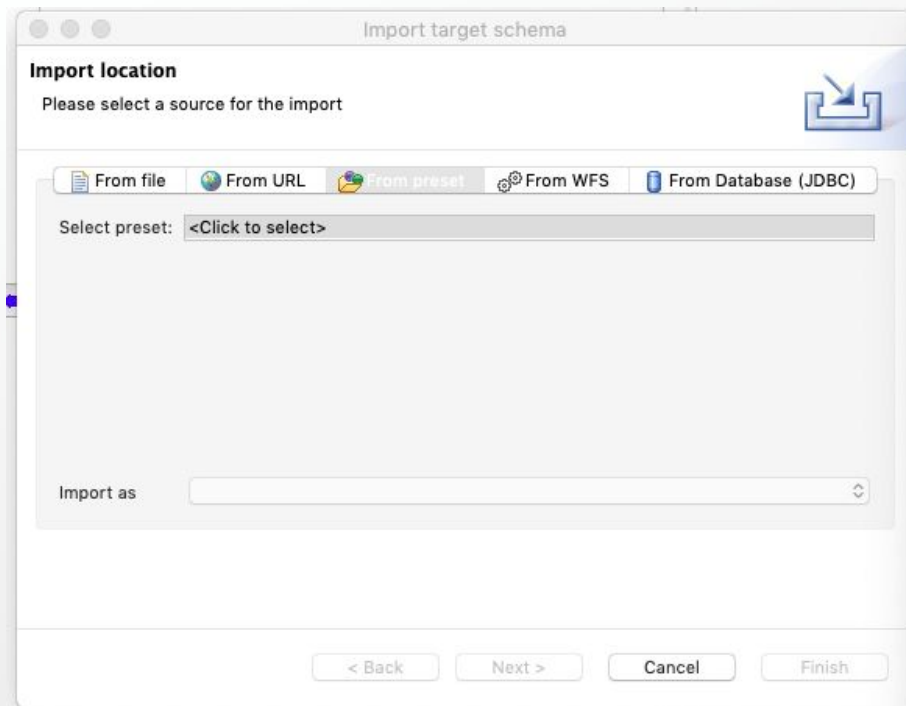


Prezentacja atrybutów przykładowego obiektu źródłowego

▼ T ObszarySpecjalnejOchronyPolygon	+
☰ kod	PLB180002
☰ filename	ObszarySpecjalnejOchronyPolygon
8 gid	45593
☰ kodinspire	PL.ZIPOP.1393.N2K.PLB180002.B
☰ nazwa	Beskid Niski
🕒 the_geom	{CRS=ETRS89 / Poland CS92} MULTIPOLYGON (((669315.801783
▼ 📄 Metadata	+
🔗 Identifier	cfe9d2cd-c913-4de2-acb0-584466bad300

File -> Import -> Target Schema

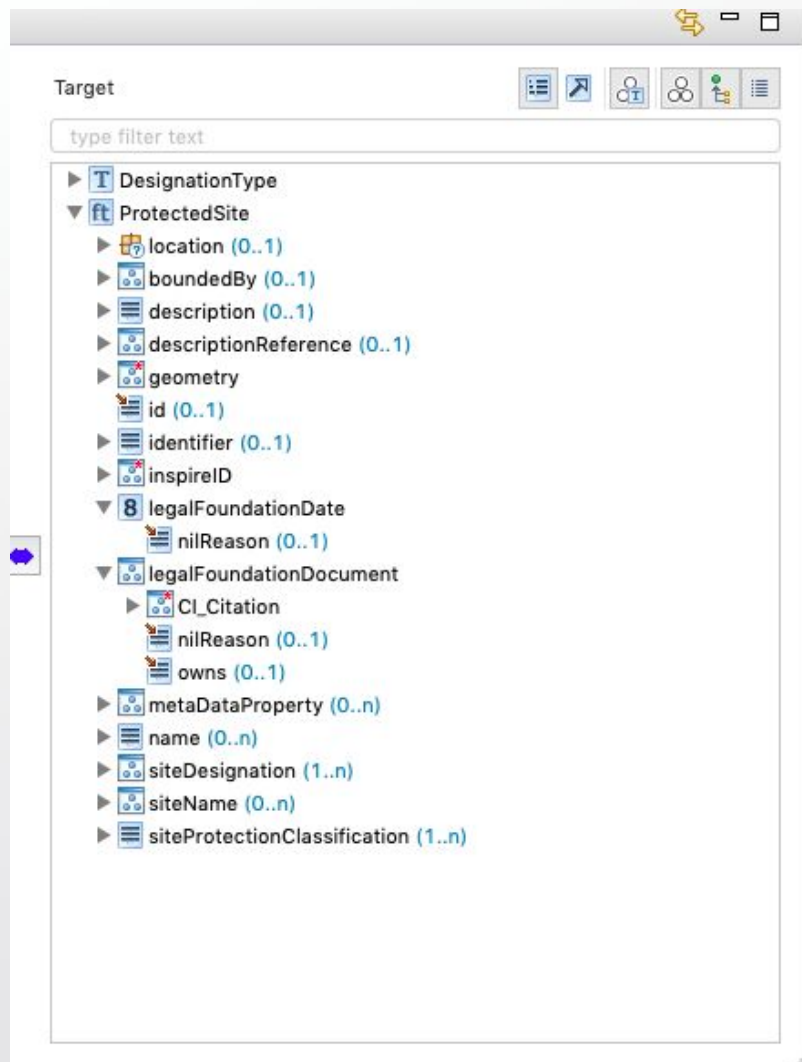
Schematy INSPIRE są dostępne w zakładce  From preset



Schemat "INSPIRE Protected Sites Simple 4.0"

Schematy INSPIRE są schematami tzw. Complex Features, co oznacza, że atrybuty mogą być zagnieżdżone - atrybut może zawierać swoje własne atrybuty.


Taka struktura jest prosta w realizacji w formatach XML (GML) i JSON (GeoJSON), ale nie pasuje dobrze do plików płaskich np. SHP i relacyjnych baz danych.

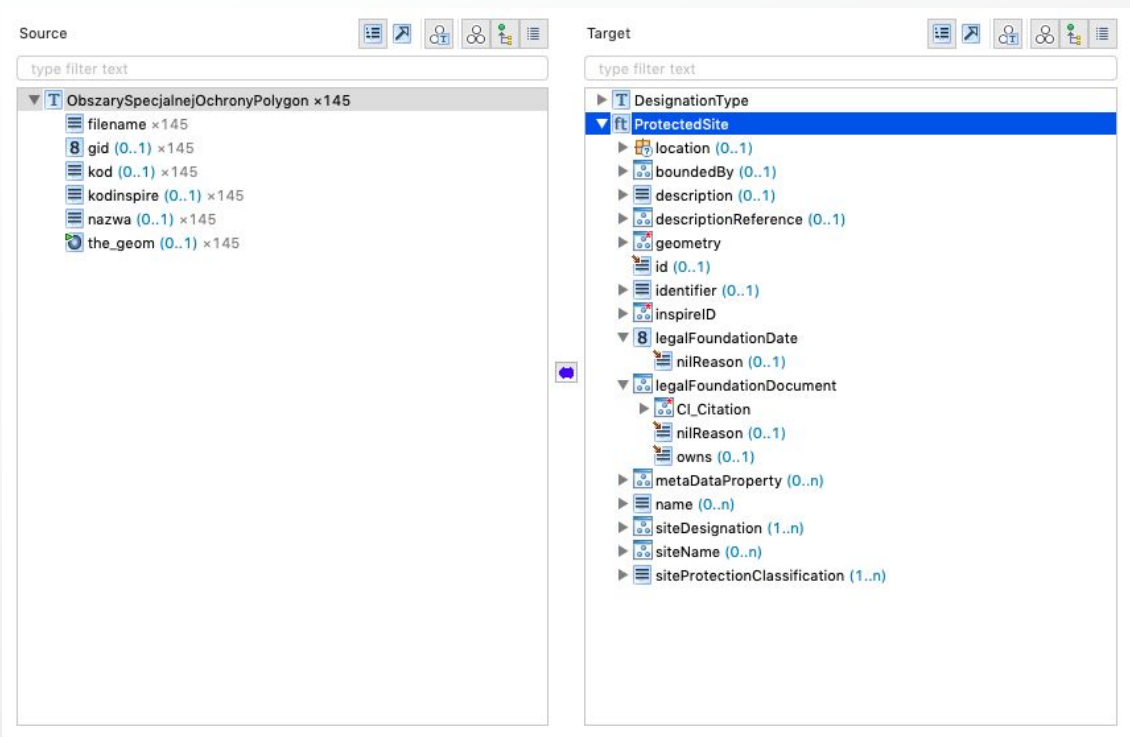


Po wczytaniu schematu wejściowego i wynikowego, można przystąpić do przygotowania mapowania (**Alignment**).

Wczytanie danych źródłowych nie jest obowiązkowe, ale umożliwia podgląd wyników i ewentualnych błędów na bieżąco.

Pierwszym krokiem jest mapowanie typów danych. Najprostszym przykładem jest przeniesienie jednego zbioru wejściowego w jeden zbiór zharmonizowany - zostanie w tym celu użyta funkcja **Retype**.

Należy kliknąć typ źródłowy w panelu **Source**, typ wynikowy w panelu **Target** i kliknąć przycisk  , wybrać **Retype**.

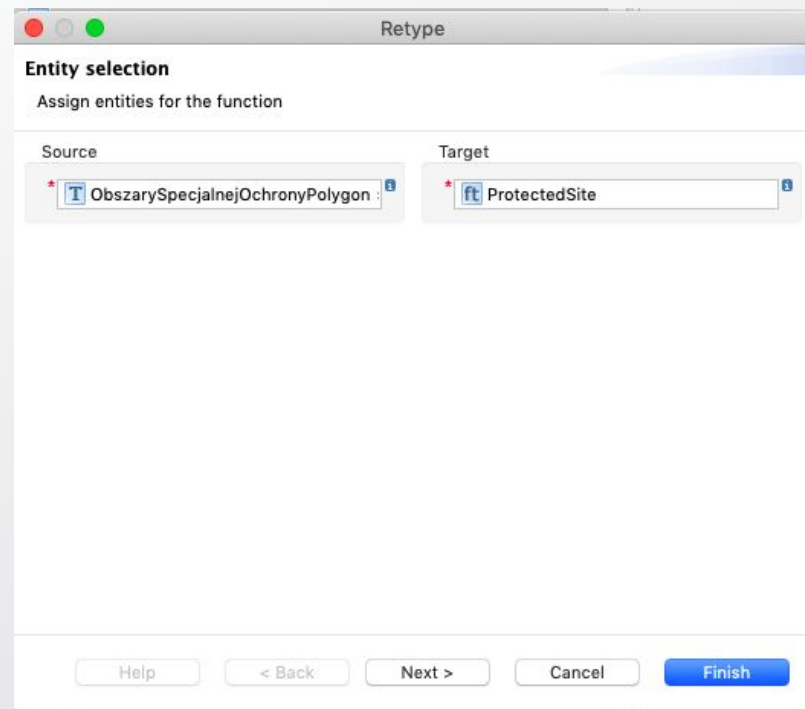


Funkcja **Retype** dokonuje mapowania jednego typu wejściowego na jeden typ wynikowy. Inne opcje to:

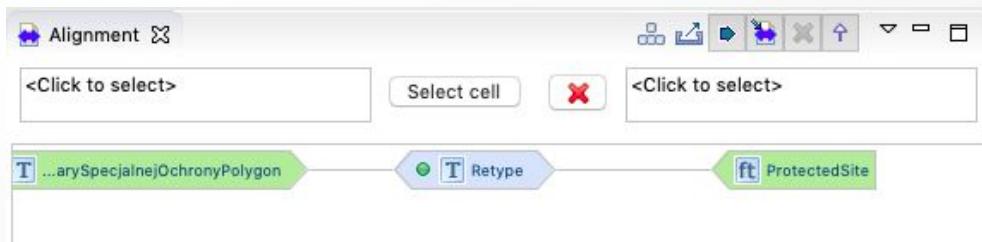
Groovy Retype - mapowanie 1:1 wspomagane skryptami w języku Groovy

Groovy Merge - scalenie n:1 wspomagane skryptami w języku Groovy

Merge - scalenie n:1




Dodanie funkcji **Retype** dodaje **Mapping Cell** do tabeli mapowań - można wówczas rozpocząć dodawanie kolejnych mapowań dla atrybutów.

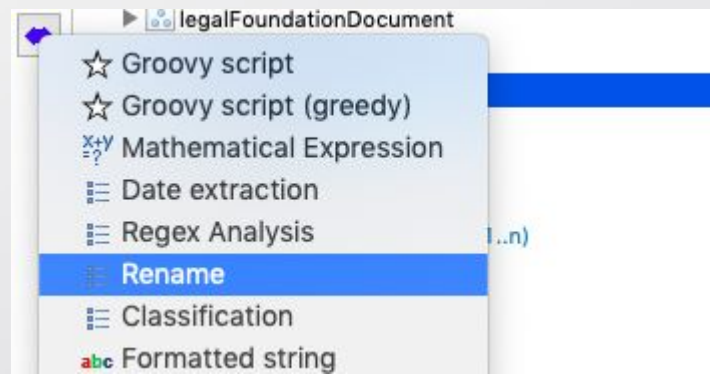
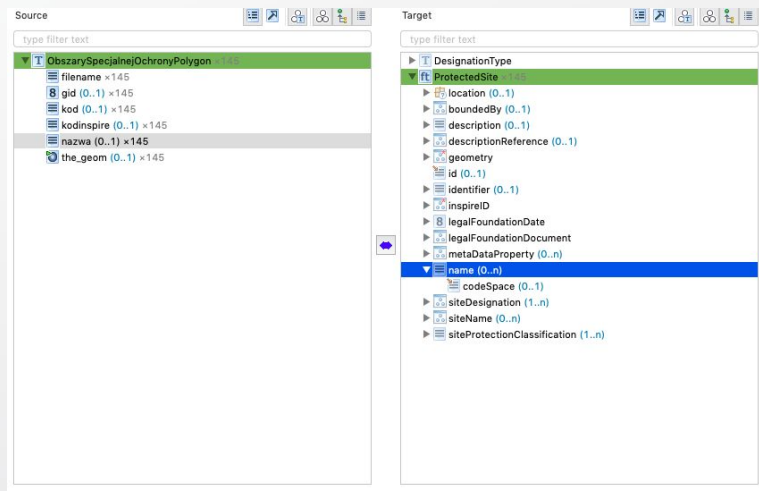


Mapowanie atrybutów może być:

- z schematu wejściowego do wynikowego
- przypisanie wartości stałej do schematu wynikowego

Mapowanie atrybutów polega na wyborze atrybutu w schemacie źródłowym, następnie wyborze atrybutu w schemacie wynikowym, kliknięciu  i wyborze funkcji mapowania.

Najprostszą funkcją jest **Rename**, czyli przeniesieniu wartości atrybutu A w schemacie wejściowym w niezmienionej formie do atrybutu B w schemacie wynikowym. Przykład - "nazwa" na "name"



Po każdej edycji mapowania następuje wykonanie transformacji i jej walidacja. Błędy walidacji można przeczytać, klikając dwukrotnie w

Instance validation

The screenshot shows a software interface with two main windows. The 'Instance validation' window on the left displays a tree view of warnings. The 'Report List' window on the right shows a list of tasks and their completion times.

Instance validation window:

- Report
- Warnings
 - ProtectedSite (290 warnings)
 - geometry (145 warnings)
 - Cardinality (showing 5 of 145)
 - 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
 - inspireID (145 warnings)

Report List window:

Task	Time
16:59 2019-08-29	
Instance validation	18:44.39
Instance transformation	18:44.38
Instance validation	18:36.30
Instance transformation	18:36.29
XML schema import	18:18.00
Load data into database	18:06.16
Shapefile import	18:06.16
Shapefile import	17:58.47
16:35 2019-08-29	
13:42 2019-08-27	
09:49 2019-08-27	

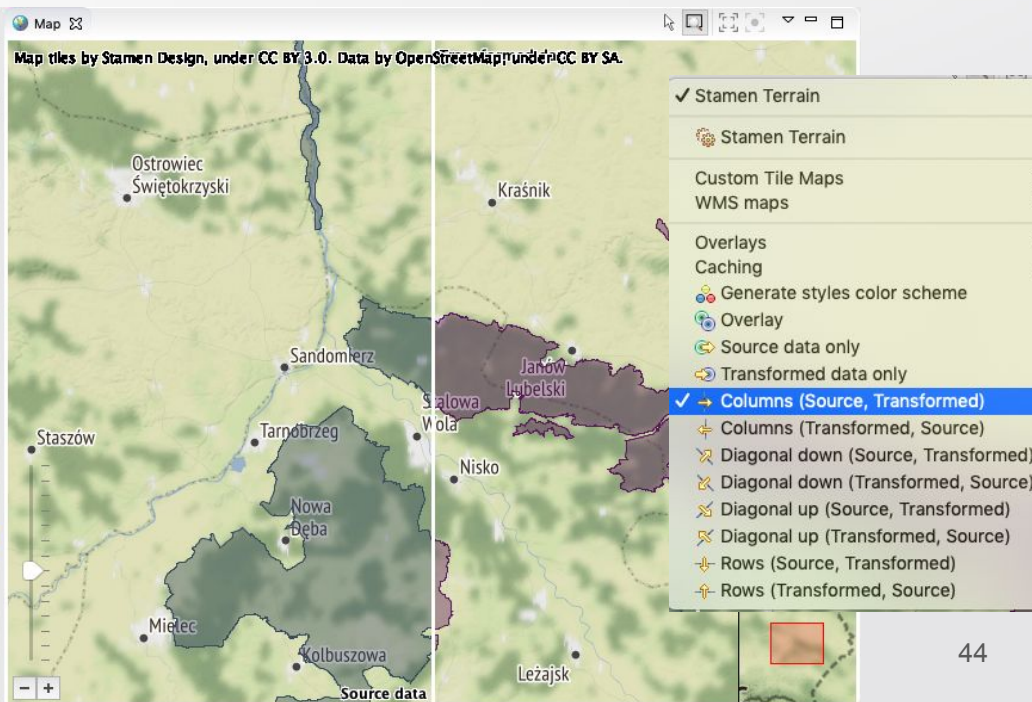
Minimalny zestaw mapowań, jaki pozwala na poprawne zaliczenie walidacji to:




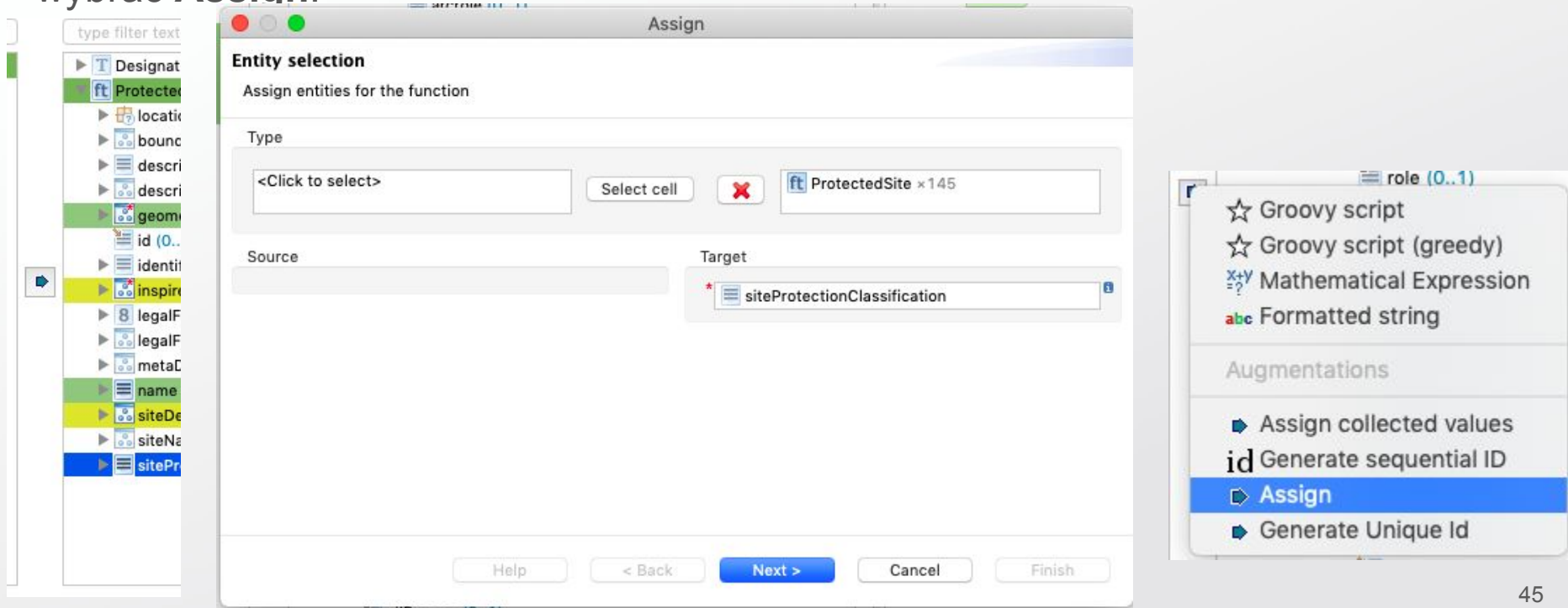
Szczegółowy wykaz znajduje się w zeszycie ćwiczeń.

Ustawiono podgląd mapy w kolumnach - po lewej stronie dane oryginalne, po prawej zharmonizowane.

Transformed Data					
ProtectedSite					
ProtectedSite	1	2	3	4	5
ProtectedSite	+	+	+	+	+
location	no value	no value	no value	no value	no value
boundedBy	no value	no value	no value	no value	no value
description	no value	no value	no value	no value	no value
description...	no value	no value	no value	no value	no value
geometry	+	+	+	+	+
id	no value	no value	no value	no value	no value
identifier	no value	no value	no value	no value	no value
inspireID	+	+	+	+	+
Identifier	+	+	+	+	+
loc...	PLB180002	PLB320015	PLB320016	PLB080001	PLB300013
na...	PL.ZIPOP.1393.N2	PL.ZIPOP.1393.N2K.F	PL.ZIPOP.1393.N2K.F	PL.ZIPOP.1393.N2K.F	PL.ZIPOP.1393.N2K.F
ver...	no value	no value	no value	no value	no value
legalFoun...	no value	no value	no value	no value	no value
legalFoun...	no value	no value	no value	no value	no value
metaData...	no value	no value	no value	no value	no value
name	Beskid Niski	Ostoja Witnicko-Dębna	Lasy Puszczy nad D...	Puszcza Barlinecka	Dolina Samicy
siteDesig...	no value	no value	no value	no value	no value
siteName	no value	no value	no value	no value	no value
siteProte...	no value	no value	no value	no value	no value
Metadata	+	+	+	+	+
Identifier	4352ea16-8ca8-417	340eb289-63a0-453	cd250a4c-5771-4bf1	3e1e0b0a-248e-4af4	dd833627-62a0-41e
SourceID	cfe9d2cd-c913-4de2	9f5a325f-a86d-4675	dc06c9c8-2c26-4e3	c7c7d5ce-b902-409	1fb63cb6-33dd-4aee

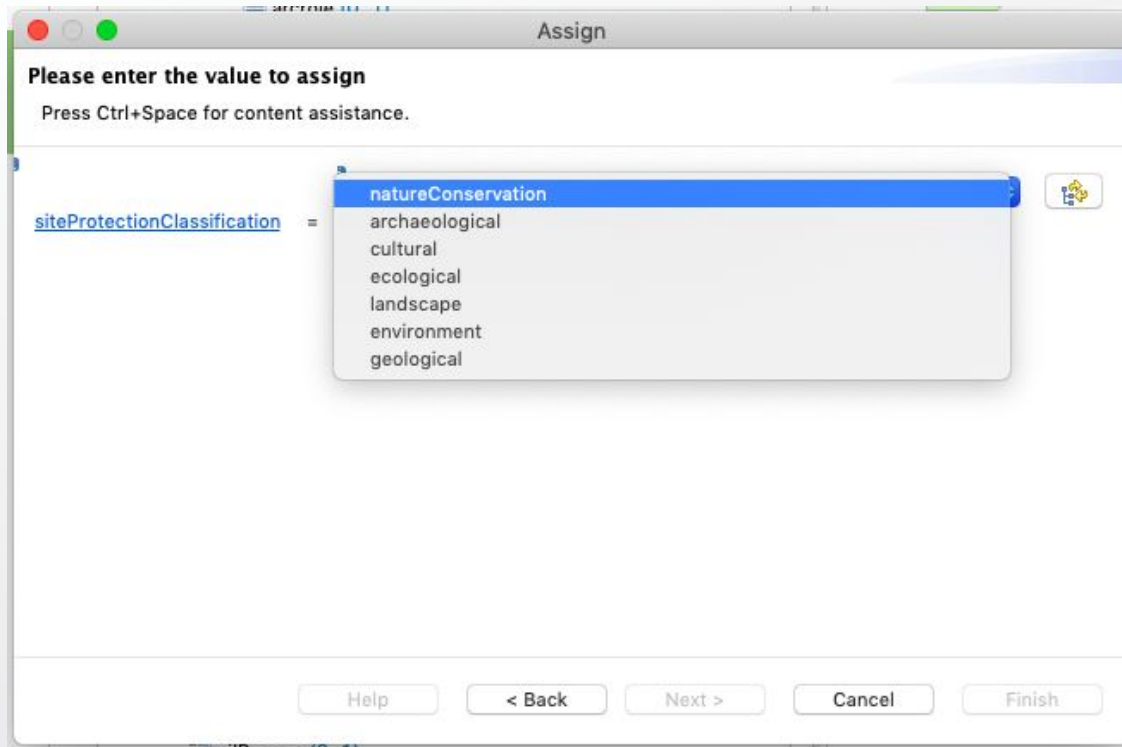


Aby przypisać wartość stałą do danych wynikowych, należy zaznaczyć atrybut w schemacie wynikowym, następnie kliknąć  między panelami **Source** i **Target** i wybrać **Assign**.



The screenshot shows the 'Assign' dialog box in a data mapping tool. The dialog has a title bar 'Assign' and a 'Entity selection' section with the text 'Assign entities for the function'. Below this, there is a 'Type' field with a dropdown menu showing '<Click to select>'. To the right of the 'Type' field is a 'Select cell' button and a red 'X' button. The 'Source' field is empty. The 'Target' field is set to 'siteProtectionClassification'. On the right side of the dialog, there is a list of 'Augmentations' with the following items: 'Groovy script', 'Groovy script (greedy)', 'Mathematical Expression', 'Formatted string', 'Assign collected values', 'Generate sequential ID', 'Assign' (highlighted in blue), and 'Generate Unique Id'. At the bottom of the dialog, there are buttons for 'Help', '< Back', 'Next >', 'Cancel', and 'Finish'.

Dla atrybutów, dla których wartości są typu **enumeration** (niezmienny słownik) HALE pokaże listę dopuszczalnych wartości.

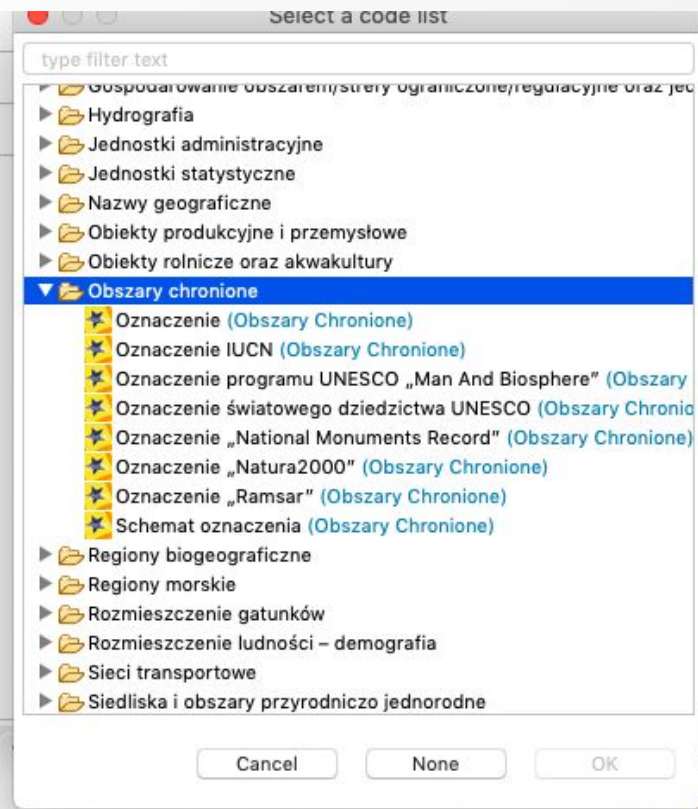
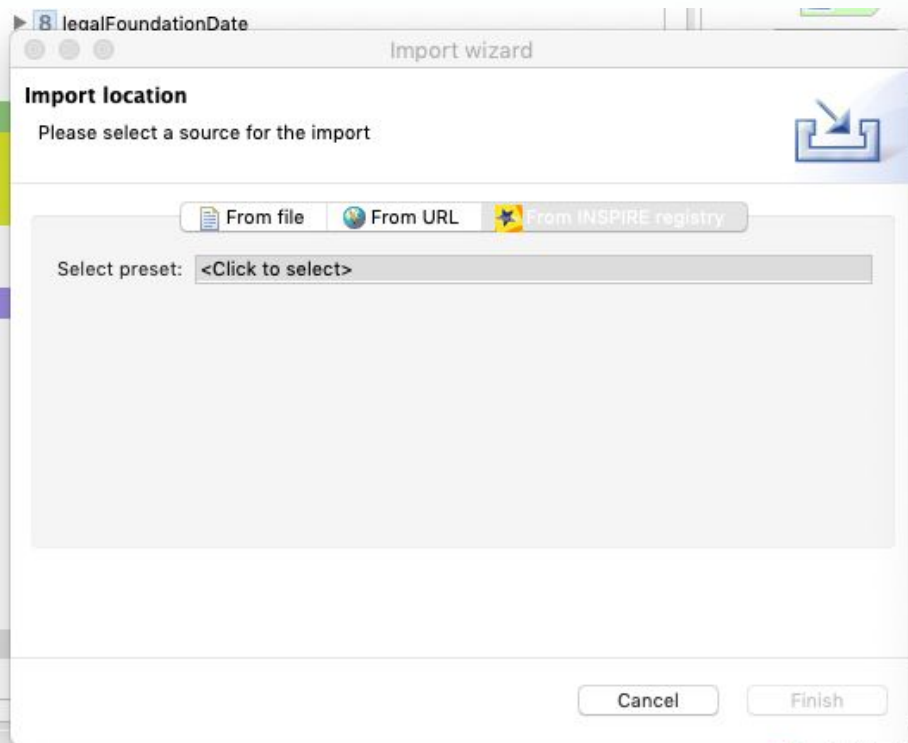


W przypadku atrybutów, gdzie lista wartości to **Enumeration** (katalog zamknięty), lista możliwych wartości jest dostępna od razu w HALE.

Dla katalogów otwartych (**code list**) konieczny jest wcześniejszy import listy z rejestru INSPIRE. Import odbywa się poprzez narzędzie **Import -> Code list -> From INSPIRE Registry**.

Gdy do danych wynikowych będzie przypisana jedna, stała wartość z listy kodowej należy użyć **Assign**, natomiast w przypadku mapowania słownika wartości na listę kodową należy użyć funkcji **Classification**.

Import list kodowych

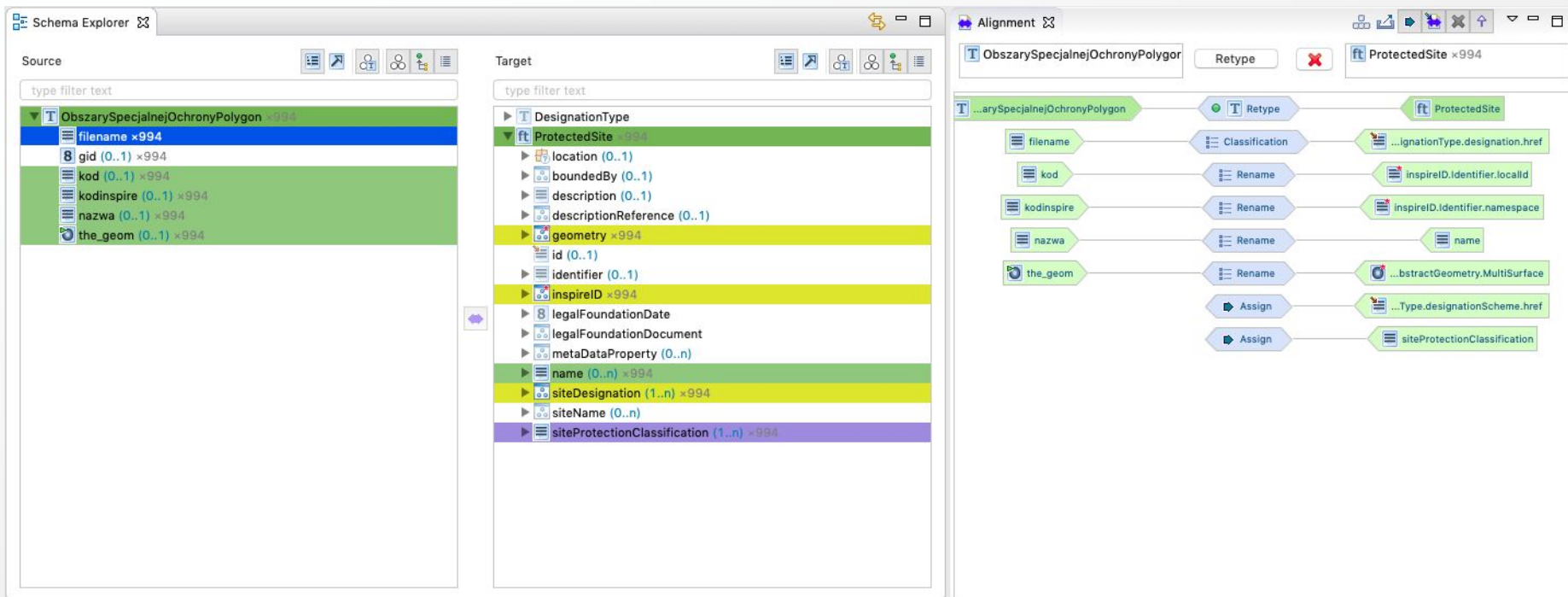


Mapowanie danych z kilku źródeł jednocześnie - przygotowanie danych i struktury do takiej operacji.

- jeden schemat INSPIRE, jedna klasa INSPIRE, wiele plików źródłowych o identycznym schemacie
 - wówczas wystarczy importować schemat źródłowy jeden raz i kolejne dane dodawać za pomocą funkcji **Import source data**
- jeden schemat INSPIRE, jedna klasa INSPIRE, wiele plików źródłowych o różnym schemacie
 - wówczas należy importować schemat wejściowy dla każdego źródła danych (**Import source schema**) i wykonać mapowanie **Retype** dla każdego źródła osobno, wybierając ten sam schemat docelowy
- jeden schemat INSPIRE, wiele klas INSPIRE, wiele plików źródłowych
 - wówczas należy wykonać osobne mapowania dla każdej relacji.

Przypadek 1

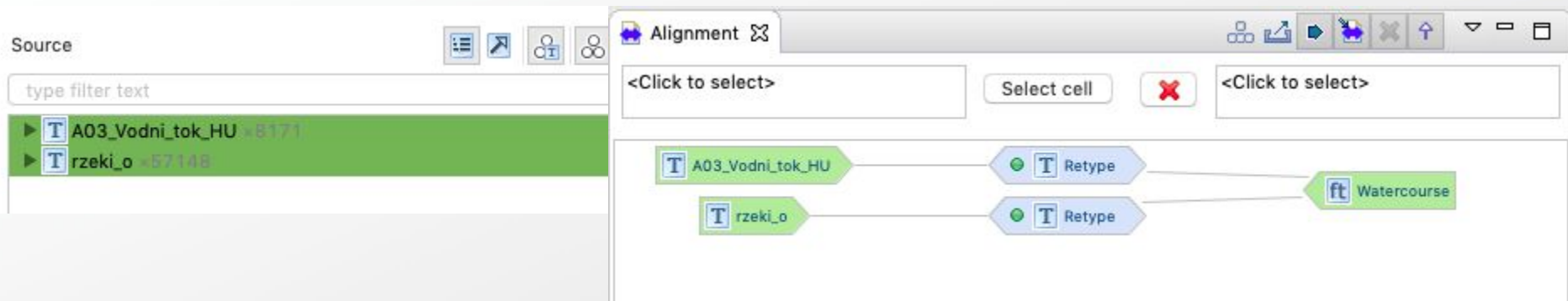
Mapowanie dwóch plików SHP z obszarami OSO i SOO na klasę INSPIRE **ProtectedSite**



Mapowanie prawie nie różni się od mapowania dla pojedynczego pliku - jedyna różnica to wykorzystanie atrybutu **filename**

Przypadek 2

Mapowanie dwóch plików SHP (czeska i polska sieć rzeczna) na jedną klasę INSPIRE **Watercourse**



Występują 2 schematy źródłowe i dwie operacje **Retype** kierujące dane do tej samej klasy

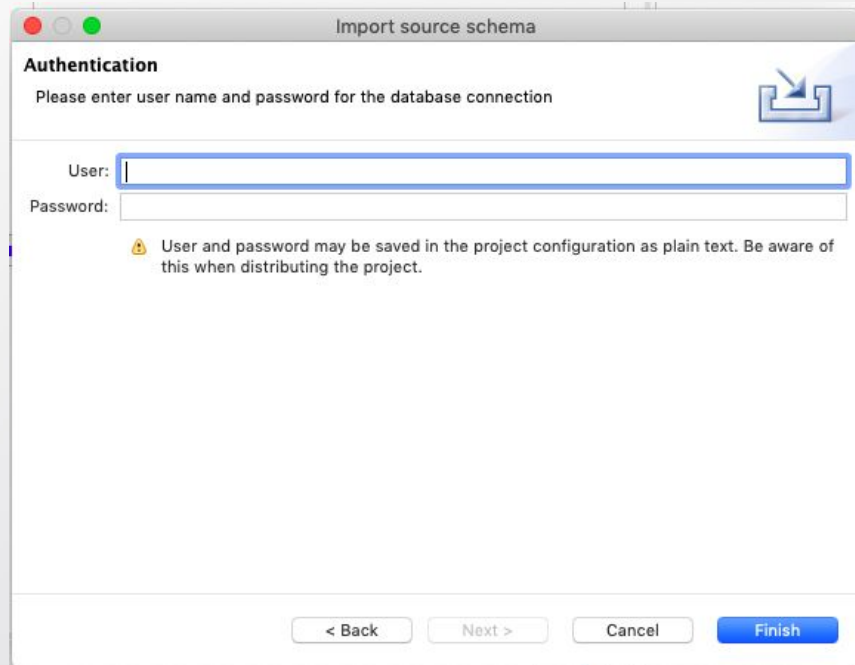
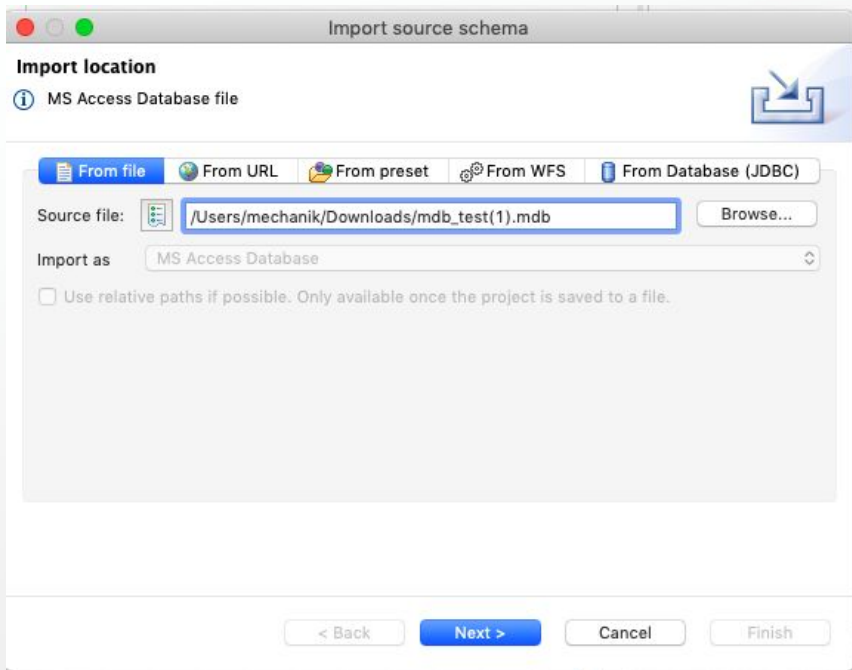
Mapowanie plików "rzeki_o.shp" na klasę **Watercourse**, "zlew_el.shp" na **DrainageBasin** oraz "jeziora.shp" na klasę **StandingWater**



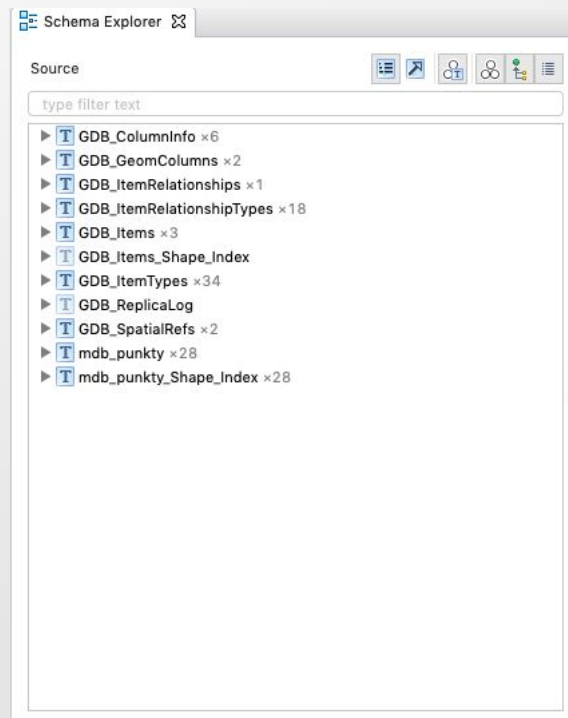
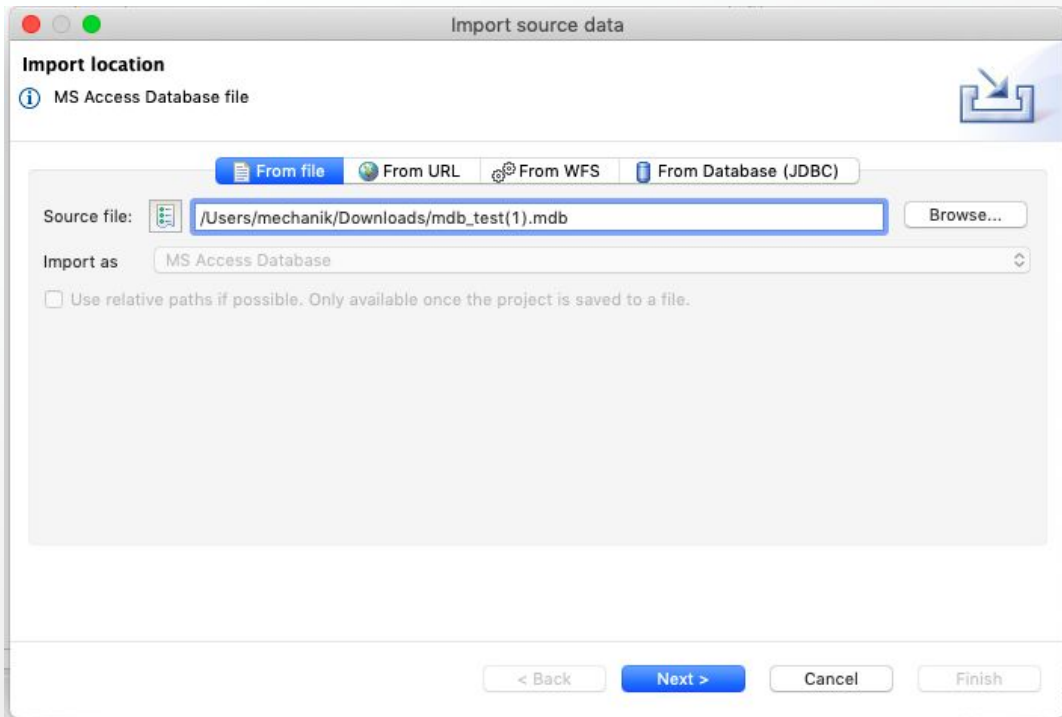
Występują 3 schematy wejściowe i 3 klasy wynikowe

Wykorzystanie danych i struktury zapisanych w bazach danych

Import struktury i danych z geobazy osobistej MDB odbywa się podobnie do importu schematu z plików SHP, to jest przez narzędzie **Import -> Source schema -> From file**. Jediną różnicą jest dialog podania loginu i hasła, które należy pozostawić puste.



Dane należy załadować narzędziem **Import -> Source data -> From file.**



Bazy SQLite i SpatiaLite teoretycznie są obsługiwane przez HALE, w praktyce nie udało się importować żadnego rzeczywistego przykładu.

Import source schema

Please select a source for the import

From file From URL From preset From WFS **From Database (JDBC)**

Driver PostgreSQL/PostGIS

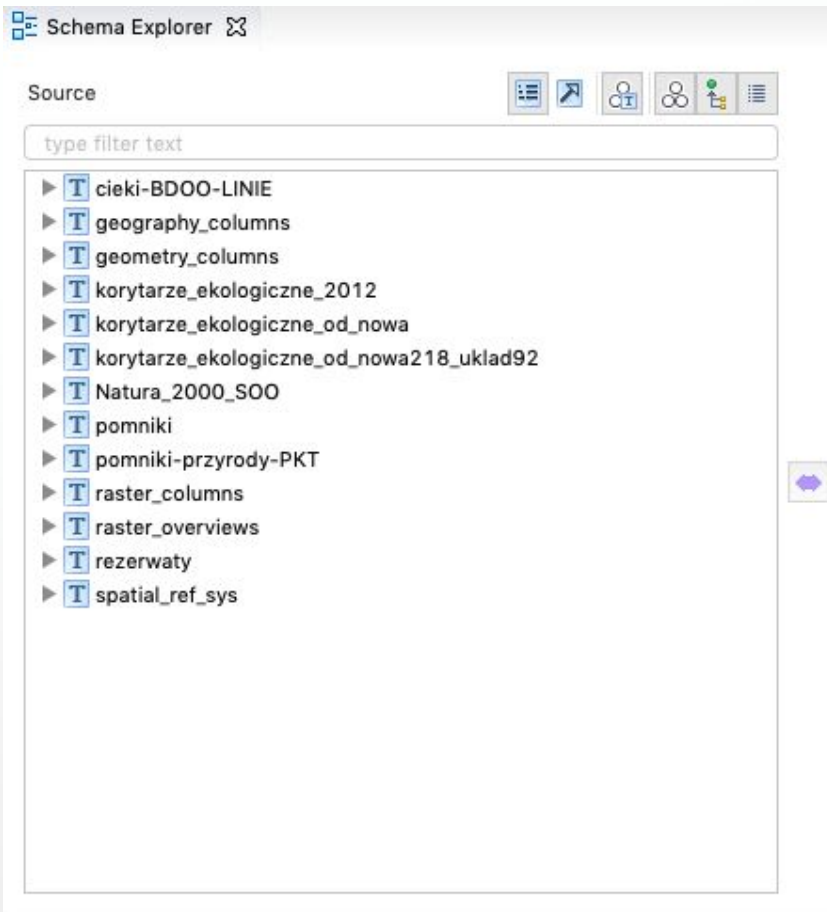
Host(:Port) dragon.gis-support.pl:5432

Database inspire

Import as Database Schema

< Back **Next >** Cancel Finish

Do importu struktury i danych z bazy PostGIS (lub Microsoft SQL Server) służy opcja **From Database (JDBC)**.



Każda tabela z bazy jest importowana jako jedna klasa źródłowa.

Dalszy proces mapowania przebiega identycznie jak w przypadku mapowania z danych zapisanych w pliku.

Generowanie plików GML i XML

File -> Export -> Transformed data

GML (FeatureCollection)

GML (Inspire SpatialDataSet) - przestarzały

GML (WFS 2.0 FeatureCollection)

XML użytkownika



Do wyboru z listy elementów XML
OpenGIS



Opcje obowiązkowe:

- wybór pliku docelowego

Export destination

 GML file 

Target file:

Validation

Validator	Status
Project validator	OK

- wybór układu współrzędnych, kolejności wierzchołków, prefiksu dla definicji układu współrzędnych


Convert to CRS

☒ Convert all geometries to the given target CRS:

EPSG prefix

☐ Use a specific EPSG prefix for SRS names

Unify winding order



Opcje dodatkowe:

- pretty print - formatowanie drzewa XML (znaki nowej linii, wcięcia)
- redukcja 1-elementowych geometrii Multi do prostych geometrii
- ominięcie atrybutu nilReason (przyczyna braku danych) dla niepustych atrybutów
- wybór formatowania wartości dziesiętnych

XML

☐ Pretty print XML

Simplify geometries

☒ Use single geometries for geometry collections with only one element
(for example for a MultiPolygon with only one Polygon use only the contained Polygon)

nilReason

☒ Omit nilReason attributes for elements that are not nil

Output formatting

☐ Use a formatted number output for geometry coordinates

Format:
(e.g. 00000.000)

☐ Use a formatted output for decimal values

Format:
Test: 123456789.6543 will be represented as 1.234567896543E8
(e.g. use 0.000## to write at least 3 and at most 5 decimal places)

Partycjonowanie danych - zapis wyniku do wielu plików GML

Dostępne opcje to:

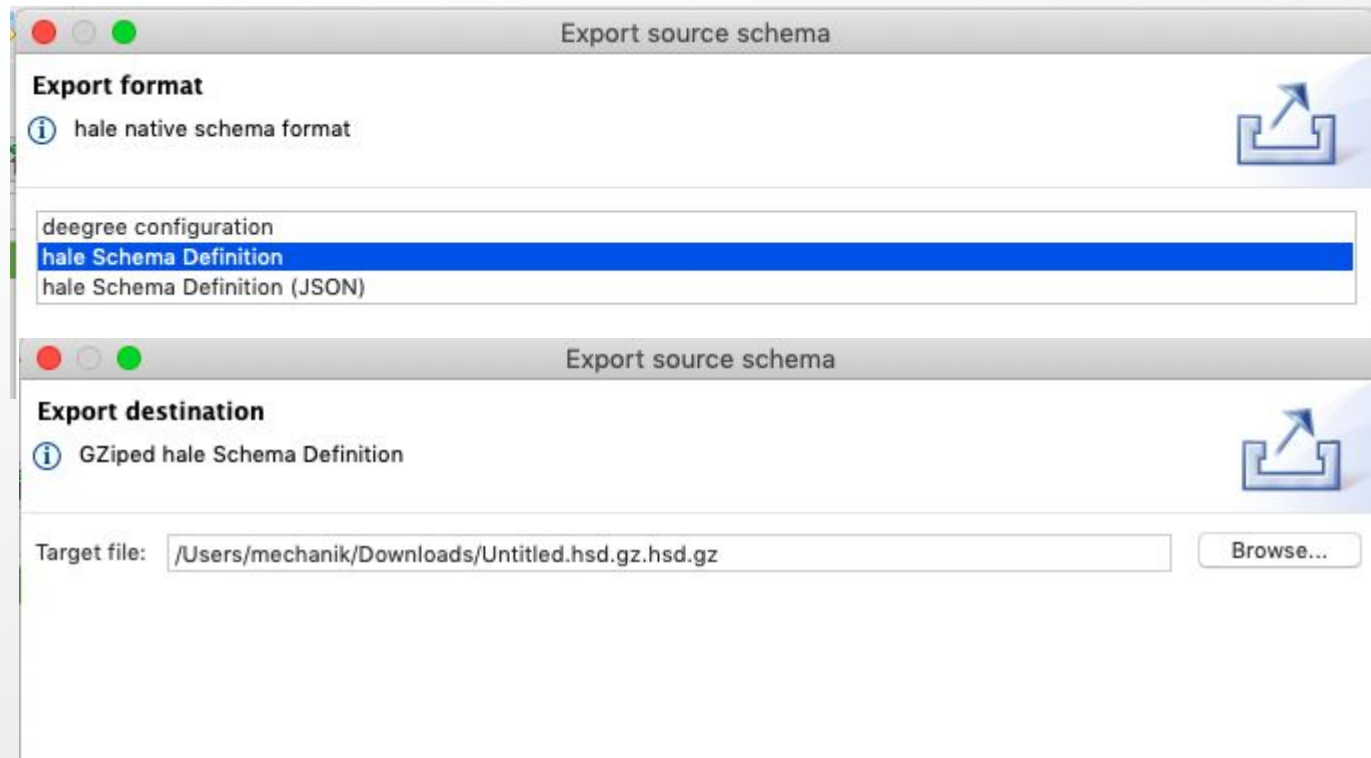
- podział co każde N obiektów
- podział na klasy obiektów
- podział przestrzenny (na arkusze)



Transformacje schematów danych do innych formatów

File -> Export -> Source schema

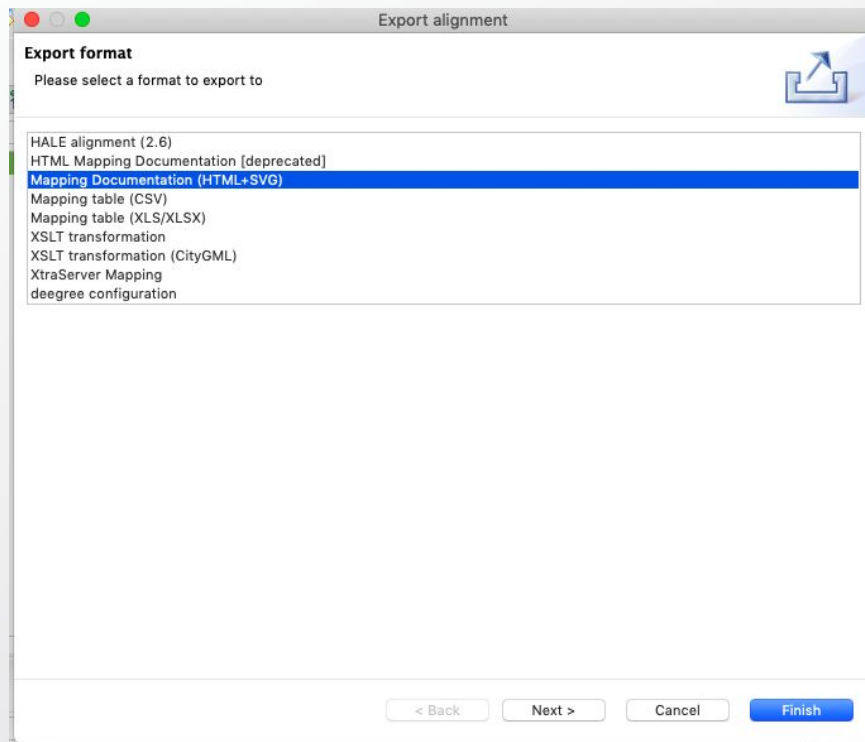
File -> Export -> Target schema



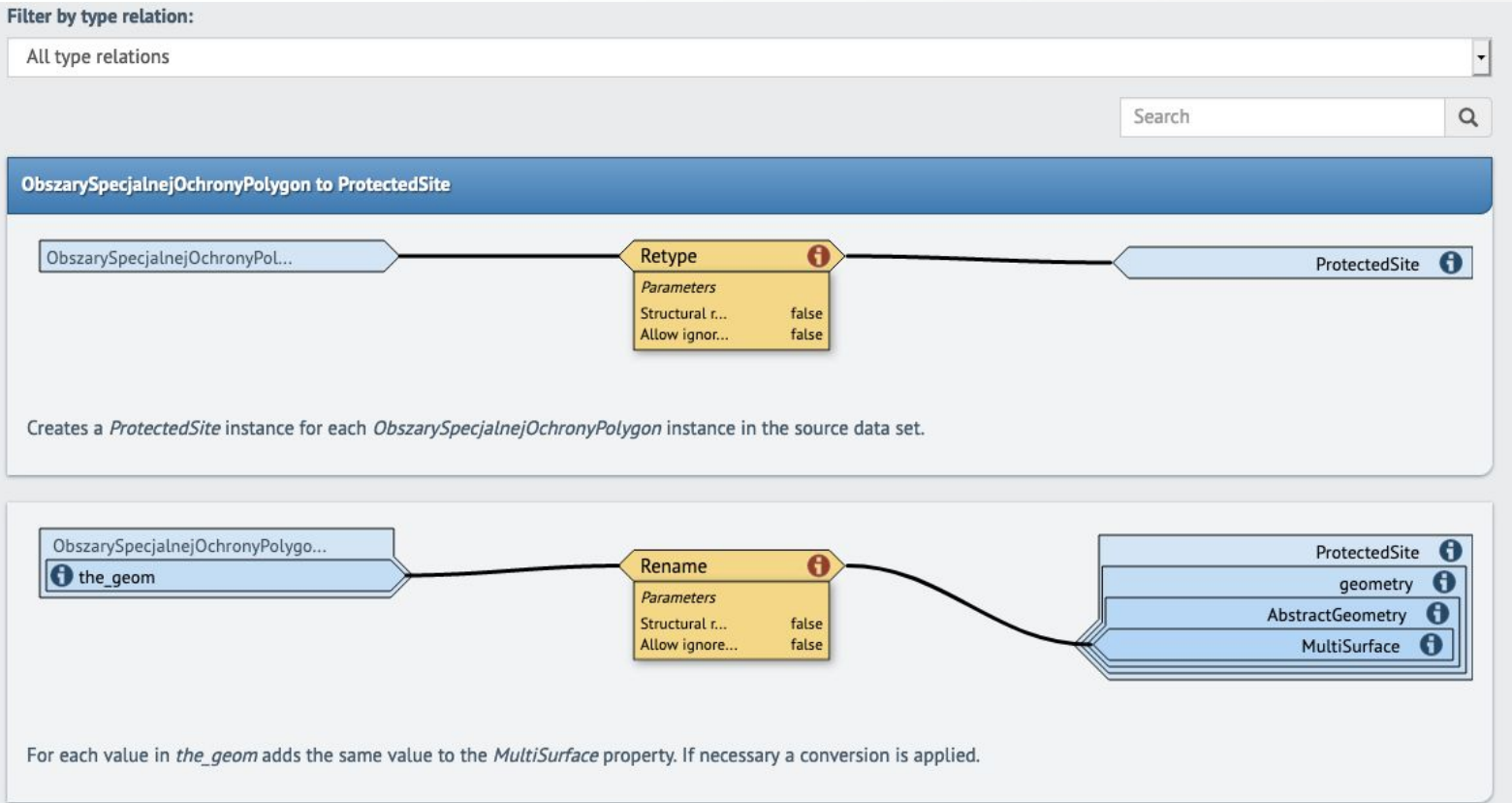
Jako formaty wynikowe dostępne są wyłącznie formaty czytelne dla HALE lub deegree. **Nie ma możliwości eksportu schematu XSD z HALE Studio.**

```
<?xml version="1.0" encoding="UTF-8"?><hsd:schemas xmlns:hsd="http://www.esdi-humboldt.eu/hale/schema">
<hsd:schema namespace="http://www.esdi-humboldt.eu/hale/shp">
<hsd:type-index>
<hsd:entry index="0">
<hsd:name namespace="http://www.esdi-humboldt.eu/hale/shp">gid</hsd:name>
</hsd:entry>
<hsd:entry index="1">
<hsd:name namespace="http://www.esdi-humboldt.eu/hale/shp">kod</hsd:name>
</hsd:entry>
<hsd:entry index="2">
<hsd:name namespace="http://www.esdi-humboldt.eu/hale/shp">kodinspire</hsd:name>
</hsd:entry>
<hsd:entry index="3">
<hsd:name namespace="http://www.esdi-humboldt.eu/hale/shp">MultiPolygon</hsd:name>
</hsd:entry>
<hsd:entry index="4">
<hsd:name namespace="http://www.esdi-humboldt.eu/hale/shp">nazwa</hsd:name>
</hsd:entry>
<hsd:entry index="5">
<hsd:name namespace="http://www.esdi-humboldt.eu/hale/shp">ObszarySpecjalnejOchronyPolygon</hsd:name>
</hsd:entry>
</hsd:type-index>
<hsd:mapping-relevant>5</hsd:mapping-relevant>
<hsd:types>
<hsd:type index="0">
<hsd:name namespace="http://www.esdi-humboldt.eu/hale/shp">gid</hsd:name>
<hsd:constraint type="binding" value="java.lang.Integer"/>
<hsd:constraint type="has-value" value="true"/>
<hsd:constraint type="mappable" value="false"/>
<hsd:constraint type="mapping-relevant" value="false"/>
</hsd:type>
```

Większe możliwości występują w zakresie eksportu mapowania - możliwe są formaty HALE, ale też XSLT, CSV, XLSX oraz dokumentacja graficzna (HTML z obrazem SVG)



Widok przykładowego eksportu schematu mapowania do HTML+SVG - do przeglądania w przeglądarce internetowej.



Przygotowanie szablonów mapowania z parametrami

Przygotowanie szablonu do zautomatyzowanego przetwarzania nie różni się w większości od zwykłego projektu mapowania, z wyjątkiem ustawień eksportu.

Z uwagi na dużą liczbę parametrów dostępnych dla eksportu, HALE umożliwia tworzenie tzw. **export configurations** (dokumentacja posługuje się pojęciem **export preset**) które w trakcie automatycznego przetwarzania będą wywoływane po nazwie.

File -> Export -> Create custom date export



The screenshot shows a window titled "Create export configuration". Inside, there's a section "Custom export configuration settings" with the instruction "Configure the export configuration". A small icon of a document with an arrow is in the top right. The form has three fields: "Name:" with the value "GML 4258", "Description:" with the text "Eksport do formatu GML w układzie EPSG:4258.", and "Format:" with a dropdown menu showing "GML file". At the bottom, there are four buttons: "< Back", "Next >", "Cancel", and "Finish".

Tworzenie szablonu eksportu przebiega tak samo, jak konfiguracja eksportu, za wyjątkiem tego, że zamiast podania nazwy pliku wynikowego podaje się nazwę i opis szablonu, a sam eksport nie jest wykonywany.

Wykorzystanie projektów mapowania z parametrami w Hale CLI

CLI jest skrótem od **C**ommand **L**ine **I**nterface.

HALE Studio po przygotowaniu projektu i szablonu eksportu może zostać wykorzystane do wykonania transformacji bez środowiska graficznego - w sposób zautomatyzowany.

Istnieją 4 sposoby na wywołanie HALE z linii komend:

- dla wszystkich platform: poprzez wirtualną maszynę Javy
- dla Windows: poprzez plik HALE.exe
- dla Linux/Mac: poprzez osobny program `hale-cli`
(<https://github.com/halestudio/hale-cli>)
- poprzez kontener Docker

Należy uruchomić wiersz polecenia (cmd), a następnie przeciągnąć do niego plik HALE.exe

a następnie dopisać parametry obowiązkowe: `-nosplash -console -application hale.transform`

Parametry:

- project - obowiązkowo: ścieżka do projektu
- source - obowiązkowo: ścieżka do danych źródłowych (źródło zdefiniowane w projekcie jest pomijane)
- filter - opcjonalnie: filtr do danych źródłowych w formacie CQL
- target - obowiązkowo: ścieżka do pliku wynikowego
- preset - opcjonalnie: nazwa szablonu eksportu
- validate - opcjonalnie: oznaczenie walidatora

Szczegółowy opis transformacji z linii komend znajduje się w zeszycie ćwiczeń.

Masowe przetwarzanie danych

HALE Studio nie posiada własnego systemu masowego przetwarzania danych.

Niemniej jednak można wykorzystać w tym celu plik Batch w Windows, lub skrypt shell w innych systemach.

Pętla po wszystkich plikach SHP w katalogu, przypisuje każdy plik do zmiennej %%i

```
for %%i in (*.shp) do  
    C:\Users\inspire\WeTransform\HALE.exe -nosplash  
-console -application hale.transform -project  
projekt.halez -source %%i -target %%i.gml -preset GML4258
```

Wywołanie HALE

ścieżka projektu

źródło (ze zmiennej)

nazwa
szablonu

wynik (ze zmiennej + rozszerzenie GML)

Skrypt należy zapisać w pliku z rozszerzeniem .bat, następnie uruchomić dwuklikiem lub z poziomu wiersza poleceń.

Publikacja danych

Dane zharmonizowane w HALE Studio można automatycznie opublikować w:

GeoServer

XtraServer

deegree

usłudze Hale Connect

Do eksportu do GeoServer jest niezbędna instalacja rozszerzenia AppSchema po stronie GeoServer.

Opis: <https://docs.geoserver.org/stable/en/user/data/app-schema/index.html>

Pobieranie:

<https://sourceforge.net/projects/geoserver/files/GeoServer/2.15.2/extensions/geoserver-2.15.2-app-schema-plugin.zip/download>

Instalacja polega na pobraniu archiwum, rozpakowaniu go i skopiowaniu do katalogu GeoServera i podkatalogu /WEB-INF/lib.

WAŻNE: po instalacji należy zapewnić min. 1 GB pamięci (Java Heap Space) dla GeoServer, inaczej eksport się nie powiedzie z błędem Out of Memory.

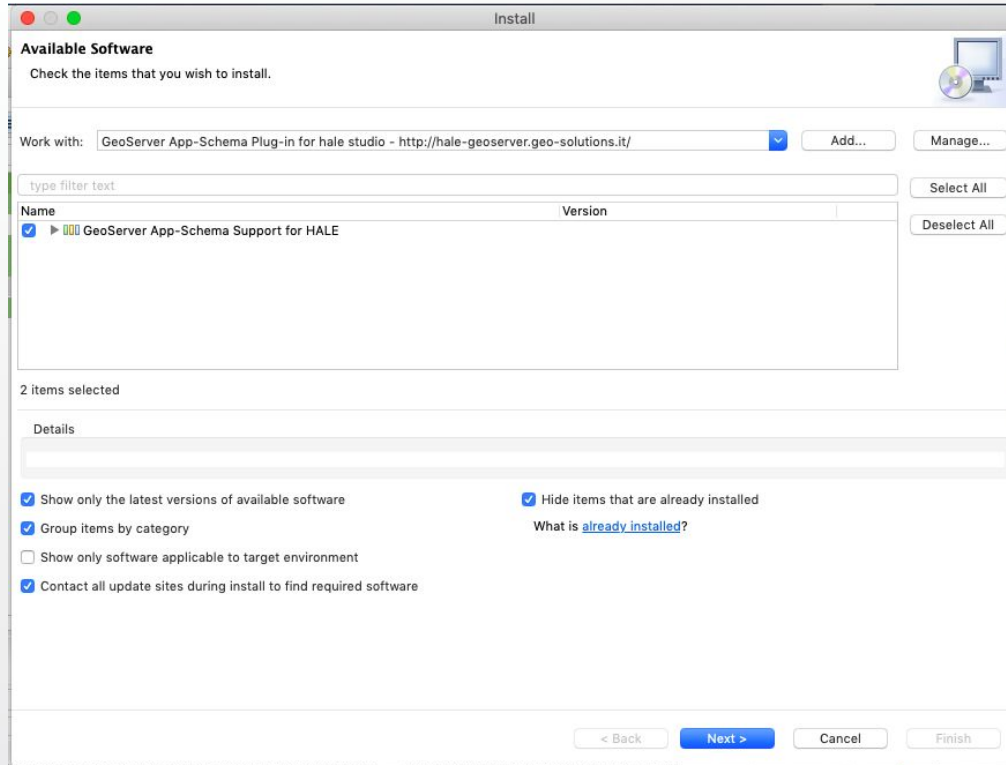
Zasadą działania publikacji danych za pomocą HALE Studio jest wykorzystanie już istniejącego rozszerzenia do obsługi schematów aplikacyjnych - App Schema.

W przypadku publikacji na serwer wysyłana jest jedynie **transformacja** danych, natomiast nie wysyła się żadnych danych wynikowych. Generowane przez GeoServer pliki GML są przykładem transformacji na żądanie - *on the fly*, w warstwie bazodanowej dane są przechowywane w dotychczasowym schemacie, który nie musi być zgodny z INSPIRE.

Wadą tego podejścia jest obsługa tylko jednego formatu danych źródłowych - musi być to baza PostGIS.

Po stronie HALE Studio należy zainstalować rozszerzenie GeoServer:

Help -> Install New Software...

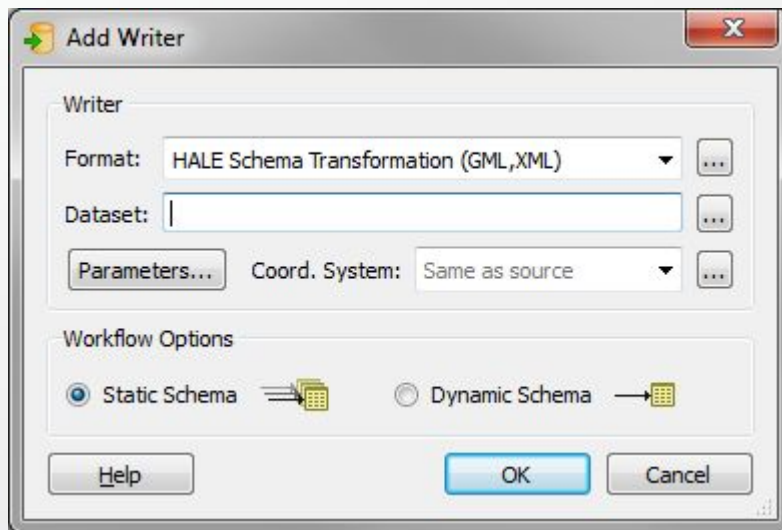


Opis procesu publikacji znajduje się w zeszycie ćwiczeń.

Rozszerzenia HALE Studio

GeoServer AppSchema
FME Plugin
XtraServer Plugin

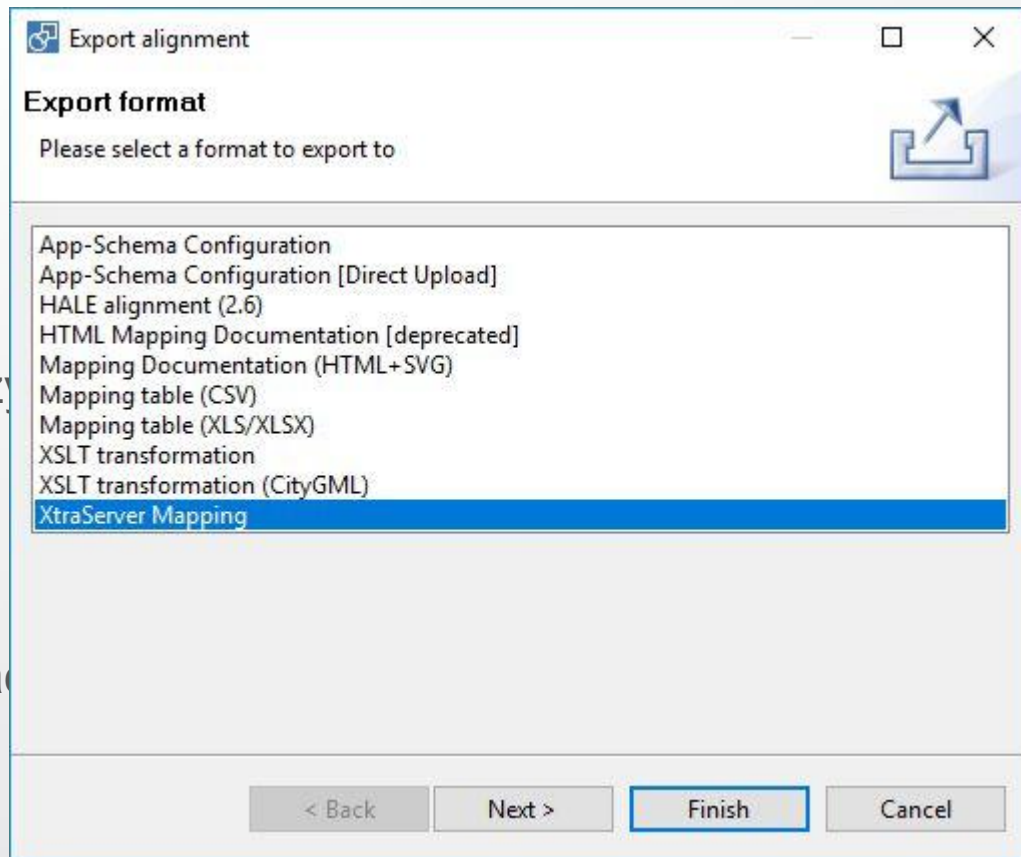
HALE Studio może współpracować z FME. Ideą rozszerzenia jest dodanie nowego formatu wynikowego - HALE Schema Transformation do FME.



XtraServer jest produktem firmy interactive instruments, który służy do publikacji usług sieciowych WMS WFS.

Aby wykorzystać rozszerzenie należy włączyć tryb "XtraServer compatibility mode" w HALE.

Po utworzeniu mapowania (Alignment) można je wyeksportować w formacie XtraServer.



Groovy jest językiem skryptowym działającym w środowisku wirtualnej maszyny Java.

Może być użyty do tworzenia własnych funkcji w HALE Studio (tak jak Python w QGIS).



support

Dziękuję za uwagę