



HARMONIZACJA ZBIORÓW DANYCH PRZESTRZENNYCH (HALE) – WARSZTATY (POZIOM ZAAWANSOWANY)

2017 r.





CELE BUDOWY INSPIRE

Cel główny: stworzenie zdolności współdziałania (**interoperacyjności**) w zakresie informacji przestrzennej w Europie.

Sposobem na realizację i osiągnięcie powyższego celu jest **harmonizacja** – działania o charakterze technicznym, organizacyjnym i prawnym, mające na celu doprowadzenie do wzajemnej spójności zbiorów danych przestrzennych i usług geoinformacyjnych.



WPROWADZENIE KLUCZOWE POJĘCIA

Dyrektywa INSPIRE

Dyrektywa z dnia 14. marca 2007 ustanawiająca jednolitą infrastrukturę informacji przestrzennej dla Unii Europejskiej i państw EFTA.

Specyfikacja danych

Dokument ustalający wymagania techniczne, które powinien spełniać wyrób, proces lub usługa.



WPROWADZENIE KLUCZOWE POJĘCIA

Interoperacyjność

W kontekście infrastruktury informacji przestrzennej oznacza możliwość łączenia zbiorów danych przestrzennych bez powtarzalnej interwencji manualnej.

Integracja danych

Działania mające na celu umożliwienie jednoczesnego łączenia wielu zbiorów, usunięcie z nich powtórzeń i elementów zbędnych.



WPROWADZENIE KLUCZOWE POJĘCIA

Harmonizacja danych

Proces mający na celu zapewnienie dostępu do danych przestrzennych umożliwiającego łączenie ich w sposób spójny z innymi zharmonizowanymi danymi.

Mapowanie

Proces przyporządkowania obiektów i ich atrybutów ze zbioru wejściowego do schematu danych wyjściowych.



ZNACZENIE I TREŚCI SPECYFIKACJI

Specyfikacje techniczne dzielą się na dwie podstawowe grupy:

- Dokumenty ramowe, dotyczące ogólnego schematu budowy danych INSPIRE oraz szczegółowo objaśniające sposób konstrukcji klas i atrybutów podstawowych.
- Szczegółowe specyfikacje dla specjalistycznych tematów zebranych w załączniki.

The screenshot displays the INSPIRE Knowledge Base website. The main header includes the European Commission logo and the text 'INSPIRE KNOWLEDGE BASE Infrastructure for spatial information in Europe'. The navigation menu includes 'Home', 'Learn', 'Implement', 'Participate', 'Use', and 'Toolkit'. The 'Implement' section is active, showing a list of documents under 'Data Specifications' and a 'Quick search' bar. The main content area is titled 'Data Specifications > Themes' and lists various themes organized into three annexes: ANNEX 1, ANNEX 2, and ANNEX 3. Each theme is represented by an icon and a text label.

Wszystkie specyfikacje techniczne znajdują się na stronie www: <https://inspire.ec.europa.eu/Themes/Data%20Specifications/2892>



ZNACZENIE I TREŚCI SPECYFIKACJI SPECYFIKACJE PODSTAWOWE

Do czego służą specyfikacje podstawowe?

- Ogólne omówienie modelu danych INSPIRE i przełożenie go na język maszynowy.
- Pełnią rolę nadrzędną wobec specyfikacji specjalistycznych (dla konkretnych tematów).
- Zawierają szczegółowe opisanie typów danych i sposobu ich harmonizacji dla tych klas i atrybutów, które powtarzają się w wielu tematach.



ZNACZENIE I TREŚCI SPECYFIKACJI SPECYFIKACJE PODSTAWOWE

Czego dotyczą specyfikacje podstawowe?

- Definicje Tematów INSPIRE
- Ogólny model danych
- Metodologia tworzenia specyfikacji
- Kodowanie danych przestrzennych
- Model danych Observations & Measurements
- Model podstawowy Coverage Types
- Model podstawowy Activity Complex



ZNACZENIE I TREŚCI SPECYFIKACJI SPECYFIKACJE TEMATYCZNE

Załączniki i tematy INSPIRE:

Annex I:

- Adresy
- Jedn. Administracyjne
- Działki
- Nazwy Geograficzne
- Hydrografia
- Obszary chronione
- Sieć transportowa

Annex II*:

- Ukształtowanie terenu
- Użytkowanie terenu
- Ortoobrazy
- Geologia

Annex III*:

- Budynki
- Obiekty produkcyjne i przemysłowe
- Zasoby naturalne
- Warunki atmosferyczne i meteorologiczno-geograficzne
- Warunki oceanograficzno – geograficzne
- Regiony morskie
- Gleba
- i inne....

* Załączniki tematycznie drugi i trzeci występują obecnie razem jako Annex II i III



ZNACZENIE I TREŚCI SPECYFIKACJI ZAWARTOŚĆ SPECYFIKACJI INSPIRE

1. Ogólne wytyczne dotyczące harmonizacji danych (wspólne dla wszystkich tematów)
2. Informacje o charakterze danych, których dotyczy temat, ich strukturze i metadanych, charakterystyka potencjalnych zbiorów źródłowych oraz sposób udostępniania w postaci usługi zebrane w rozdziałach:
 - a) Scope
 - b) Overview
 - c) Specification scopes
 - d) Identification information
 - e) Data content and structure
 - f) Reference systems, unit of measure and grids
 - g) Data quality
 - h) Dataset-level metadata
 - i) Delivery
 - j) Data capture
 - k) Portrayal
3. Dodatkowe informacje w postaci załączników do specyfikacji.



ZNACZENIE I TREŚCI SPECYFIKACJI ZAWARTOŚĆ SPECYFIKACJI INSPIRE

Ogólne wytyczne dotyczące harmonizacji danych (wspólne dla wszystkich tematów)

Są to wszystkie części specyfikacji przedstawione na szarym tle, mogą występować we wszystkich rozdziałach, a także w załącznikach. Zawierają najbardziej podstawowe informacje, takie jak: odwołania do aktów prawnych, specyfikacji podstawowych oraz definicje niektórych pojęć, a także reguły harmonizacji i integracji danych, które powinny spełniać zbiory źródłowe i zharmonizowane.



ZNACZENIE I TREŚCI SPECYFIKACJI ZAWARTOŚĆ SPECYFIKACJI INSPIRE

Articles 3, 4 and 5 of the Implementing Rules lay down the requirements for the content and structure of the data sets related to the INSPIRE Annex themes.

IR Requirement

Article 4

Types for the Exchange and Classification of Spatial Objects

1. For the exchange and classification of spatial objects from data sets meeting the conditions laid down in Article 4 of Directive 2007/2/EC, Member States shall use the spatial object types and associated data types, enumerations and code lists that are defined in Annexes II, III and IV for the themes the data sets relate to.
2. Spatial object types and data types shall comply with the definitions and constraints and include the attributes and association roles set out in the Annexes.
3. The enumerations and code lists used in attributes or association roles of spatial object types or data types shall comply with the definitions and include the values set out in Annex II. The enumeration and code list values are uniquely identified by language-neutral mnemonic codes for computers. The values may also include a language-specific name to be used for human interaction.

2.3 Normative References

- [Directive 2007/2/EC] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)
- [Directive 2008/98/EC] Directive 2008/98/EC of the European Parliament and of the Council of 19 November 2008 on waste and repealing certain Directives
- [ISO 19107] EN ISO 19107:2005, Geographic Information – Spatial Schema
- [ISO 19108] EN ISO 19108:2005, Geographic Information – Temporal Schema
- [ISO 19108-c] ISO 19108:2002/Cor 1:2006, Geographic Information – Temporal Schema, Technical Corrigendum 1
- [ISO 19111] EN ISO 19111:2007 Geographic information - Spatial referencing by coordinates (ISO 19111:2007)
- [ISO 19113] EN ISO 19113:2005, Geographic Information – Quality principles
- [ISO 19115] EN ISO 19115:2005, Geographic information – Metadata (ISO 19115:2003)
- [ISO 19118] EN ISO 19118:2006, Geographic information – Encoding (ISO 19118:2005)
- [ISO 19123] EN ISO 19123:2007, Geographic Information – Schema for coverage geometry and functions
- [ISO 19125-1] EN ISO 19125-1:2004, Geographic Information – Simple feature access – Part 1: Common architecture

8.2.1 Coordinate Reference System

Metadata element name	Coordinate Reference System
Definition	Description of the coordinate reference system used in the dataset.
ISO 19115 number and name	13. referenceSystemInfo
ISO/TS 19139 path	referenceSystemInfo
INSPIRE obligation / condition	mandatory
INSPIRE multiplicity	1..*
Data type (and ISO 19115 no.)	186. MD_ReferenceSystem
Domain	To identify the reference system, the referenceSystemIdentifier (RS_Identifier) shall be provided. NOTE More specific instructions, in particular on pre-defined values for filling the referenceSystemIdentifier attribute should be agreed among Member States during the implementation phase to support interoperability.
Implementing instructions	
Example	referenceSystemIdentifier: code: ETRS_89 codeSpace: INSPIRE RS registry
Example XML encoding	<pre><gmd:referenceSystemInfo> <gmd:MD_ReferenceSystem> <gmd:referenceSystemIdentifier> <gmd:RS_Identifier> <gmd:code> <gco:CharacterString>ETRS89 </gco:CharacterString> </gmd:code> <gmd:codeSpace> <gco:CharacterString>INSPIRE RS registry</gco:CharacterString> </gmd:codeSpace> <gmd:RS_Identifier> </gmd:referenceSystemIdentifier> </gmd:MD_ReferenceSystem> </gmd:referenceSystemInfo></pre>
Comments	



ZNACZENIE UML DLA HARMONIZACJI

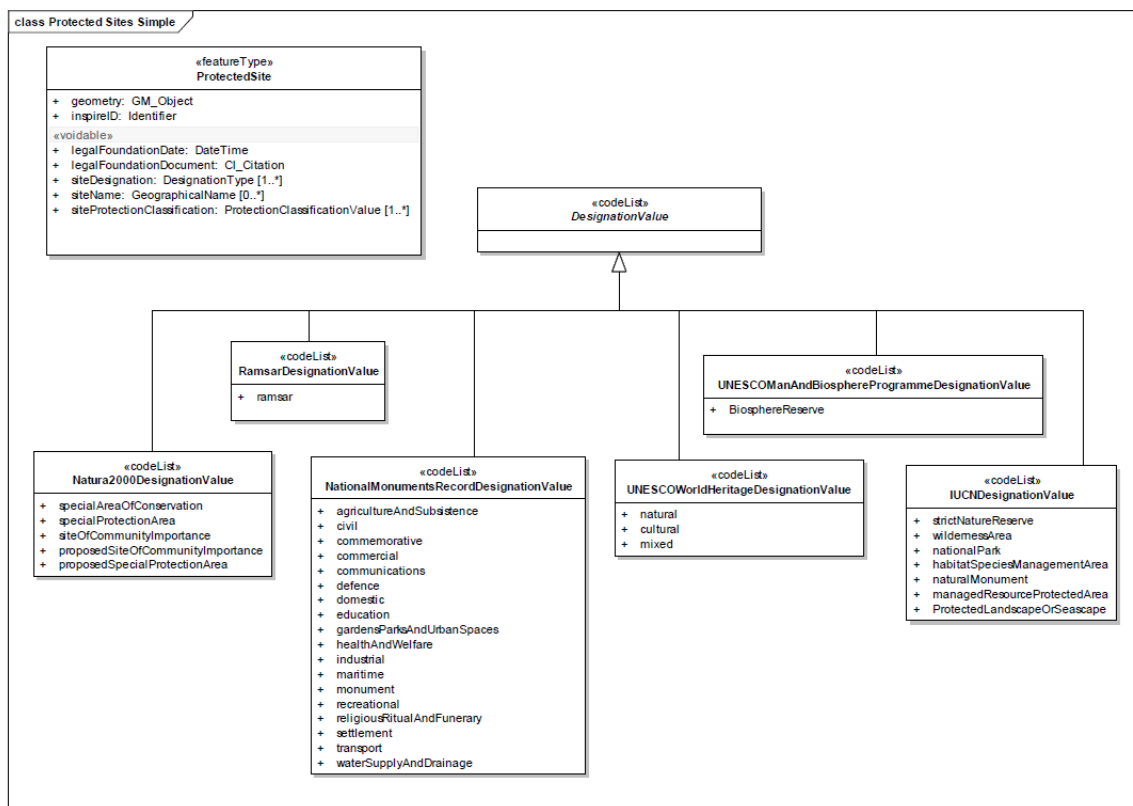
Struktura danych INSPIRE jest w specyfikacjach przedstawiana zarówno w formie katalogów obiektów...

ProductionFacility	
Value type:	GM_Surface
Definition:	Spatial property of the spatial object with a value that is a vector geometry type from ISO 19107.
Description:	This is the secondary geometry property for the Production Facility and it is set as an optional property to identify the location area for the Facility. It is intended, if available, to furnish a more detailed spatial information in addition to the basic mandatory geometry.
Multiplicity:	0..1
Stereotypes:	«voidable»
Attribute: riverBasinDistrict	
Value type:	RiverBasinDistrictValue
Definition:	Code identifier and/or name assigned to the basin district of a watercourse.
Description:	NOTE Information required (not registered in the Hydrography Theme) according to Article 3(1) of Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy (OJ L 327, 22.12.2000, p. 1). Directive as amended by Decision No 2455/2001/EC (OJ L 331, 15.12.2001, p. 1).
Multiplicity:	0..1
Attribute: status	
Value type:	StatusType
Definition:	The state or condition of the facility, with regard to the functional and operational order, in which it is arranged for a limited or extended time period.
Multiplicity:	1..*
Stereotypes:	«voidable»
Association role: groupedPlot	
Name:	A Plot is grouped by one or more Facilities (this means a Plot may be shared by different Facilities)
Value type:	ProductionPlot
Multiplicity:	0..*
Stereotypes:	«voidable»
Association role: hostingSite	
Name:	A Site hosts one or more Facilities
Value type:	ProductionSite
Multiplicity:	0..1
Stereotypes:	«voidable»
Association role: groupedInstallation	
Name:	An Installation is grouped by one Facility
Value type:	ProductionInstallation
Multiplicity:	0..*
Stereotypes:	«voidable»
Association role: groupedBuilding	
Value type:	ProductionBuilding
Multiplicity:	0..*
Stereotypes:	«voidable»



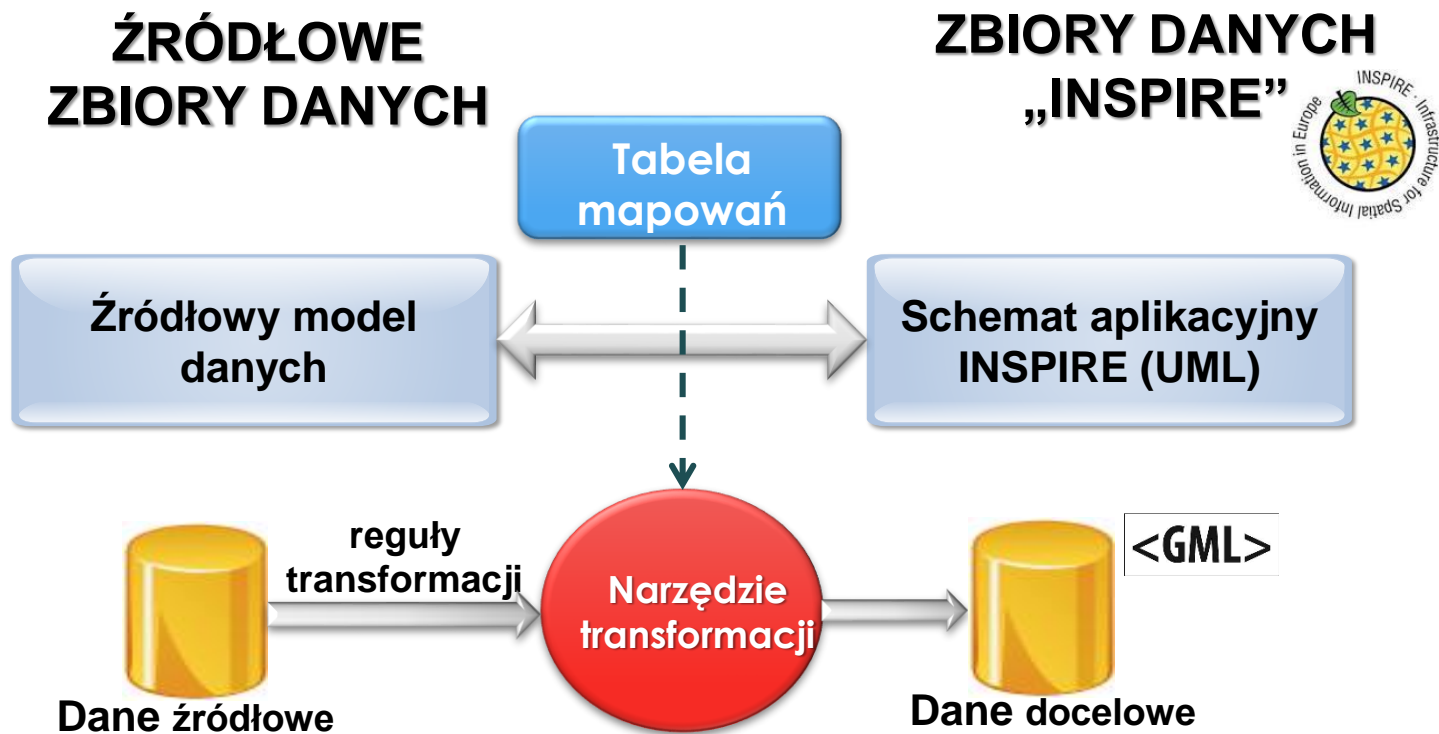
ZNACZENIE UML DLA HARMONIZACJI

... jak i diagramów UML:





UPROSZCZONY PROCES HARMONIZACJI





ETAPY HARMONIZACJI

- 1. Identyfikacja zbiorów danych przestrzennych**, która polega na wyborze zasobów danych przestrzennych, które powinny podlegać procesowi harmonizacji zgodnie z wytycznymi INSPIRE. Na tym etapie istotna jest analiza istniejących zbiorów danych, w tym struktury danych i zakresu informacyjnego danych źródłowych pod kątem spełnienia wymagań specyfikacji INSPIRE dla określonego tematu.
- 2. Przekształcenie danych.** Etap który polega na dostosowaniu wybranych danych źródłowych do wymagań wynikających z dyrektywy INSPIRE. Obejmuje zadania mające na celu przemodelowanie danych ze struktur stosowanych w schematach implementacyjnych danych źródłowych do struktur wymaganych przez model wymiany danych zgodnie ze specyfikacją. Składa się z następujących etapów:
 - a) przygotowanie danych do transformacji**, obejmujące stworzenie tabeli mapowań pomiędzy źródłowym a schematem docelowym INSPIRE,



ETAPY HARMONIZACJI

- b) transformacja danych ze schematu źródłowego do schematu docelowego INSPIRE,
- c) transformacja danych z lokalnego układu współrzędnych do układu wspólnego dla danych publikowanych w INSPIRE,
- d) pozyskanie dodatkowych informacji, które nie są dostępne w zbiorze źródłowym, a konieczne są do spełnienia wymogów narzuconych w specyfikacji dla konkretnego tematu. Spełnienie wymagań dla określonego tematu INSPIRE może wymagać przetworzenia danych pochodzących z wielu zbiorów danych przestrzennych. W takich przypadkach w procesie harmonizacji można podejmować działania na rzecz integracji tych zbiorów. Integracja ma celu wówczas uzyskanie możliwe kompletnego zbioru danych zharmonizowanych (zgodnie ze specyfikacją INSPIRE dla określonego tematu).



ETAPY HARMONIZACJI

3. Publikacja danych poprzez usługi sieciowe obejmuje prace polegające na opublikowaniu zharmonizowanych danych za pomocą geoinformacyjnych usług sieciowych, w szczególności usług pobierania i przeglądania. Jednym z zadań realizowanych w ramach tego etapu jest opracowanie metadanych, które opisują zharmonizowane dane i warunki korzystania z tych danych. Metadane powinny być opublikowane w katalogu tak, by możliwe było wyszukanie informacji o zharmonizowanych zbiorach w krajowej infrastrukturze informacji przestrzennej za pomocą dostępnych usług wyszukiwania.



DROGA DO HARMONIZACJI ROZPOZNANIE TEMATU

- Harmonizacja danych zaczyna się już na etapie podjęcia decyzji.
- Rozpoznanie tematu dokonujemy poprzez zapoznanie się z kluczowymi informacjami zawartymi w specyfikacji i konfrontację ich z własnym doświadczeniem.
- Wiedza o tym, co chcemy otrzymać na wyjściu może być kluczowa dla odpowiedniego dobrania danych wejściowych:
- Należy podjąć decyzję, pozyskanie których danych jest obligatoryjne dla stworzenia satysfakcjonującego zbioru danych zharmonizowanych
- Istotnym czynnikiem jest ocena czasochłonności pracy związanej z integracją danych z potencjalnie wielu źródeł
- Mając świadomość tego, co chcemy uzyskać, możemy zacząć szukać danych do harmonizacji (zbiorów źródłowych)



DROGA DO HARMONIZACJI MAPOWANIE

- Mapowanie to „tłumaczenie” jednego modelu danych na drugi
- Nie wszystko można przetłumaczyć
- Wykonanie poprawnego mapowania jest kluczowym etapem, który warunkuje sukces całego procesu harmonizacji danych
- Podczas tworzenia mapowania warto bazować na zbiorach dostępnych (najlepiej będących w dyspozycji organizacji), nie zapominając, że ich kolejne wersje mogą zawierać więcej informacji (zgodnych z modelem danych źródłowych)



DROGA DO HARMONIZACJI TRANSFORMACJA DANYCH

- Transformacja danych źródłowych do modelu INSPIRE powinna być jedynie zwieńczeniem całego procesu – przełożeniem dobrego mapowania na język maszynowy.
- Koniecznym elementem transformacji jest uwzględnienie formatu danych wejściowych, który determinuje dobór narzędzi.
- Narzędzia użyte do harmonizacji powinny zostać dobrane w sposób optymalny, tak, by transformacja danych nie była zbyt czasochłonna, ale zakończyła się uzyskaniem w pełni zharmonizowanego zbioru.



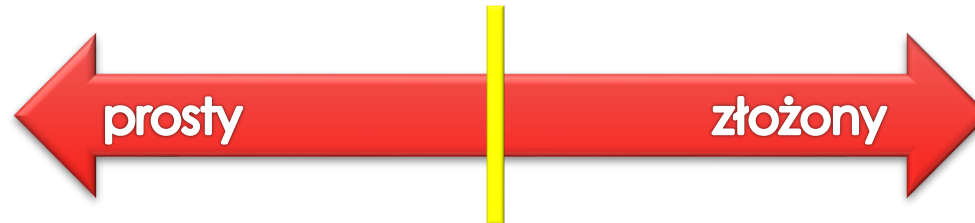
ETL: TRANSFORMACJA DANYCH

Transformacja danych to zbiór funkcji oraz reguł, które pozwalają dane źródłowe przekształcić do postaci zgodnej ze strukturami oczekiwanymi docelowymi.

Przykłady transformacji danych:

- tłumaczenie kodowanych wartości,
- transformacja współrzędnych,
- filtrowanie, sortowanie, agregacja,
- selekcja, łączenie,
- mapowanie list kodowych.

▶▶ JAKI POZIOM HARMONIZACJI JEST WŁAŚCIWY?



Zbyt prosty:

- Utrata informacji
- Mało korzyści z wykorzystania informacji

Zbyt złożony:

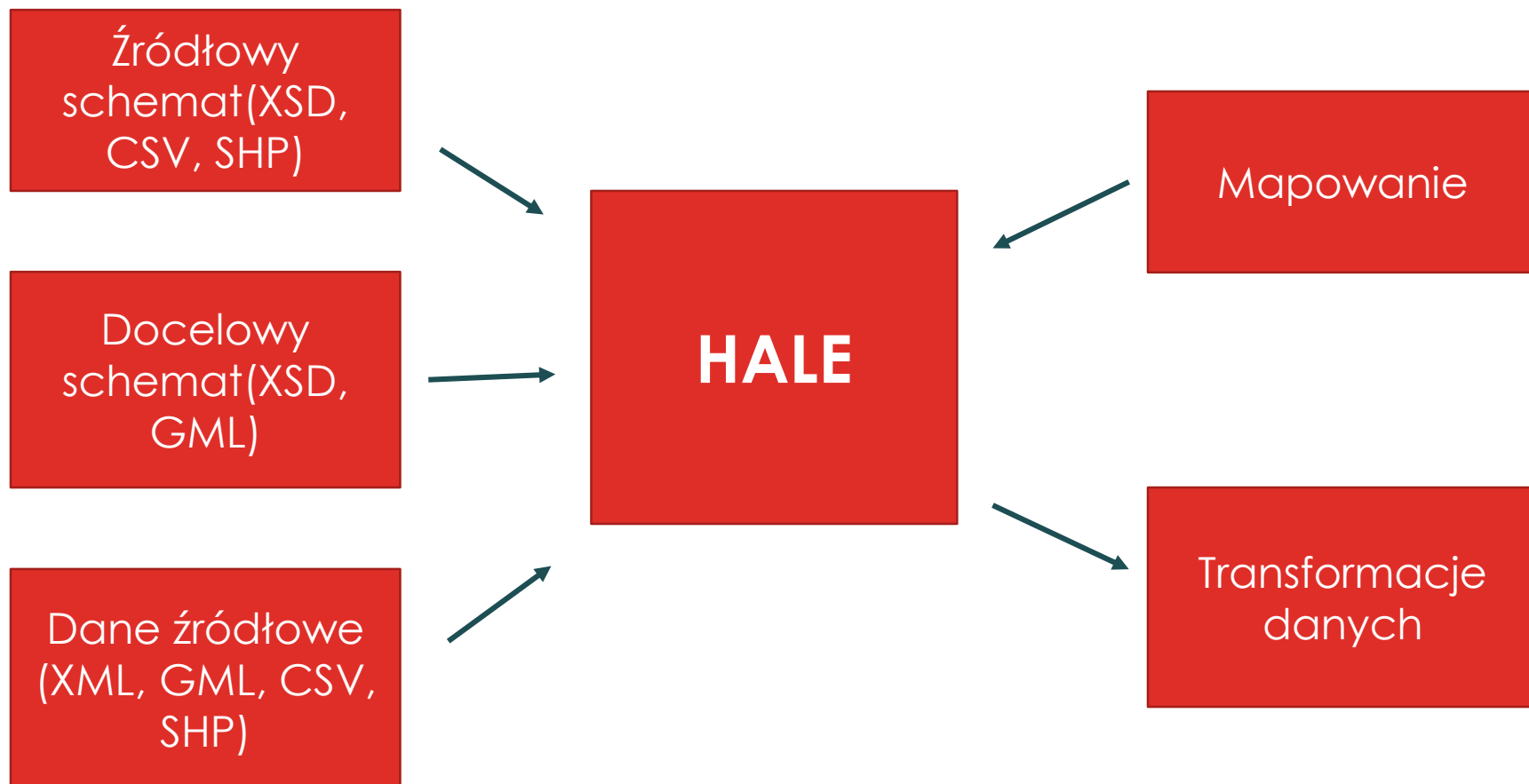
- Trudności w implementacji
- Wysokie koszty

Wymagania

- Proces iteracyjny
- Konieczne dobrze określone założenia
- Zrozumienie danych poddawanych procesowi harmonizacji
- Testowanie



WPROWADZENIE DO HALE





WPROWADZENIE DO HALE

- HALE zapewnia rozbudowany graficzny interfejs przystosowany do potrzeb specjalistów GIS, dzięki czemu proces mapowania oraz samej transformacji jest czytelniejszy, a co za tym idzie łatwiejszy do zrozumienia, niż w przypadku transformacji XSL.
- Dodatkowo zapewniany jest bezpośredni podgląd na rezultaty transformacji zapewniający porównanie pomiędzy danymi wejściowymi, a wyjściowymi.
- Oprogramowanie rozwijane z myślą o przeprowadzaniu harmonizacji danych przestrzennych INSPIRE.



HALE - WORKFLOW

1. Import danych źródłowych wraz z ich schematem.
2. Import docelowego schematu danych.
3. Identyfikacja odpowiadających sobie klas obiektów oraz typów danych.



HALE - WORKFLOW

4. Identyfikacja oraz dodanie w pliku projektowym powiązań oraz relacji pomiędzy źródłowymi, a docelowymi typami danych.

- Jakie informacje są dostępne w źródłowym zbiorze danych?
- Jakie informacje są niezbędne dla zasilenia zbioru docelowego?
- W jaki sposób zasilić zbiór docelowy ze źródła?



HALE - WORKFLOW

5. Po każdej zmianie wprowadzonej w mapowaniu powinna nastąpić weryfikacja czy wyniki transformacji są zadowalające. Należy proces powtarzać iteracyjnie, aż rezultat będzie zadowalający.
6. Zapis przetransformowanych danych do pliku. Istnieje możliwość zdefiniowania samego mapowania, w celu późniejszego przetworzenia zbioru danych.



HALE – OKNO PROGRAMU

The screenshot shows the HALE Studio 3.3.1 interface with the following components:

- 1**: Source schema explorer showing a tree structure for 'parki x122' with elements like 'akt_prawny', 'data', 'filename', 'gid', 'kodinspire', 'nazwa', 'OBJECTID', 'pow', and 'the_geom'.
- 2**: Target schema explorer showing a tree structure for 'ProtectedSite x122' with elements like 'DesignationType', 'location', 'boundedBy', 'description', 'descriptionReference', 'geometry', 'id', 'inspireID', 'legalFoundationDate', 'legalFoundationDocument', 'metaDataProperty', 'name', 'siteDesignation', 'siteName', and 'siteProtectionClassification'.
- 3**: Alignment pane showing a mapping diagram between source and target elements. It includes operations like 'Rename', 'Assign (Bound)', and 'Formatted string'.
- 4**: Error Log pane showing a table of messages with columns for Message, Plug-in, and Date.
- 5**: Type hierarchy pane showing a list of events with columns for Name, Date, and Time.

Message	Plug-in	Date
Found 1 possible FeatureCollection elements, using element (http://www.opengis.net/...)	eu.esdihumboldt.ha...	10.10.2017, 14:34
Found 1 possible FeatureCollection elements, using element (http://www.opengis.net/...)	eu.esdihumboldt.ha...	10.10.2017, 14:34
All validations completed successfully.	eu.esdihumboldt.ha...	10.10.2017, 14:32
All configured validations were successful.	eu.esdihumboldt.ha...	10.10.2017, 14:32
GML file export	eu.esdihumboldt.ha...	10.10.2017, 14:32
[14:32:00] Geometry structure match: MultiPolygon - (http://www.opengis.net/gml/...)	eu.esdihumboldt.ha...	10.10.2017, 14:32
[14:32:00] Geometry structure match: MultiPolygon - (http://www.opengis.net/gml/...)	eu.esdihumboldt.ha...	10.10.2017, 14:32

Name	Date	Time
14:13 2017-10-10		
Project validator		14:32:03
GML file export		14:32:00
Instance validation		14:31:12
Instance transformation		14:31:10
Load data into database		14:31:08
Shapelite import		14:31:07
Project validator		14:23:53
GML file export		14:23:49
Instance validation		14:14:31

- 1) okno schematu źródłowego
- 2) okno schematu docelowego
- 3) okno transformacji
- 4) rejestr błędów/ właściwości
- 5) wydarzenia



HALE – OKNO MAPY

The screenshot displays the HALE software interface with several key components:

- Map 1:** A satellite map showing source data points (blue circles) in the Poznań region.
- Map 2:** A map showing transformed data points (red circles) in the Lublin region.
- Source Data Table:** A table with columns for 'oczyszczalnie_sciekow' and 'ProductionFacility'. It lists attributes like 'Adres', 'geometryProperty', 'ID_baza', 'Id_tysn', 'Nazwa', 'Odzysk_bio', and 'Powiat'.
- Transformed Data Table:** A table with columns for 'ProductionFacility' and 'ProductionFacility'. It lists attributes like 'Identifier', 'namespace', 'versionId', 'metaDataProperty', 'name', 'riverBasinDistrict', 'status', and 'surfaceGeometry'.
- Help Window:** A window titled 'What is a Schema?' explaining schema levels (Logical, Conceptual) and mapping processes.

- 1) mapa z danymi źródłowymi
- 2) Mapa z danymi po transformacji
- 3) Okno pomocy

- 4) Przykład danych wejściowych
- 5) Przykład danych wyjściowych



HALE – SCHEMAT DANYCH – PODSTAWOWE ELEMENTY

Typy



Typ danych



Abstrakcyjny typ danych



Obiekt przestrzenny



Abstrakcyjny obiekt przestrzenny

Atrybuty



Tekstowy



Numeryczny



Geometria



Złożony

Target

type filter text

- DesignationType
 - designation
 - designationScheme
 - percentageUnderDesignation (0..1)
- ProtectedSite 122
 - location (0..1)
 - location
 - priorityLocation
 - boundedBy (0..1)
 - description (0..1)
 - descriptionReference (0..1)
 - geometry ×122
 - id ×122
 - identifier (0..1)
 - inspireID ×122
 - legalFoundationDate ×122
 - legalFoundationDocument ×122
 - metaDataProperty (0..n)
 - name (0..n)
 - siteDesignation (1..n) ×122
 - siteName (0..n) ×122
 - siteProtectionClassification (1..n) ×122