ArcGIS Data Interoperability – codzienna praca z różnymi formatami i strukturami danych

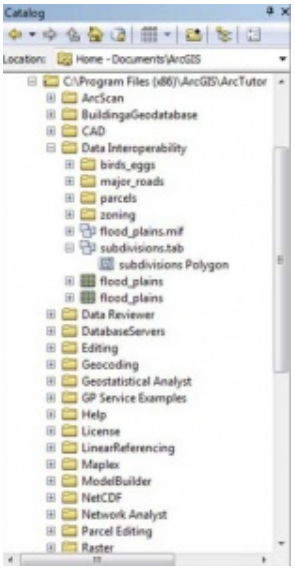
Wraz z rozwojem różnych technologii GIS pojawia się coraz więcej formatów, w których zapisywane są dane. Niemalże codziennie użytkownicy stają przed wyzwaniami związanymi z ich odczytaniem i integracją z innymi danymi. Jak odczytać dane zapisane w różnych formatach i pochodzące od różnych dostawców? W jaki sposób je zaimportować i wyeksportować do innego formatu? Jak przekształcić dużą ilość różnie zapisanych danych przestrzennych do jednego modelu, który może być opublikowany i w prosty sposób wykorzystywany przez innych? Problemy te możemy rozwiązać dzięki rozszerzeniu do ArcGIS Desktop – Data Interoperability.

Rozszerzenie to nie służy wyłącznie do prostej konwersji formatów Esri do innych formatów i odwrotnie, ale jest to jedna z wielu możliwości, jakie nam daje.

Przegląd możliwości Data Interoperability

ArcGIS Data Interoperability jest rozszerzeniem stworzonym wspólnie przez firmy Esri oraz Safe Software i opiera się na technologii Feature Manipulation Engine (FME). Ułatwia użytkownikom korzystanie ze wszelkich danych GIS w środowisku ArcGIS Desktop, bez względu na format danych. Umożliwia bezpośredni odczyt ponad 100 formatów danych przestrzennych, m.in.: GML, XML, WFS, Autodesk®, DWG™/DFX™, MicroStation® Design, MapInfo®, MID/MIF, TAB, Oracle®, Oracle Spatial, Intergraph® GeoMedia® Warehouse, oraz eksport do ponad 70 formatów zapisu danych. Rozszerzenie daje możliwość bezpośredniego odczytu i przeglądania, a także eksportu i importu danych

zapisanych w różnych formatach, dzięki czemu możliwa jest ich wizualizacja i analiza z użyciem narzędzi dostępnych w środowisku ArcGIS Desktop. Dzięki aplikacji Workbench możliwa jest budowa złożonych przestrzennych narzędzi ETL (Extract – Transform – Load), pozwalających na pozyskanie danych, transformację ich atrybutów i geometrii oraz dopasowanie ich do docelowego modelu danych i zasilenie go tymi danymi.



Rys. 1. Widok danych w formacie MapInfo TAB (MITAB) w drzewie katalogu.

Rozszerzenie do ArcGIS Desktop Data Interoperability oferowane jest na dwóch poziomach licencjonowania:

- instalacja rozszerzenia bez autoryzacji, dająca możliwość bezpośredniego dostępu do usługi WFS (Web Feature Service) i do danych zapisanych w formacie GML – SF (Geography Markup Language – Simple Features); opcja ta daje możliwość utworzenia połączenia do WFS i GML, dodania ich do mapy i użycia ich z narzędziami geoprzetwarzania, jak również użycia narzędzi konwersji,

znajdujących się w skrzynce narzędziowej Data Interoperability do odczytu usług WFS i do odczytu i zapisu GML,
- instalacja rozszerzenia i jego autoryzacja licencją, umożliwiająca dodanie i korzystanie z dodatkowych, dostępnych z pudełka 100 formatów danych przestrzennych i dająca dostęp do aplikacji Workbench.

Bezpośredni odczyt formatów i tworzenie połączeń

Bezpośredni odczyt formatów umożliwia pracę z różnymi formatami danych bez użycia pośredniego formatu lub narzędzi do konwersji. Są one tłumaczone (przekształcane, odczytywane) na bieżąco jako dane tylko do odczytu i wyświetlane w pamięci, podczas gdy plik źródłowy pozostaje na dysku w swoim natywnym formacie. Aby skorzystać z tej funkcjonalności wystarczy po zainstalowaniu i autoryzacji rozszerzenia Data Interoperability uruchomić je z poziomu aplikacji, której chcemy użyć do bezpośredniego odczytu danego formatu. Dokonać tego możemy, wybierając opcję Customize (z głównego menu aplikacji) Extensions i dodając nasze dane do okna aplikacji. Jako przykład poniżej prezentowany jest widok danych w formacie MapInfo TAB (MITAB) w drzewie katalogu.

Wyświetlenie i widoczność w drzewie katalogu niektórych formatów może wymagać utworzenia połączenia interoperacyjnego. Jest ono stworzonym przez użytkownika połączeniem (linkiem) do danych źródłowych, które są zarządzane w drzewie katalogu jako pojedynczy zbiór danych do bezpośredniego odczytu. Połączenie pozwala określić dane źródłowe, format odczytu danych (tzw. reader FME) i inne parametry obsługiwane przez wybrany format. Jego wykorzystanie może być przydatne przy:

- połączeniu do usługi WFS (Web Feature Service) i dodawaniu do mapy typów obiektów przez nią opublikowanych,

- definiowaniu układu współrzędnych dla danych źródłowych,
- uzyskaniu szybkiego dostępu do danych chronionych hasłem,
- połączeniu zawartości wielu plików w pojedynczy zbiór danych,
- przechowywaniu określonych parametrów baz danych, takich jak połączenia i hasło.

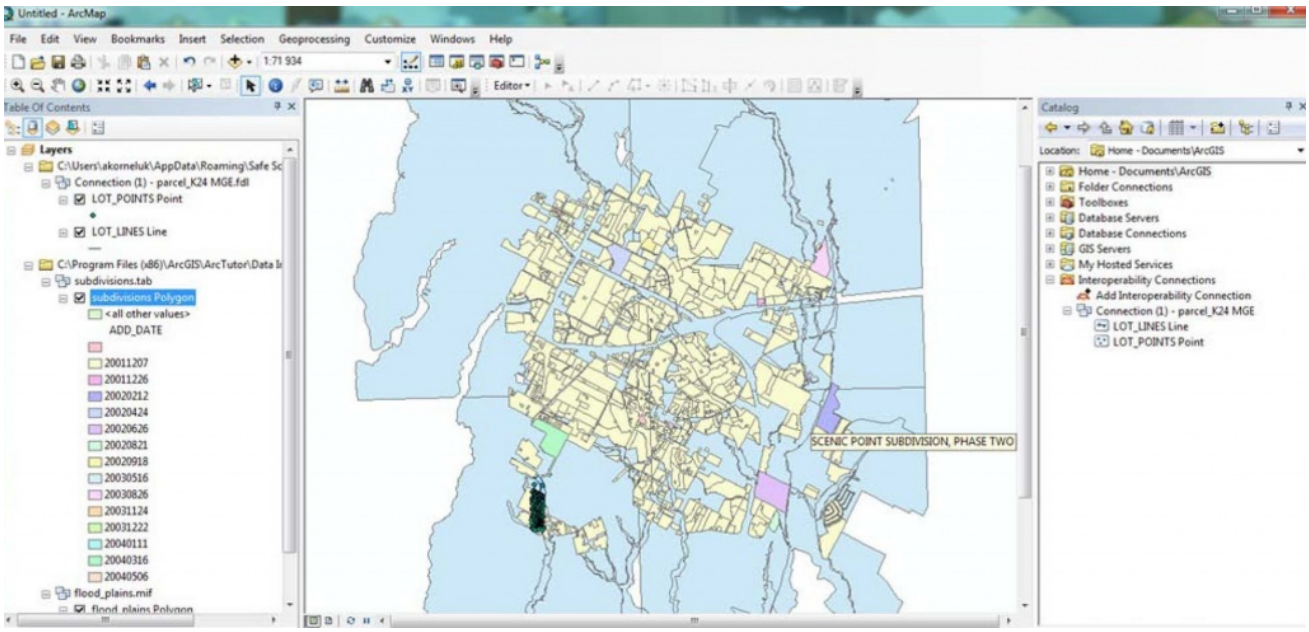
Utworzenie połączenia interoperacyjnego przebiega w kilku prostych krokach. Najpierw rozwijamy folder Interoperability Connections znajdujący się w drzewie katalogu, po czym wybieramy opcję Add Interoperability Connection, w wyniku czego zostaje wyświetlone okno dialogowe Interoperability Connection, w którym dokonujemy wszystkich potrzebnych ustawień. Wybieramy format danych, dla którego chcemy utworzyć nasze połączenie, lokalizację danych źródłowych oraz, w razie potrzeby, definiujemy układ współrzędnych.

Przykładem zastosowania może być dodanie połączenia do formatu Intergraph MGE (Modular GIS Environment), który do reprezentowania geometrii obiektów wykorzystuje standardowe elementy MicroStation (korzysta z silnika Microstation). Po otwarciu okna Interoperability Connection jako format odczytu danych ustawiony zostanie wybrany z galerii (okno FME Reader Gallery) format MGE, dzięki któremu możliwy będzie odczyt plików i powiązanej z nimi bazy danych. Format DGN Microstation pozwala na zapisywanie plików bez rozszerzenia, a także na zapisywanie plików z rozszerzeniem o dowolnej długości. W rezultacie pliki DGN mogą być zapisane z rozszerzeniem, które może sugerować, jakiego typu dane są w nim przechowywane. W opisywanym przykładzie zostanie wybrany z lokalizacji plik z rozszerzeniem .par, które reprezentuje działki (ang. parcels). Dla danych źródłowych zostanie określony układ współrzędnych, w jakim się znajdują. Ostatnim etapem będzie zdefiniowanie parametrów formatu odczytu. Podane zostaną parametry bazy mdb, w której przechowywane są dane.

Rys. 2. Schemat dodania połączenia interoperacyjnego.

Dane odczytane – co dalej?

W momencie dodania do aplikacji ArcMap (a także ArcScene czy ArcGlobe) bezpośrednio odczytanych formatów lub też formatów zdefiniowanych poprzez połączenie interoperacyjne dostępne stają się wszystkie standardowe funkcje mapy. Możemy przeglądać tabelę atrybutów, używać opcji etykietowania, snapować do geometrii, zmieniać symbolizację i użyć jej ze wszystkimi narzędziami geoprzetwarzania, które jako dane wejściowe akceptują klasy obiektów lub warstwy. Dla przykładu poniżej prezentowany jest widok dodanych do okna aplikacji ArcMap formatów danych MAPINFO TAB z obszarami podziału (subdivisions.tab) i Intergraph MGE przechowujący działki. Dodatkowo została nadana symbolizacja dla obszarów podziału, na podstawie unikalnych wartości, którymi było pole przechowujące datę podziału obszaru.



Rys. 3. Widok danych w formacie MITAB (z nadaną symbolizacją wg unikalnej wartości) i MGE.

Skrzynka narzędziowa Data Interoperability

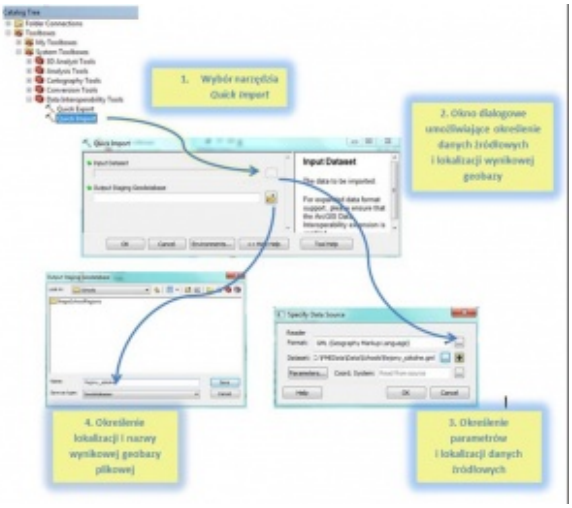
Skrzynka narzędziowa Data Interoperability zawiera narzędzia pozwalające na import i eksport danych pomiędzy geobazą a nienatywnym formatami danych, z wykorzystaniem formatów odczytu i zapisu FME (ang. Reader/Writer FME). Narzędzia te stanowią idealne rozwiązanie, kiedy potrzebujemy szybko przekonwertować nasze dane. Mogą być one stosowane jako samodzielne narzędzia geoprzetwarzania, używane przy budowaniu modeli w ModelBuilder, a także wykonywane w oknie lub skrypcie Pythona.

Importowanie i eksportowanie danych

Wykorzystanie narzędzi Quick Import i Quick Export stanowi dobre rozwiązanie, gdy interesują nas szybki import lub eksport danych bez modyfikacji geometrii czy atrybutów. Użycie szybkiego importu pozwoli na konwersję dowolnych danych źródłowych wspieranych przez Data Interoperability i zasilenie nimi geobazy. Z kolei wykorzystanie szybkiego eksportu

umożliwi konwersję danych przechowywanych w geobazie do formatów obsługiwanych przez to rozszerzenie.

Jako przykład praktycznego wykorzystania posłużyć może import pliku GML z rejonami szkół różnych typów do geobazy plikowej. Uruchamiamy narzędzie Quick Import, w wyniku czego wyświetlone zostaje okno dialogowe, w którym wybieramy przycisk obok pola Input Dataset. W otwartym oknie Specify Data Source określamy format odczytu danych (pole Format), którym w naszym przypadku jest format GML, a następnie określamy ich lokalizację (pole Dataset). Aby zatwierdzić wybrane parametry, wybieramy przycisk OK. Po dokonaniu ustawień danych źródłowych możemy następnie określić lokalizację i nazwę geobazy plikowej, do której będą one importowane. W oknie dialogowym Quick Import wybieramy przycisk znajdujący się obok pola Output Staging Geodatabase, a następnie określamy nazwę geobazy i ścieżkę do folderu, w którym zostanie ona utworzona. Uruchomienie narzędzia nastąpi po wybraniu przycisku OK.



Rys. 4. Schemat użycia narzędzia Quick Import.

Importowany plik GML przechowywał cztery typy obiektów poligonowych reprezentujących rejony różnych typów szkół, dlatego też efektem końcowym wykonanej pracy będą cztery klasy obiektów poligonowych przechowujących odpowiednio: rejony szkół podstawowych, rejony gimnazjów, rejony szkół średnich i

rejony szkolne przechowujące rejony wszystkich trzech typów szkół.

Rys. 5. Efekt końcowy importu pliku GML do geobazy plikowej.

Zaawansowana konwersja formatów i struktur danych

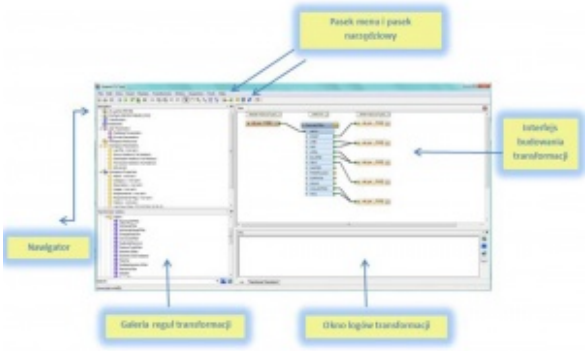
Szybki import czy eksport danych może jednak być niewystarczającym rozwiązaniem, kiedy oprócz zmiany samego formatu staniemy przed wyzwaniem związanym z modyfikacją geometrii lub atrybutów. Gdy chcemy przekształcić dużą liczbę różnie zapisanych danych przestrzennych do ujednoliconego modelu, filtrować dane i połączyć je z innymi klasami obiektów, użyć wartości atrybutów do stworzenia w locie unikalnych klas obiektów lub wyselekcjonować dane przy użyciu różnych kryteriów, wówczas z pomocą mogą przyjść nam narzędzia przestrzennego ETL (ang. Spatial ETL tools).

Narzędzia przestrzennego ETL są stworzonymi przez użytkownika narzędziami geoprzetwarzania, które pozwalają na pozyskanie danych ze źródłowego modelu i dopasowanie ich do docelowego modelu danych dzięki przekształceniom dokonywanym na ich atrybutach i geometrii. Można je tworzyć w obszarze roboczym aplikacji FME Workbench, charakteryzującej się prostym i intuicyjnym interfejsem, w którym użytkownik za pomocą schematu graficznego może ułożyć cały proces transformacji.

Uruchomienie i budowa aplikacji FME Workbench

Aby rozpocząć pracę z aplikacją FME Workbench, a tym samym utworzyć narzędzie przestrzennego ETL, należy z poziomu drzewa

katalogu stworzyć nową skrzynkę narzędziową (ang. Toolbox). Po jej utworzeniu klikamy ją prawym klawiszem myszy i wybieramy New – Spatial ETL Tool, w wyniku czego zostanie otwarty interfejs Create Translation Workspace Wizard, który pozwoli na ustawienie formatu danych wejściowych, wyjściowych, parametrów danych i ich lokalizacji. W wyniku wykonanych ustawień wyświetlona zostanie aplikacja FME Workbench.



Rys. 6. Interfejs użytkownika aplikacji FME Workbench.

Interfejs użytkownika aplikacji FME Workbench podzielony jest na następujące, główne komponenty:

- pasek menu i pasek narzędziowy, zawierający narzędzia pozwalające na poruszanie się po przestrzeni roboczej aplikacji, narzędzia kontroli, narzędzia dodawania i usuwania źródłowych i docelowych danych, narzędzia uruchamiania transformacji, a także odniesienie do pomocy online,
- interfejs budowania transformacji, nazywany także obszarem roboczym, w którym użytkownicy mogą zdefiniować transformację danych, wykorzystując przy tym prosty schemat graficzny; domyślnie przestrzeń robocza interpretowana (czytana) jest od lewej do prawej; po lewej stronie definiowane są dane źródłowe; narzędzia transformacji znajdują się w środkowej części aplikacji, a po prawej definicje danych wyjściowych (docelowych);

połączenia pomiędzy poszczególnymi elementami w przestrzeni roboczej reprezentują przepływ danych i mogą rozgałęziać się w różnych kierunkach,

- nawigator, będący rodzajem drzewa katalogowego, który pokazuje definicję i lokalizację danych źródłowych i docelowych wraz z ich wszystkimi dodatkowymi ustawieniami,
- galeria reguł transformacji umożliwiająca zlokalizowanie, wybór i uzyskanie informacji o konkretnej regule transformacji FME, której chcemy użyć; wszystkie reguły pogrupowane są w kategorie odpowiadające ich przeznaczeniu; najważniejszymi kategoriami są: Calculators (obliczające wartości i dodające je do nowego atrybutu), Database (umożliwiające interakcję z zewnętrznymi, profesjonalnymi bazami danych, pozwalające na składanie zapytań SQL, pozyskanie danych z bazy i połączenie ich z innymi obiektami), Filters (pozwalające na wykonywanie testów geometrii i atrybutów obiektów i wydzielające odpowiednie ścieżki w zależności od ich wyników), Geometric Operators (przetwarzające geometrię obiektów), Infrastructure (zapewniające interakcję z silnikiem FME, dające możliwość zwracania i zapisu logów, możliwość połączenia aplikacji mapowej do podglądu przekształconych danych, zawierające reguły transformacji do tworzenia, usuwania i kopiowania atrybutów), Lists (pracujące z listami atrybutów), Manipulators (modyfikujące geometrię i atrybuty obiektów), Rasters (pracujące z rastrami), String (tworzenie, modyfikacja, usuwanie atrybutów będących ciągiem znaków), Surfaces (pracujące z powierzchniami, np. tworzenie konturów), Web Services (komunikujące się z usługami Web z użyciem protokołu http),
- okno logów transformacji, w którym po uruchomieniu transformacji prezentowane są jej wyniki; znajdziemy w nim informacje o ewentualnych błędach w transformacji, statusie, czasie jej trwania i liczbie przekształconych

obiektów.

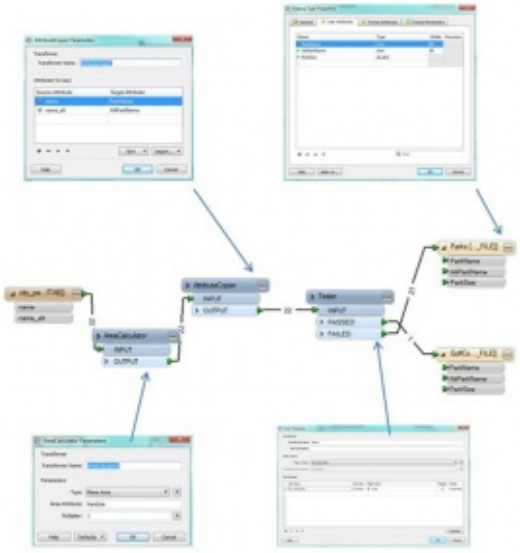
Praktyczne wykorzystanie narzędzi przestrzennego ETL

W omawianym przykładzie jako dane wejściowe dodane zostaną obszary parków zapisane w formacie MITAB, które w dalszej części będą podlegały transformacjom, a na końcu zostanie zasilona nimi geobaza plikowa, dodana jako format danych wyjściowych. Po ustawieniu wszystkich niezbędnych parametrów w oknie Create Translation Workspace Wizard wyświetli się aplikacja FME Workbench z widocznymi w jej centralnej części danymi wejściowymi połączonymi z wyjściowymi klasami obiektów geobazy plikowej za pomocą reguły GeometryFilter, która dokona odpowiedniej filtracji danych, uwzględniając typ geometrii, i pozwoli zasilić nimi odpowiednią klasę obiektów. Gdyby na tym etapie uruchomić transformację, efektem byłaby klasa obiektów poligonowych o atrybutach name oraz name_alt znajdujących się w danych źródłowych.

Aby zapoznać się z podstawowymi możliwościami aplikacji, dokonamy małych przekształceń. Załóżmy, że użytkownik chciałby wyselekcjonować park, który niegdyś był polem golfowym, i utworzyć z niego odrębną klasę obiektów, a następnie obliczyć powierzchnie wszystkich parków i zapisać wyniki obliczeń w nowo utworzonym atrybucie wyjściowych klas obiektów. Aby to zrobić, uporządkujemy najpierw interfejs i usuniemy regułę GeometryFilter oraz wszystkie niepotrzebne klasy obiektów, z wyjątkiem klasy obiektów poligonowych, do której będą ładowane nasze dane. Następnie wybierzemy właściwości docelowej klasy obiektów, klikając ją prawym klawiszem myszy i wybierając opcję Properties. W otwartym oknie Feature Type Properties, w zakładce General, w polu Feature Type Name zmienimy nazwę z domyślnej na Parks, a następnie z poziomu zakładki User Attributes dodamy nowy atrybut ParkSize o typie danych double, który będzie przechowywał obliczone w późniejszym kroku powierzchnie parków. Możemy zmienić także nazwy dwóch źródłowych atrybutów. Kolejnym etapem będzie utworzenie nowej klasy obiektów, która przechowywać będzie pole golfowe.

Dokonać tego możemy, kopiując klasę obiektów Parks i zmieniając jej nazwę na GolfCourses (innym sposobem jest utworzenie nowej klasy obiektów z poziomu paska menu, poprzez wybór opcji Writers – Add Feature Type). Następnie pomiędzy danymi źródłowymi i docelowymi dodana zostanie reguła AttributeCopier, która pozwoli na przeniesienie wartości atrybutów źródłowych do atrybutów o nowych nazwach w klasach obiektów. Łączymy dane wejściowe z portem INPUT dodanej reguły, a połączenie z portu OUTPUT dodajemy do wyjściowych klas obiektów. Kolejną czynnością będzie dodanie reguły Tester, pozwalającej na wyselekcjonowanie parku, który dawniej był polem golfowym. Po wybraniu reguły z galerii reguł transformacji umieszczamy ją pomiędzy AttributeCopier a klasami obiektów (port OUTPUT AttributeCopier połączony zostaje z portem INPUT reguły TESTER, której porty PASSED i FAILED połączone zostaną odpowiednio do GolfCourses i Parks). Aby jednak móc wyselekcjonować nasze pole golfowe, musimy ustalić kryteria testu w parametrach reguły. W sekcji Test Clauses podwójnie klikamy pole Left Value, a następnie z rozwijalnej listy wybieramy opcję Set To Attribute Value i wskazujemy atrybut name_alt. Jako operator wybrany zostanie operator Contains, a jako Right Value wpisane zostanie słowo kluczowe Golf. Ustalone kryterium testowe będzie selekcjonować obiekty, które w starej nazwie zawierają słowo Golf. Na zakończenie dodamy regułę transformacji AreaCalculator, którą połączymy z danymi źródłowymi i regułą AttributeCopier. Jako parametry nowo dodanej reguły transformera ustawiony zostanie typ obliczanej powierzchni na Plane Area i nazwa atrybutu, który zostanie utworzony w wyniku przetwarzania. Nazwa zostanie ustalona jako ParkSize, dzięki czemu zostanie ona automatycznie zmapowana i zapisana w atrybucie ParkSize docelowych klas obiektów. Gdy schemat transformacji został już utworzony, nie pozostaje nic innego jak uruchomienie jej z poziomu paska narzędziowego przy pomocy narzędzia Run. W wyniku przekształcenia dwie nowe klasy obiektów zostaną utworzone w geobazie plikowej, której nazwa i lokalizacja zostały ustalone w początkowym etapie. Nowe klasy obiektów

Parks (składająca się z 22 obiektów) i GolfCourses (jeden obiekt) zawierają dwa przemapowane atrybuty z danych źródłowych i jeden atrybut nowo utworzony w wyniku obliczenia powierzchni parków.



Rys. 7. Schemat transformacji wraz z ustawionymi parametrami reguł transformacji .

Podsumowanie

Bezpośredni odczyt, połączenie interoperacyjne, szybki import i eksport oraz narzędzia przestrzennego ETL to sposoby na ułatwienie codziennej pracy z danymi GIS, zapisanymi w różnych formatach i strukturach. Przedstawione przykłady stanowią jedynie niewielką część możliwości rozszerzenia ArcGIS Desktop Data Interoperability, a zwłaszcza aplikacji FME Workbench, która stanowi potężne narzędzie, dzięki któremu ujednolicenie struktury danych już nigdy nie będzie trudnym zadaniem.