Otwarte standardy zapisu danych geograficznych a ArcGIS

Wspólnym problemem twórców i użytkowników oprogramowania GIS jest przekazywanie informacji geograficznych. Różnorodność danych doprowadziła do powstania dużej liczby formatów ich zapisu. Często zdarza się, że w codziennej pracy z systemami informacji geograficznej otrzymujemy lub odnajdujemy przydatne dane, które niestety zapisane są w formacie nieobsługiwanym natywnie przez używane oprogramowanie. Takie sytuacje doprowadziły do powstania standardów zapisu danych, wykorzystywanych do ich wymiany. W świecie systemów informatycznych takim standardem jest otwarty format zapisu danych XML, natomiast jego modyfikacja, czyli format GML, jest przeznaczona ściśle do zapisu danych geograficznych.

Otwarty format zapisu i wymiany danych XML

XML – eXtensible Markup Language – w wolnym tłumaczeniu oznacza rozszerzalny język znaczników.

Ten prosty i elastyczny sposób tekstowego opisu i przechowywania danych przeznaczony jest do reprezentowania różnego rodzaju informacji w sposób ustrukturyzowany, np. dokumentów, opisów konfiguracji, książek, transakcji, faktur, danych liczbowych, danych geograficznych oraz wielu innych.

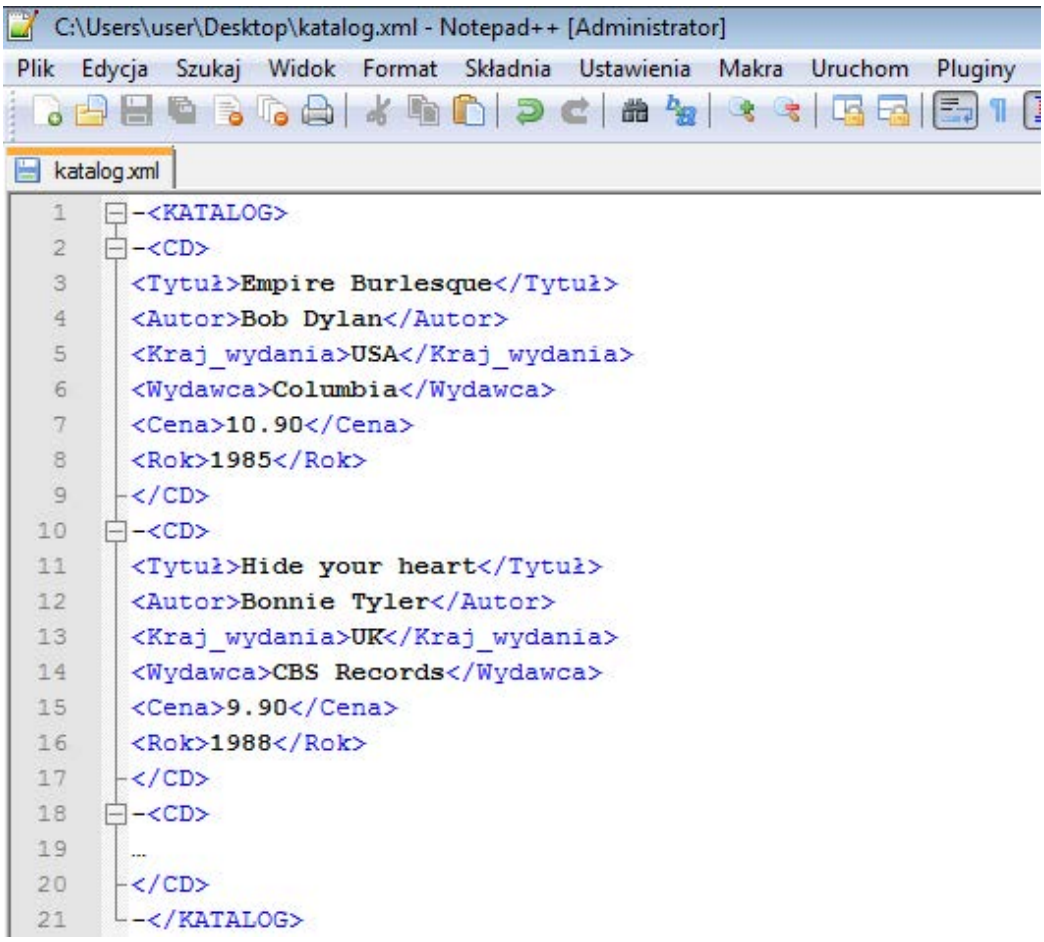
Język XML jest niezależny od platformy programowej i sprzętowej, co znacznie ułatwia wymianę dokumentów pomiędzy różnymi systemami. Obecnie, w dobie Internetu i zwiększającej się z roku na rok liczby informacji przesyłanych w postaci elektronicznej, XML i jego pochodne należą do najpopularniejszych formatów wymiany danych. Oprócz możliwości

przesyłania ustrukturyzowanych informacji pomiędzy programami, XML pozwala na sprecyzowanie jednoznacznych zasad, według których aplikacje będą te dane zapisywać i odczytywać. Każdy twórca oprogramowania, dzięki językowi XML, nie musi być zdany na korzystanie narzuconych z góry sposobów zapisu danych, lecz może stworzyć własny, autorski format dopasowany do potrzeb.

Język XML charakteryzują następujące cechy:

- jest językiem opartym na znacznikach lub tagach, np. <tytuł></tytuł>;
- został zaprojektowany, aby opisywać i przechowywać dane;
- znaczniki w języku XML nie są predefiniowane; użytkownik może (a nawet powinien) samodzielnie definiować tagi, które będą wykorzystywane w utworzonych przez niego formatach zapisu danych;
- jest językiem, który opisuje sam siebie (*self-descriptive*);
- dokument XML zawiera tylko tekst, nie może zawierać danych binarnych;
- model dokumentu XML to struktura drzewa, którego węzłami są elementy lub encje, a liśćmi – pola tekstowe zawierające ostateczne dane.

Na rys. 1 przedstawiono przykładowy dokument XML opisujący katalog płyt CD. Element „katalog” jest elementem nadrzędnym, elementy CD są elementami podrzędnymi. Tagami opisującymi dane są np. <Tytuł></Tytuł> oraz pola tekstowe zawierające właściwe informacje (np. tytuł płyty *Empire Burlesque*).



Rys. 1. Dokument XML zawierający strukturę danych katalogu płyt CD. Dokumenty XML można edytować w każdym edytorze tekstowym, w tym przypadku jest to aplikacja Notepad++.

Schematy i walidacja – XML Schema

Język XML jest luźnym standardem, bardzo elastycznym i dającym dużą swobodę każdemu, kto zechce z niego skorzystać. Powoduje to, że dane zapisane w formacie XML przez jednego użytkownika nie zawsze będą opisane w ten sam sposób, co dane w formacie XML opisane przez innego użytkownika. Struktury zapisu danych różnych autorów mogą się różnić, mimo że autorzy będą opisywali te same dane. Aby ujednolicić i jednoznacznie uzgodnić format zapisu, utworzono schematy struktury dokumentów XML. Jednym ze standardów służących do definiowania struktury dokumentu XML jest XML Schema Definition Language. Dokumenty posiadające definicje XML Schema zapisuje się zwykle

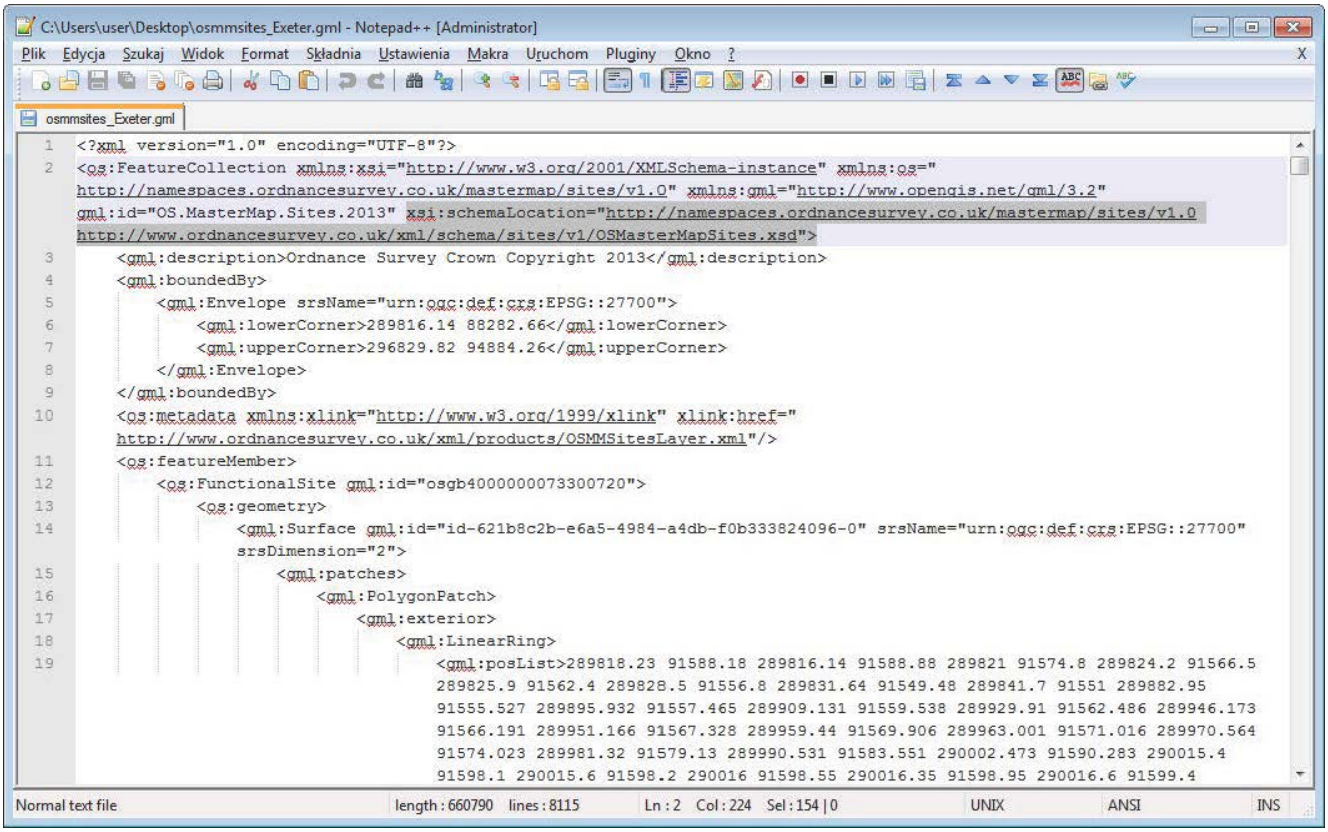
w plikach o rozszerzeniu .xsd (XML Schema Definition).

Plik XSD jest dokumentem, który ściśle określa strukturę dokumentów XML oraz opisuje elementy i typy danych, jakie mogą wystąpić w ostatecznym pliku XML. To zestaw instrukcji, które jasno określają, w jaki sposób należy zapisywać i odczytywać dane w plikach XML. Definiuje on ewentualne wartości domyślne oraz ograniczenia (nie pozwala na zapisanie każdej wartości, dopóki nie spełnia ona konkretnych wymagań). Standard XSD jest również wykorzystywany przez programistów i twórców aplikacji do weryfikowania, czy każdy element opisany w pliku XML jest zgodny z zasadami założonymi przez twórcę danego formatu. Każdy plik XML musi posiadać w części nagłówkowej dokumentu odniesienie do pliku XSD, definiującego jego strukturę. Odniesienie może wskazywać na adres internetowy serwera, który przechowuje plik .xsd lub ścieżkę lokalną komputera, na którym znajdują się pliki XML oraz ich schemat XSD. Do jednego schematu XSD może się odnosić wiele plików XML.

Dzięki temu zagwarantowana jest pełna zgodność i kompatybilność dwóch lub więcej plików XML utworzonych według jednego schematu XSD.

Schemat XSD charakteryzują następujące cechy:

- określa elementy, jakie mogą się pojawić w dokumencie;
- określa atrybuty, jakie mogą wystąpić w dokumencie;
- określa, które elementy są elementami nadrzędnymi, a które podrzędnymi (*parent – child*);
- określa kolejność występowania elementów podrzędnych;
- określa liczbę elementów podrzędnych, jaka może wystąpić w każdym elemencie nadrzędnym;
- określa, kiedy element może pozostać niewypełniony (pusty) lub musi zawierać jakąś wartość (oraz pozwala na ustawienie wartości domyślnej);
- definiuje typy danych dla elementów i atrybutów;
- definiuje wartości domyślne i wartości stałe (*default and fixed values*) dla elementów oraz atrybutów.



Rys. 2. Podgląd pliku danych zapisanych w postaci GML z widocznym odniesieniem do schematu XSD. Dane pochodzą ze strony brytyjskiego wydziału do spraw map – Ordnance Survey (https://www.ordnancesurvey.co.uk/business-and-government/products/mastermap-products.html).

GML – Geography Markup Language

GML to oparty na XML język opisu danych geograficznych, opracowany w 2001 roku przez Open Geospatial Consortium. GML jest uniwersalnym standardem do kodowania, rozpowszechniania i gromadzenia informacji geograficznych. Jest niezależny od jednej, konkretnej platformy programowej – implementacja obsługi formatu GML zależy od autora aplikacji. W 2007 roku specyfikacja języka GML została zatwierdzona przez Międzynarodową Organizację Normalizacyjną jako norma międzynarodowa (ISO 19136, 2007), którą następnie przyjął do oficjalnego zbioru polskich norm Polski Komitet Normalizacyjny (PKN).

Tak jak w przypadku większości formatów opartych na języku

XML, format GML składa się z dwóch części: schematu, który opisuje strukturę dokumentu, oraz właściwego pliku przechowującego dane w odpowiedniej strukturze. Inaczej mówiąc, dokument GML jest opisany schematem GML Schema. Tak więc dokument GML może przechowywać informacje dotyczące poszczególnych obiektów geograficznych, takich jak punkty, linie i poligony. Wraz z rozwojem dokumentacji GML powstało wiele dodatkowych rozszerzeń aplikacyjnych tego formatu, które ułatwiają wymianę danych pomiędzy użytkownikami działającymi w konkretnej dziedzinie. Przykładami takich rozszerzeń mogą być: CityGML – wspólny model danych wykorzystywany przez aplikacje odwzorowujące dane 3D dotyczące miast, LanGML – format wykorzystywany przez inżynierów i geodetów w USA czy INSPIRE – format danych zgodnych z wytycznymi dyrektywy INSPIRE. Używając rozszerzeń aplikacyjnych formatu GML, użytkownicy mogą posługiwać się dokładniejszą terminologią. Terminy ogólne, takie jak punkty, linie i poligony, można zastąpić określeniami bardziej szczegółowymi, które lepiej odzwierciedlają dane używane przez pracowników danej branży, jak np. mosty, skrzyżowania, autostrady, lasy czy zbiorniki wodne.

GML jest mechanizmem, który może być wykorzystany do zakodowania prawie każdego rodzaju danych przestrzennych. Pozwala na zapis zarówno geometrii, jak i atrybutów dotyczących obiektów geograficznych. Dzięki temu umożliwia organizacji zajmującej się danymi GIS publikowanie własnego modelu / standardu informacyjnego, bez względu na stopień jego skomplikowania. Wykorzystanie otwartych formatów zapisu danych zdejmuje również z publikującego potrzebę konwertowania i udostępniania tych samych danych w wielu formatach w celu umożliwienia ich odczytu jak największej liczbie interesariuszy. Wystarczy, że użytkownik końcowy dysponuje oprogramowaniem wyposażonym w parser XML / GML, aby móc odczytać pliki bez korzystania z dodatkowego oprogramowania lub konwertowania ich do pośrednich formatów zapisu danych.

Praca ze zbiorami danych XML / GML w ArcGIS

W celu wykorzystania zbiorów danych zapisanych w formatach XML / GML w ArcGIS, należy użyć rozszerzenia ArcGIS Data Interoperability. Rozszerzenie to zostało utworzone we współpracy z firmą Safe Software i jest oparte na mechanizmie FME (Feature Manipulation Engine), służącym do konwersji danych pomiędzy różnymi formatami ich zapisu.

Rozszerzenie ArcGIS Data Interoperability pozwala na:

- dodawanie obiektów GML bezpośrednio do mapy w programie ArcMap;
- używanie danych GML jako danych źródłowych dla narzędzi geoprzetwarzania (a także generowanie wyniku geoprzetwarzania jako pliku GML);
- konwertowanie innych formatów danych GIS do GML;
- używanie danych GML w aplikacjach 3D – ArcGlobe oraz ArcScene;
- publikowanie danych GML w Internecie – jako usługi WFS przy pomocy ArcGIS for Server (rozszerzenie ArcGIS Data Interoprability for Server);
- bezpośredni odczyt wielu innych formatów danych geograficznych oraz eksport do ponad 50 różnych formatów danych GIS.

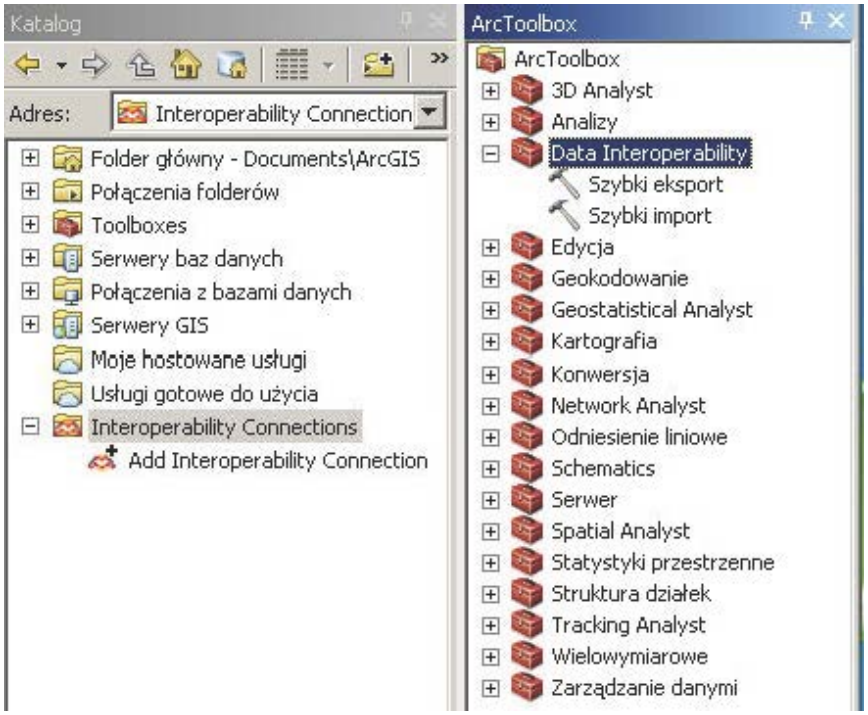
ArcGIS Data Interoperability umożliwia tworzenie schematów aplikacyjnych GML, w których można zdefiniować własną strukturę opisu danych odpowiadającą konkretnym potrzebom. Również w przypadku posiadania zestawu danych GML, który nie jest obsługiwany domyślnie przez rozszerzenie Data Interoperability, możliwość utworzenia i dodania własnego formatu GML do listy pomaga rozwiązać problem z wczytaniem danych z formatu nieobsługiwanego natywnie przez aplikację.

Po zainstalowaniu rozszerzenia ArcGIS Data Interoperability w ArcToolbox pojawią się dwa dodatkowe narzędzia:

- Quick Import – konwertuje dowolny format danych wspieranych przez Data Interoperability do formatu klasy obiektów;
- Quick Export – konwertuje jedną lub więcej klas obiektów lub warstw obiektów ArcGIS do dowolnego formatu wspieranego przez Data Interoperability.

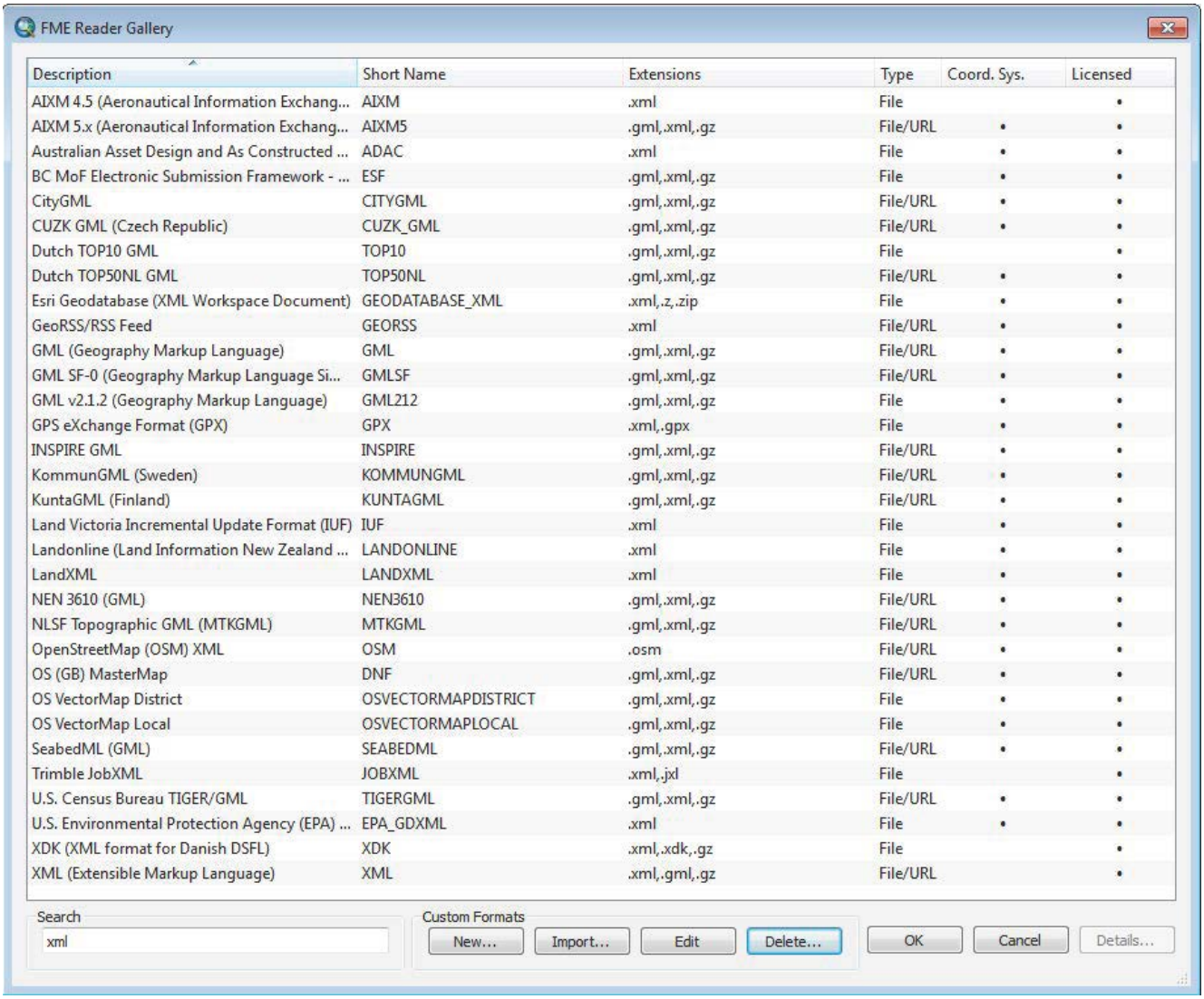
Narzędzia te mogą być użyte jako samodzielne narzędzia geoprzetwarzania, element modelu w ModelBuilder, uruchomione w oknie lub skrypcie Python lub opublikowane jako usługa geoprzetwarzania ArcGIS for Server.

Ponadto w oknie Katalog, zawierającym katalog połączeń do folderów lokalnych, baz danych i serwerów GIS, pojawi się pozycja Interoperability Connections z opcją Add Interoperability Connection. Dzięki temu rozwiązaniu użytkownik ma możliwość „przypięcia" na stałe w oknie Katalog odniesienia do konkretnego zestawu danych (lokalnego lub sieciowego) i nie musi przy każdorazowym uruchomieniu ArcMap odszukiwać danych na dysku lokalnym lub nawiązywać na nowo połączenia ze zdalnym źródłem danych.



Rys. 3. Narzędzia Data Interoperability w aplikacji ArcMap

Po wskazaniu zestawu danych do odczytu lub zapisu można wybrać konkretny format GML lub jego modyfikację z listy FME Writer / Reader Gallery.



Rys. 4. Lista formatów XML / GML oraz ich modyfikacji obsługiwanych przez rozszerzenie ArcGIS Data Interoperability

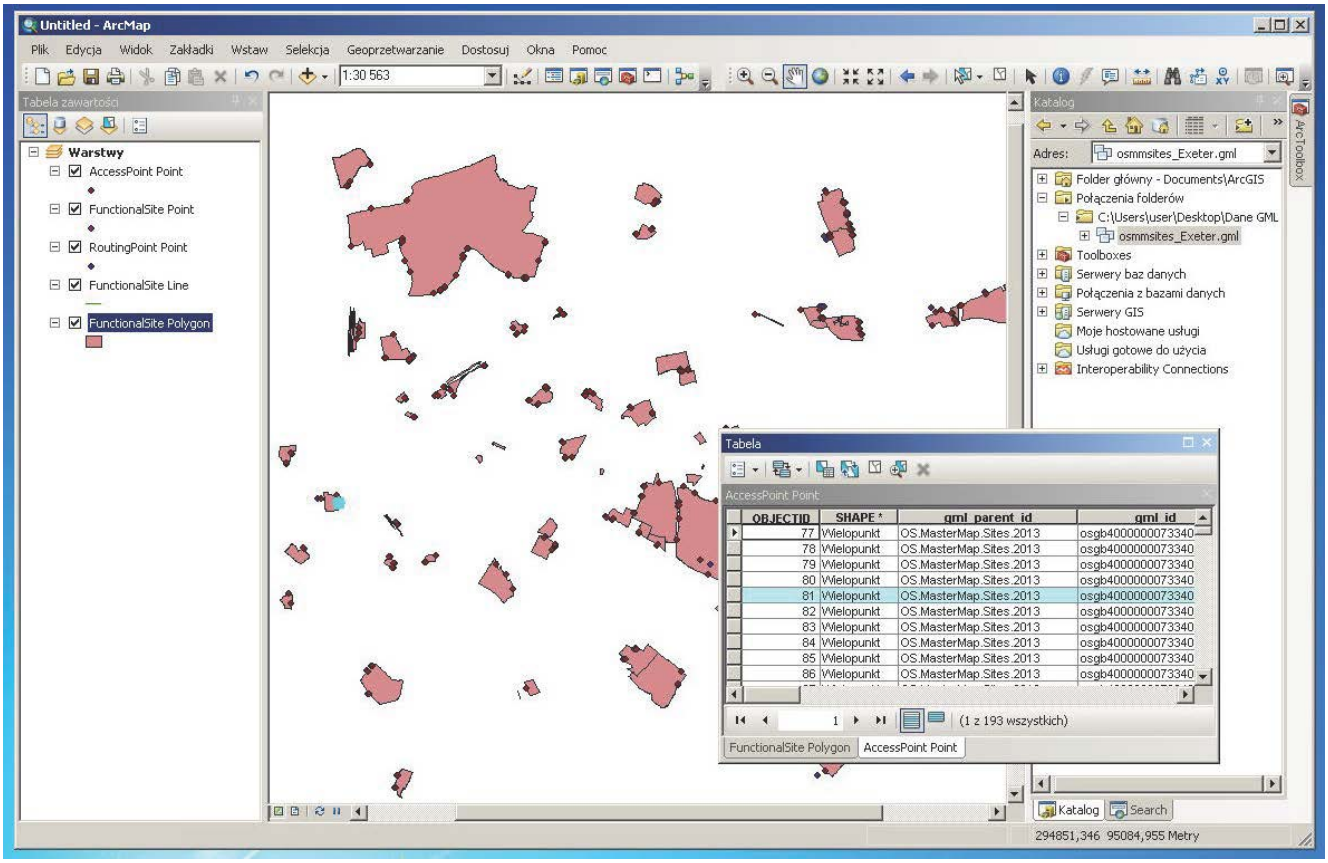
Data Interoperability może być używane na dwóch poziomach licencjonowania.

- Licencja podstawowa – jest bezpłatna i nie wymaga przeprowadzania procesu autoryzacji. Wystarczy samo zainstalowanie rozszerzenia na maszynie roboczej, na której znajduje się ArcGIS for Desktop. W ten sposób uzyskujemy dostęp do serwisów WFS (Web Feature Service – serwis internetowy używający do transportu obiektów

geograficznych technologii XML i GML) oraz wsparcie formatu GML-SF (GML Simple Features – nieco uproszczony w stosunku do oryginalnego formatu GML standard zapisu danych geograficznych opracowany przez OGC). Dane zawarte w serwisach WFS oraz plikach GML-SF można łatwo dodawać do mapy oraz używać ich w narzędziach geoprzetwarzania ArcGIS. Istnieje również możliwość użycia narzędzi konwertujących z ToolBoxa Data Interoperability Tools do wczytywania serwisów WFS oraz odczytu i zapisu plików GML-SF.

- Licencja pełna – pełna funkcjonalność odblokowana po przeprowadzeniu autoryzacji rozszerzenia ArcGIS Data Interoperability for Desktop oraz aktywowania go w aplikacji ArcMap lub ArcCatalog (zakładka Customize>Extensions). W ten sposób otrzymujemy wsparcie oraz możliwość korzystania z ponad 100 formatów wymiany danych GIS, CAD, danych rastrowych, formatów baz danych oraz wielu innych. Ponadto odblokowany zostaje dostęp do aplikacji FME Workbench umożliwiającej tworzenie własnych formatów zapisu danych oraz automatyzacji procesów konwersji.

Dzięki narzędziom Data Interoperability użytkownicy mogą bezpośrednio odczytywać, przeglądać i analizować dane, zapisane nie tylko w formatach XML i GML, lecz także w wielu innych, wykorzystując wszystkie narzędzia dostępne w środowisku ArcGIS Desktop.



Rys. 5. Dane z pliku GML wczytane do aplikacji ArcMap (źródło danych: https://www.ordnancesurvey.co.uk/)

Pełną listę formatów obsługiwanych przez ArcGIS Data Interoperability można znaleźć na stronie http://www.esri.com/software/arcgis/extensions/datainteroperability/supported-formats.