

## Ucywilizować metadane

Tomasz Nałęcz<sup>1</sup>



Na początku lat 90. XX w., kiedy większość z nas dopiero zaczynała przysięgę z systemami informacji przestrzennej, wszystko wydawało się proste. Zasoby danych były bardzo skromne, każdy wiedział, kto i jakimi danymi dysponuje. Informacje o nowych warstwach informacyjnych rozchodziły się dość szybko wśród niewielkiej społeczności, ale też nie było tak wielkiej potrzeby poszukiwania danych, gdyż większość zespołów działała samodzielnie. Obecnie mamy do czynienia raczej z nadmiarem danych, a na pewno coraz częściej pojawiają się trudności z wyborem tych najlepszych pod względem jakości i dokładności. Dziś raczej dyskutuje się o obszernych zasobach danych – Big Data – i sposobach efektywnego zarządzania bazami. Obecnie typowy użytkownik może dotrzeć zarówno do danych wytwarzanych przez organizacje publiczne (np. takie jak: PIG, GDOŚ, GUGiK), jak i skorzystać z tych oferowanych komercyjnie lub na licencji otwartej (*OpenStreetMap*). Ale czym należy się kierować, żeby wybrać najbardziej wartościowe dane, które posłużą do naszych projektów i analiz? To trochę tak, jak w przypadku przeglądania gazety z ogłoszeniami (dziś raczej internetowego portalu ogłoszeniowego!) lub poszukiwania książki w ogromnej bibliotece. Obydwa przykłady są adekwatne do opisywanej sytuacji. Gazeta reklamuje nam produkty z krótką charakterystyką i kieruje do sprzedającego. W bibliotece natomiast musimy skorzystać z katalogu, aby wyszukać numer i charakterystykę poszukiwanego woluminu. W przypadku danych przestrzennych pomagają metadane, które porządkują informacje o danych w postaci katalogu, jak również reklamują poszczególne produkty poprzez ich charakterystykę. W dodatku dzięki metadaniom nie tylko można uzyskać informację o danych, ale też jesteśmy kierowani do producenta (wytwórcy). Metadane to krótka charakterystyka cyfrowego zbioru danych (np. mapa, zdjęcia lotnicze, ortofotomapa) – to chyba najprostsza definicja, wskazująca jasno, że metadane służą użytkownikom do szybkiego wyszukiwania informacji przestrzennej przez odpowiedź na zasadnicze pytania: kto? gdzie? jak? kiedy? Użytkownik otrzymuje szybką i zwięzłą informację o zbiorze danych, czy są to dane rastrowe czy wektorowe, jaki mają układ współrzędnych, kto jest wytwórcą, jaka jest ich dostępność i gdzie można uzyskać bliższe informacje, itp.

Metadane są podstawowym elementem zapewniającym sprawne działanie infrastruktury informacji przestrzennej (PG-2014-07), stanowią także istotny cel dyrektywy INSPIRE (PG-2014-04), umożliwiając sprawne zarządzanie zharmonizowanymi danymi geoprzestrzennymi.

Wyróżniamy trzy poziomy stosowania metadanych i odpowiadające im rodzaje metadanych:

– metadane wyszukiwania (ang. *discovery metadata*), podstawowe informacje o zbiorze danych umożliwiające użytkownikowi odnalezienie zbioru;

– metadane rozpoznania (ang. *exploration metadata*), posiadają rozszerzony zakres pozwalający m.in. określić jakość danych oraz nawiązać kontakt z wytwórcą;

– metadane stosowania (ang. *exploitation metadata*), charakteryzują te elementy zbioru, które są niezbędne do odczytania danych, ich transferu oraz interpretację i praktyczne wykorzystanie przez środowiska systemów użytkownika.

Metadane geoinformacyjne są tworzone dla bardzo różnych zakresów danych. Dyrektywa INSPIRE wymaga tworzenia metadanych dla zbiorów danych przestrzennych, serii danych przestrzennych oraz usług. Jednakże metadane mogą odnosić się do znacznie bardziej szczegółowych poziomów, tj. klas obiektów, instancji obiektu, a nawet typu atrybutów. Im bardziej szczegółowo opisany jest zbiór danych, tym łatwiej przeprowadzić zaawansowane wyszukiwanie prowadzące do uzyskania trafnej informacji.

O tym, jak ważne są metadane dla budowy infrastruktury informacji przestrzennej, niech świadczą słowa zapisane w preambule dyrektywy INSPIRE: „Główną przeszkodą dla pełnego wykorzystania dostępnych danych jest czasochłonność i kosztowność poszukiwania istniejących danych przestrzennych lub sprawdzanie, czy mogą one być użyte w danym celu. Państwa członkowskie powinny zatem dostarczać opisy dostępnych zbiorów danych przestrzennych oraz usług w formie metadanych”. Dyrektywa poświęca metadaniom o wiele więcej miejsca, poprzez dwa szczegółowe dokumenty opisujące zasady wdrażania dyrektywy w zakresie metadanych, opisane poniżej:

– Rozporządzenie Komisji (WE) NR 1205/2008 z dnia 3 grudnia 2008 r. w sprawie wykonania dyrektywy 2007/2/WE Parlamentu Europejskiego i Rady w zakresie metadanych – ustanawia formalne zasady tworzenia profilu metadanych INSPIRE.

– Zasady wdrażania dyrektywy INSPIRE w zakresie metadanych: wytyczne techniczne oparte o EN ISO 18115 i EN ISO 19119 – oficjalna dokumentacja implementacyjna profilu metadanych określająca zależności (mapowanie) pomiędzy profilem INSPIRE a profilem ISO.

Wszystkie te techniczne zasady są oczywiście przeznaczone dla ekspertów i nie powinny w żaden sposób dotyczyć użytkownika. Dlaczego zatem metadane stanowią dla większości odbiorców istotną barierę i są omijane z daleka? Wynika to głównie z niewielkiej popularyzacji metadanych, może być także efektem błędnego ich kojarzenia z zapisem w kodowaniu języka XML (ryc. 1). Język XML posiada niezaprzeczone zalety, głównie poprzez hierarchiczną strukturę, która pozwala na szybkie i sprawne wyszukiwanie informacji. Jednakże jest to język, co trzeba podkreślić, czytelny dla maszyn (komputerów) i raczej

<sup>1</sup> Państwowy Instytut Geologiczny – Państwowy Instytut Badawczy, Rakowiecka 4, 00-075 Warszawa; tomasz.nalecz@pgi.gov.pl

```

<gmd:CI_ResponsibleParty>
  <gmd:organisationName>
    <gco:CharacterString>Państwowy Instytut Geologiczny-Państwowy Instytut Badawczy
    (PIG-PIB)</gco:CharacterString>
  </gmd:organisationName>

  <gmd:contactInfo>
    <gmd:CI_Contact>
      <gmd:phone>
        <gmd:CI_Telephone>
          <gmd:voice>
            <gco:CharacterString>+48224592000</gco:CharacterString>
          </gmd:voice>
        </gmd:CI_Telephone>
      </gmd:phone>
    </gmd:CI_Contact>
  </gmd:contactInfo>

```

Ryc. 1. Przykład kodowania danych w hierarchicznej strukturze języka XML

może być trudny w odbiorze przez użytkowników. Należy w tym momencie rozwiązać wszelkie wątpliwości i kategorycznie stwierdzić, że XML jest językiem przeznaczonym dla ekspertów. Użytkownicy przez klienta usługi sieciowej (np. geoportal) otrzymują tłumaczenie XML na czytelną formę HTML, czyli w postaci standardowej strony internetowej z pogrupowanymi informacjami. Odpowiednia aplikacja pośredniczy między użytkownikiem a zapisami metadanych w XML.

Stworzono dwa narzędzia umożliwiające pominięcie bezpośredniego kontaktu użytkownika z zapisami w XML, są to – katalog metadanych (prezentujący w czytelnej formie wyniki wyszukiwania metadanych) oraz edytor metadanych (służący do wprowadzania informacji przez wytwórców danych, którzy wypełniają zasoby katalogowe). Obydwa te zagadnienia zostaną przedstawione w kolejnym felietonie.

Metadane choć zapisane w niezbyt przyjemnym na pierwszy rzut oka formacie (XML) można łatwo ucywilizować, poprzez stosowanie szeregu przyjaznych użytkownikowi aplikacji, które w sposób intuicyjny prowadzą proces zapytań, a także prezentują wyniki w formie czytelnych dokumentów. Kolejną fazą promocji metadanych są prace, które mają na celu taką transformację z XML do HTML, aby zapytania można było prowadzić bezpośrednio z poziomu wyszukiwarki internetowej (np. google, yahoo), a nie poprzez katalog metadanych.

Budowa infrastruktury informacji przestrzennej (iip) jest złożonym zadaniem międzynarodowym, a wręcz globalnym wymagającym precyzji i spójności pomiędzy poszczególnymi elementami składowymi. Metadane są istotnym, żeby nie powiedzieć podstawowym, czynnikiem zapewniającym działanie całego rozległego systemu gwarantującego interoperacyjność informacji przestrzennej. Dlatego tak ważne jest, aby były tworzone systematycznie i w sposób ujednolicony, zgodnie z przyjętymi standardami. Należy tu wspomnieć o licznych zaletach wynikających z opisu zbiorów danych przez metadane.

Istotną zaletą metadanych jest ułatwienie zarządzania danymi przestrzennymi zarówno na poziomie iip, jak i w samej organizacji, odpowiedzialnej za ich wytworzenie.

Jak się okazuje, opisywanie danych przestrzennych metadanymi znacząco wpływa na usprawnienie zarządzania nimi w samej organizacji, szczególnie w przypadku ogromnych ilości informacji gromadzonych na serwerach. Dostęp do metadanych przez jeden punkt dostępowy pozwala uzyskiwać informację dziedzinową, regionalną czy też historyczną dla całego zakresu danych pochodzących z różnych źródeł. Takie rozwiązanie znacznie ułatwia możliwości dotarcia do danych, ale także wpływa na popularyzację usług geoinformacyjnych.

Należy jednak pamiętać, że metadane to nie tylko korzyści, ale też obowiązki nałożone na organy wiodące w zakresie ich wytwarzania i późniejszej aktualizacji. Zgodnie z zapisami dyrektywy INSPIRE, metadane dla Załącznika I, II i III powinny być gotowe w krajach członkowskich już pod koniec 2013 r.

Metadane, choć pozornie nieprzyjemne użytkownikowi, szczególnie ze względu na stronę techniczną z nimi związaną, mogą w niedalekiej przyszłości stanowić o sile infrastruktury informacji przestrzennej. Już niedługo, w związku z dalszym ogromnym przyrostem danych przestrzennych, metadane staną się kluczowym elementem do pozyskiwania odpowiedniej i oczekiwanej informacji. Dlatego też należy jak najszybciej rozwiązać mity funkcjonujące wokół metadanych i promować je jako proste i użyteczne narzędzie do wyszukiwania informacji przestrzennej. Do sukcesu metadanych niezbędne są dwa elementy: przełamanie przez użytkowników pewnej niechęci do korzystania z nich oraz rzetelne opisywanie metadanymi wytwarzanych zasobów przez wytwórców.

Na zakończenie, żeby zachęcić użytkowników do korzystania z metadanych, należy jeszcze raz posłużyć się przykładem katalogu bibliotecznego. Obecnie wchodząc do biblioteki i korzystając z katalogu, nie przeszukujemy tylko zasobów lokalnych, lecz mamy dostęp do całej sieci i możemy otrzymać odpowiedź wskazującą, że poszukiwana przez nas pozycja jest w innej dzielnicy lub innym mieście. W przypadku metadanych mamy ten sam mechanizm, docieramy do zasobów wytworzonych w różnych miejscach, nie wstając od biurka, przy którym pracujemy. To naprawdę potężne narzędzie.