

## **Budowa modelu jakości i jego rola w opisie jakości danych przestrzennych**

Development of quality model and its application  
in the context of spatial data quality

**Alicja Winiarska**

DahliaMatic, Warszawa

**Słowa kluczowe: jakość danych przestrzennych, informacja przestrzenna, elementy jakości, model jakości, ISO 19157**

Keywords: spatial data quality, geographic information, quality elements, quality model, ISO 19157

### **Wstęp**

Każdy z nas posługuje się danymi przestrzennymi oraz pozyskiwaną na ich podstawie informacją. Są one wykorzystywane w analizach związanych z podejmowaniem decyzji dotyczących: kierunków rozwoju regionów, planowania przestrzennego, rozwoju gospodarki, zarządzania kryzysowego. Dane przestrzenne są także niezbędne w procesie badania trendów zmian lub zjawisk społecznych. Stanowią one nierzadko referencję dla prowadzenia rejestrów tematycznych związanych z działaniem organów administracji publicznej, ośrodków badawczych i naukowych. Są one również danymi operacyjnymi służb mundurowych, ratunkowych.

Wyżej wymieniono jedynie kilka zastosowań danych przestrzennych, jednakże należy zwrócić uwagę, że są one niebagatelne. Ważne jest, że po zakończeniu określonego procesu decyzyjnego, związanego z wykorzystaniem danych przestrzennych powinno pojawić się pytanie – na ile można zaufać wnioskowi z tej analizy? Co wpływa na odpowiedź na to pytanie? Między innymi jest to wiedza, jakiej jakości dane przestrzenne zostały wykorzystane.

### **Znaczenie informacji o jakości danych przestrzennych**

Informacja o jakości danych przestrzennych wpływa na świadome ich produkowanie, wyszukiwanie i wykorzystywanie. Można wyróżnić trzy główne grupy jej odbiorców: producenci, dystrybutorzy oraz użytkownicy. Dla każdej z tych grup będzie ona miała wpływ na inne procesy.

Producenci, znając oczekiwania z rynku dotyczące jakości konkretnego zbioru danych przestrzennych, są w stanie ocenić koszty związane z jego produkcją, na które składają się między innymi: czasochłonność opracowania, poziom wiedzy jaki jest wymagany od kadry, proces testowania i kontroli, niezbędne narzędzia do przygotowania zbioru danych. Na podstawie wymagań użytkowników i informacji o jakości, producenci mają możliwość zarządzać również procesem aktualizacji zbioru danych wymuszonym przez niespełnienie konkretnych wymagań jakościowych.

Dystrybutorzy mogą uzależniać cenę produktu od jej jakości. Informacje o niej są również elementem metadanych, które stanowią sedno usług wyszukiwania danych przestrzennych ułatwiających dostęp do informacji o zbiorach danych. Obecnie wiele profili metadanych uwzględnia już podstawowe elementy jakościowe ujęte między innymi w normie ISO 19115 jako obowiązkowe, jednak odnosi się to głównie do elementów opisowych jakości, a nie elementów policzalnych. Te ostatnie występują jako elementy fakultatywne i w praktyce właściwie nie są stosowane. Dysponowanie informacją o policzalnych elementach jakości stwarza nawet możliwości do podjęcia działań marketingowych związanych ze sprzedażą.

Ostatnią grupą odbiorców informacji o jakości danych przestrzennych, są ich użytkownicy i to w ich interesie powinna ona głównie funkcjonować na rynku. W jaki sposób mogą ją wykorzystać? Jest ona użyteczna już na etapie wyszukiwania potencjalnych zbiorów danych przestrzennych dla realizacji własnych celów – czy to przeprowadzenia analizy przestrzennej, wyboru zbioru referencyjnego dla prowadzenia własnego rejestru, przeprowadzenia procesu decyzyjnego lub też badań jakiegoś zjawiska. Użytkownik, szukając materiałów do realizacji własnych celów, dysponuje katalogiem wymagań co do konkretnego zbioru. Stanowią one kryteria wyszukiwania niezbędnych dla niego zbiorów danych przestrzennych. Umieszczając opis jakości w metadanych, pozwala się użytkownikowi na ocenę lub wyeliminowanie części zbiorów danych przestrzennych z jego obszaru poszukiwań, skrócenie czasu znalezienia odpowiednich danych, a wreszcie uniknięcie wydatków związanych z zakupem niewłaściwych materiałów. Poza poszukiwaniami, informacja o jakości jest niezbędna również w procesie wykorzystywania zbioru danych przestrzennych. Na podstawie informacji o jakości użytkownik jest w stanie ocenić do jakich celów i w jaki sposób może zastosować dane, bądź jakie powinien mieć zaufanie do końcowego wyniku swoich prac, jeśli są one związane z analizą i podjęciem decyzji.

Powyższe dowodzi, że informacja o jakości danych przestrzennych wpływa pozytywnie na użyteczność danych oraz ich interoperacyjność. Daje ona podstawy do oceny sposobu i możliwości wykorzystania konkretnego zbioru we własnych pracach, kalkulacji ewentualnych kosztów materiałów, oszacowania poziomu zaufania do wyników prac wykorzystujących ten zbiór.

Szacowanie jakości danych przestrzennych i pozyskiwanie wiedzy o niej powinno być nieodzownym elementem utrzymania danych. Należy przez to rozumieć odpowiednie planowanie produkcji, aktualizacji, dystrybucji – w ogólności zarządzanie zbiorem danych przestrzennych. Wiedza o jakości wpływa na częstotliwość i zakres aktualizacji danych, opis jakości może decydować o sposobie dystrybucji.

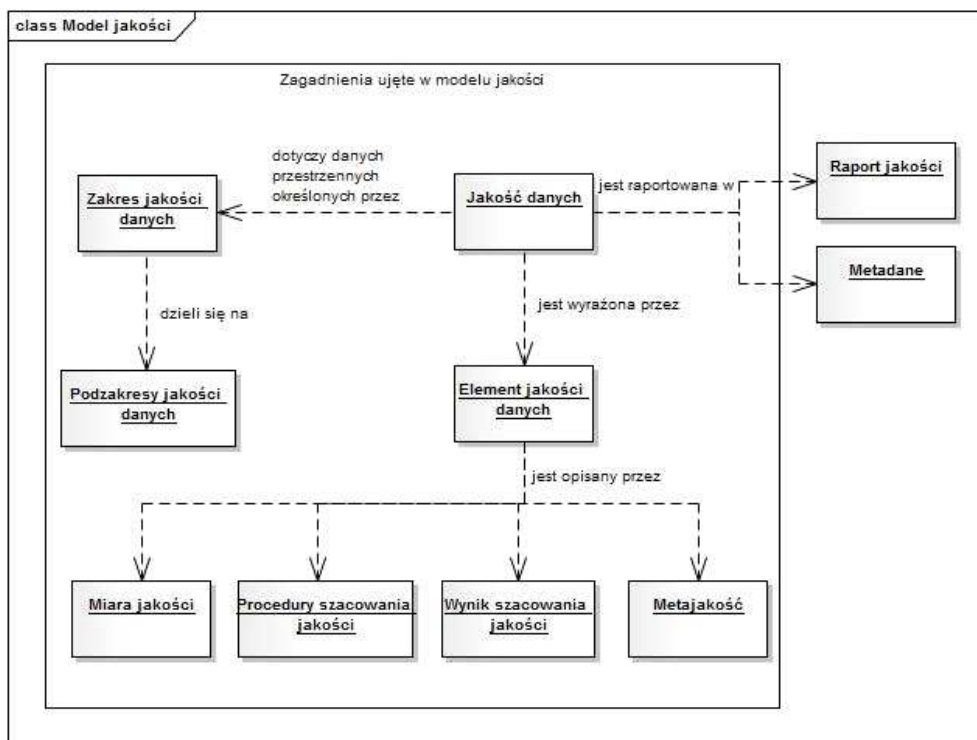
## **Zastosowanie norm ISO**

By kompleksowo i jednolicie przygotowywać opis jakości danych przestrzennych, warto stosować metodykę zawartą w normach ISO – głównie w normie ISO 19157, która zastąpiła od 2013 roku dwie funkcjonujące wcześniej normy PN-EN ISO 19113:2009 Informacja

geograficzna – Podstawy opisu jakości i PN-EN ISO 19114:2005 Informacja geograficzna – Procedury oceny jakości. Podczas przygotowywania opisu jakości warto wykorzystać też dokument *ISO/TS 19138:2006 Geographic Information – Data quality measures*, który dostarcza wiedzy w jaki sposób mierzyć jakość danych przestrzennych. Kolejna ważna norma w kontekście zagadnienia jakości danych przestrzennych to PN-EN ISO 19115:2010 Informacja geograficzna – Metadane, która może mieć zastosowanie przy zagadnieniu raportowania jakości. Zastosowanie powyższych norm w procesie określania jakości danych przestrzennych było opisywane także przez Bielecką (2010) oraz w obszernym przewodniku (Bennat i in., 2015).

## Koncepcja jakości danych przestrzennych a model jakości

W normie ISO 19157 zdefiniowana została koncepcja jakości danych przestrzennych. Na jej treść składają się: definicja zakresu jakości danych przestrzennych (*data quality scope*), elementy jakości danych przestrzennych (*data quality element*) oraz raportowanie wyrażane za pomocą elementów opisanych w metadanych lub zdefiniowanych raportów jakości. Model jakości obejmuje część elementów przytoczonej koncepcji. Zależność pomiędzy modelem a koncepcją przedstawiona jest na rysunku.



**Rysunek.** Zależność pomiędzy koncepcją danych przestrzennych przedstawioną w normie ISO 19157 a modelem jakości (opracowano na podstawie *Figure 1 Conceptual model of quality for geographic data*, ISO 19157, str. 6)

W praktyce narzędziem wspierającym odbiorców danych przestrzennych w opisywaniu, szacowaniu i rozumieniu informacji o jakości tych danych jest model jakości. Pozwala on w formalny sposób, nawiązujący do zapisów norm ISO, opisać istotne w szacowaniu jakości elementy dla konkretnego zbioru danych, a także wspiera w praktycznym wykorzystaniu opisu jakości.

Pojęcie modelu jakości nie jest nowością. Szersze prace dotyczące tego tematu zostały wykonane w ramach projektu ESDIN (*European Spatial Data Infrastructure with a Best Practice Network*), którego podsumowanie opisane zostało przez Beare M. i in. (2010). Opracowano wówczas model jakości dla zbiorów danych z tematów objętych załącznikiem nr 1 do dyrektywy INSPIRE, wykorzystując przy tym specyfikacje danych INSPIRE (*INSPIRE Data Specifications*).

Podobne działania zostały podjęte przez Główny Urząd Geodezji i Kartografii, a związane były z wypracowaniem modeli jakości dla wybranych polskich zbiorów danych przestrzennych – między innymi BDOT10k, PRG, EGiB.

Model jakości według ESDIN, zapewnia powszechne zrozumienie jakości w kontekście specyfikacji danych oraz wymagań użytkowników jak mierzyć tę jakość (Beare i in., 2010). Jest to narzędzie – dokument określany również mianem przewodnika, opisujący wymagania dotyczące danych i metody pomiaru jakości. Podczas tworzenia modelu jakości zastosowanie mają elementy jakości opisane w odpowiednich przywołanych tu normach ISO. Zawartość i zakres, kształt i zastosowanie modelu jest zdeterminowane potrzebami odbiorcy danych przestrzennych.

## Budowa modelu jakości – opis procesu

Jak zatem przygotować model jakości? Proces budowy modelu można przedstawić w następujących krokach:

- 1) określenie wymagań dotyczących danych przestrzennych,
- 2) określenie zakresu jakości danych przestrzennych,
- 3) zdefiniowanie stosowanych elementów jakości i opis ich szacowania:
  - a) zdefiniowanie miar jakości,
  - b) ustalenie granicznej dopuszczalnej wartości wskaźnika jakości,
  - c) dobór metody oceny jakości danych,
  - d) zdefiniowane procesu kontroli,
- 4) testowanie przyjętych założeń modelu jakości,
- 5) weryfikacja przyjętych założeń.

Przygotowując model jakości, w pierwszej kolejności należy zinwentaryzować wymagania dla konkretnego zbioru danych przestrzennych. Katalog wymagań powinien powstać na podstawie zapisów specyfikacji produktu (w praktyce przykładowo w przypadku zbiorów danych państwowego zasobu geodezyjnego i kartograficznego są to rozporządzenia) bądź powszechnie stosowane dobre praktyki (przykładowo zasady budowy geometrii obiektów przestrzennych w bazach danych lub modelowania informacji przestrzennej w GML). Przygotowany katalog wymagań powinien być opisany w sposób zwięzły i klarowny, co znacznie ułatwi późniejsze korzystanie z niego.

Już w trakcie opracowywania katalogu wymagań powinno pojawić się pytanie – na jakim poziomie należy szacować jakość zbioru danych? Określa się go przez zakres jakości zbioru

danych przestrzennych (*data quality scope*), którym może być seria zbioru danych, zbiór danych i podzbiór zbioru danych jednolity pod względem wybranej cechy – zakresu przestrzennego, czasowego, pochodzenia, metod opracowania zbioru, danych źródłowych lub konkretnej wartości wybranego atrybutu. Normy ISO dają dużą elastyczność w opisywaniu zakresu jakości danych przestrzennych, jednak im bardziej skomplikowany jest jego opis, tym trudniejszy proces szacowania.

Zakres jakości danych przestrzennych definiowany jest również przez próbę zbioru danych, jaka będzie poddawana procedurom kontrolnym w procesie szacowania jakości. W celu doboru odpowiedniej próby zalecane jest stosowanie metodyki norm z serii ISO 2859 oraz ISO 3951-1:2005. Dostarczają one między innymi informacji jak zdefiniować próbkę danych oraz w jaki sposób pobrać ją z tego zbioru. W normie ISO 19157 (aneks F) opisana jest interpretacja zapisów przywołanych dokumentów na potrzeby zastosowania ich dla danych przestrzennych. Próbkowanie zbioru danych nie zawsze musi mieć zastosowanie. Jest ono zalecane, gdy proces kontroli nie może być zautomatyzowany, wymaga ingerencji człowieka lub jest bardzo długotrwały. Natomiast przyjmuje się, że jeśli proces kontroli może zostać przeprowadzony w sposób automatyczny, to próba powinna obejmować 100% zdefiniowanego zakresu jakości.

Kolejnym etapem przygotowania modelu jakości jest określenie stosowanych elementów jakości i opis ich szacowania. Należy przez to rozumieć przypisanie do każdego wymagania, dotyczącego danych odpowiedniego elementu jakości, a dalej podelementu jakości, ustalenie stosowanej miary jakości, granicznej dopuszczalnej wartości wskaźnika jakości, określenie procedury kontrolnej dla weryfikacji spełnienia wymagania.

Jest to proces niełatwy, często niestety w pierwszym podejściu do przygotowania modelu jakości oparty na intuicji i weryfikowany w kolejnych krokach procesu budowy modelu jakości. Norma ISO 19157 umożliwia sześć elementów jakości: kompletność, spójność logiczna, dokładność położenia, dokładność tematyczna, dokładność czasowa, wykorzystanie, które można jeszcze uszczegółowić za pomocą 15 podelementów jakości. Dopuszczalne jest również używanie własnych elementów jakości w przypadku, gdy proponowana lista nie jest wystarczająca, jednak należy je precyzyjnie udokumentować.

Dla każdego wymagania należy zdefiniować miarę jakości w jakiej wyrażać się będzie poprawność jego spełnienia. W procesie szacowania jakości danych, a następnie porównywania jakości poszczególnych ich wersji czy też podobnych produktów, niezwykle ważne jest, by stosować zestandaryzowane miary jakości, które są dobrze udokumentowane i powszechnie stosowane. Norma ISO 19157 definiuje w aneksie D listę proponowanych do stosowania miar jakości wraz z ich szczegółowymi definicjami. Miarom jakości poświęcony jest dokument ISO/TS 19138. Dopuszczalne jest stosowanie miar jakości zdefiniowanych przez użytkowników, jednak muszą one być odpowiednio opisane oraz ich definicja powinna być ogólnie dostępna. W celu łatwego zrozumienia informacji o jakości danych przez jej odbiorców, dobrze jest stosować jak najprostsze miary jakości. Używanie bardziej skomplikowanych miar wpływa negatywnie na implementację modelu jakości w narzędziach do kontroli, problem z raportowaniem jakości, porównaniem jej, a ostatecznie z odbiorem i zrozumieniem przy bardzo rozbudowanym procesie szacowania jakości.

Kolejny krok w budowie modelu jakości, to określenie dla konkretnych wymagań granicznej dopuszczalnej wartości wskaźnika jakości. Jest to wartość miary jakości dla wybranego wymagania, której przekroczenie lub nieosiągnięcie oznacza niespełnienie kryteriów jakościowych. Przykładowo dla klasy budynków zbioru EGIB (EGB\_Budynek) model jako-

ści dopuszcza, że w zbiorze może brakować względem świata rzeczywistego pięciu budynków (miara jakości jaka ma wówczas zastosowanie to *error count* dla elementu jakości kompletność, podelementu jakości niedomiar). Jeżeli w zbiorze będzie brakować 4 budynków, to graniczna dopuszczalna wartość wskaźnika jakości nie zostanie przekroczona i zbiór spełni potrzeby użytkownika.

Warto podkreślić, iż koncepcja jakości zbiorów danych przestrzennych dopuszcza, że nie ma idealnych, bezbłędnych zbiorów danych.

Dysponując wymaganiami, przyporządkowaniem ich do konkretnych elementów jakości, określeniem miar jakości oraz granicznych dopuszczalnych wartości wskaźników jakości, należy zdefiniować procedury kontrolne. Pozwolą one na przeprowadzenie kontroli, zdiagnozowanie błędów i wyrażenie poprawności zbioru danych w zdefiniowanej mierze jakości.

Powyższe elementy składają się na procedurę szacowania jakości produktu. Norma ISO 19157 wprowadza klasyfikację metod szacowania jakości na bezpośrednie i pośrednie. Warto podział ten uwzględnić w modelu jakości, gdyż ułatwia późniejsze raportowanie jakości w metadanych. Metoda bezpośrednia polega na ocenie jakości zbioru danych przez porównanie danych z zewnętrznymi bądź wewnętrznymi danymi referencyjnymi. Z tego względu rozróżnia się dwie klasy metody bezpośredniej: wewnętrzną (*internal*) – kontrola wewnętrzna zbioru oraz zewnętrzną (*external*) – kontrola ze zbiorem zewnętrznym. Metoda pośrednia może być oparta o pełną kontrolę obiektów produktu (*full inspection*) lub o próbę obiektów (*sampling*). Pełna kontrola zalecana jest przy możliwości przeprowadzenia kontroli automatycznie. Metoda pośrednia (*indirect*) polega na ocenie jakości danych na podstawie informacji o zbiorze danych, na przykład o ich pochodzeniu.

Przez procedury kontrolne należy rozumieć opis testu, jaki zostanie zastosowany w celu weryfikacji poprawności danych – konkretnego wymagania. Najczęściej jest to analiza przestrzenna, czasem weryfikacja manualna, gdy nie ma możliwości zautomatyzowania kontroli. Najczęściej do oceny jakości ilościowej w zbiorach danych mają zastosowanie metody bezpośrednie. Można uogólnić również, że kompletność zbioru jest weryfikowana głównie przez kontrole ze zbiorami referencyjnymi, natomiast spójność logiczna weryfikowana jest głównie przez kontrole wewnętrzne zbioru.

Należy podkreślić, że sam proces budowy modelu jakości jest iteracyjny. Każdy z wyżej opisanych kroków jest ściśle powiązany z poprzednim i wpływa również na jego kształt. Zależność ta wpływa na zmianę podejścia do modelowania opisu jakości w trakcie procesu budowy modelu jakości.

Po zakończeniu opracowywania modelu jakości, niezwykle istotne jest przetestowanie go na przykładowych danych. Pozwoli to na weryfikację przyjętych założeń i dopasowanie ich do rzeczywistości, a nie jedynie bazowanie na specyfikacji i teoretycznych założeniach, co w praktyce może przynieść problemy w egzekwowaniu zapisów modelu jakości.

Wynikiem procesu szacowania jakości jest informacja o jakości produktu oraz spełnieniu poszczególnych wymagań. W praktyce, pozyskanie takiej informacji na poziomie elementów jakości, wymaga agregacji granicznych dopuszczalnych wskaźników jakości z poziomu podelementów jakości lub poszczególnych procedur kontrolnych. W całym procesie szacowania agregacja końcowa informacji o jakości jest bardzo istotnym zagadnieniem, jednak sam model jakości obejmuje weryfikację na poziomie konkretnych wymagań. Agregacja jest kolejnym krokiem po wykonaniu kontroli opisanych w modelu jakości.

W samym procesie szacowania jakości ważnym etapem jest również raportowanie informacji o jakości danych. Raportowanie dopuszczalne jest w metadanych lub w zdefiniowanych raportach, które muszą być przygotowane czytelnie dla odbiorców. Prowadzenie raportowania jakości w jednolity i niezmienny sposób ma co najmniej dwie zalety – pozwala na pozyskanie informacji o jakości danych przestrzennych, ale też na porównanie zmian w jakości konkretnego zbioru danych na przestrzeni czasu.

### **Wyzwania związane z przygotowaniem i wdrożeniem modelu jakości**

Każdy z etapów budowy modelu jakości i wdrożenie go w procesie produkcji danych powoduje określone wyzwania. Można je podzielić na organizacyjne oraz związane z budową samego modelu. Wszystkie są jednak ze sobą powiązane.

Model jakości ma największe znaczenie dla producentów danych. Dla pozostałych odbiorców informacji o jakości jest on pomocny w rozumieniu zagadnienia jakości w kontekście konkretnych danych. Dystrybutor i użytkownik danych jest już jedynie odbiorcą informacji o jakości danych, która dostarczana jest w raportach tworzonych na podstawie procesu szacowania.

W przypadku producentów danych, niezwykle ważne są wyzwania związane z budową modelu jakości i wdrożeniem go w instytucji. Natomiast dla pozostałych odbiorców informacji o jakości, niezbędna jest edukacja w zakresie rozumienia i posługiwania się modelem jakości bądź przynajmniej raportami jej dotyczącymi.

Producenci, w celu zorganizowania procesu produkcji i utrzymania danych, powinni mieć zaplanowaną strategię zarządzania jakością w instytucji lub system zarządzania jakością, których ważnym elementem jest model jakości. W strategii należy uwzględnić dostępne w instytucji zasoby ludzkie, finansowe, sprzętowe. Wdrożenie modelu jakości jako elementu tej strategii wymaga od instytucji między innymi: wiedzy, zespołu ekspertów, którzy dysponują czasem na prowadzenie procesu oceny jakości produktów, finansowania, świadomości potrzeb użytkowników danych i zdolności szybkiego reagowania na zmiany ich dotyczące.

Kolejnym niebagatelnym wyzwaniem jest wykorzystanie lub przygotowanie systemów zarządzania danymi do obsługi procesu szacowania jakości. Niezbędne jest, by aplikacje GIS będące elementem systemów zarządzania danymi implementowały modele jakości, a tym samym umożliwiły przeprowadzenie pełnego procesu szacowania jakości.

Jak wspomniano wyżej, producenci muszą się zmierzyć z wyzwaniami dotyczącymi budowy modelu jakości. Należą do nich między innymi: zapewnienie dojrzałej specyfikacji produktu, odpowiednie określenie zakresu jakości danych, granicznych dopuszczalnych wartości wskaźników jakości, przygotowanie na proces testowania i ewolucję modelu jakości.

Dysponowanie ostateczną, przetestowaną specyfikacją produktu jest niezwykle istotne ze względu na przygotowanie na jej podstawie katalogu wymagań, który będzie podlegał testowaniu w procesie szacowania jakości. Niekompletność specyfikacji powoduje trudności w opisie wymagań i w ustaleniu granicznych dopuszczalnych wartości wskaźników jakości ich dotyczących. Problemy mogą pojawić się również na etapie procesu szacowania jakości w przypadku odstępstw od specyfikacji, co wynika z różnic pomiędzy światem rzeczywistym a specyfikacją. Może to doprowadzić do uznania za błąd prawdziwej cechy świata rzeczywistego w procesie szacowania.

Kolejnym ważnym zagadnieniem podczas przygotowywania modelu jakości jest ustalenie zakresu jakości (*data quality scope*) dla danych, czyli określenie na jakim poziomie szczegółowości będzie przeprowadzana kontrola, a za tym szacowany wskaźnik jakości. Jak już wspomniano wcześniej, zbyt mały zakres jakości wpływa na skomplikowanie procesu szacowania jakości oraz powoduje wzrost kosztów procesu. Podobnie należy poprawnie wskazać poziom jakości (np. obiektu, atrybutu czy też relacji) jakiego dotyczy konkretne wymagania ze specyfikacji produktu.

Dla powodzenia przeprowadzenia oraz zrozumienia procesu szacowania jakości istotne jest również przygotowanie klarownych definicji diagnozowanych błędów. Norma ISO 19157 nie wymaga tego zabiegu, jednak znacząco wpływa to na podniesienie zrozumienia informacji dotyczącej jakości wśród jej odbiorców. Definicja taka powinna być na tyle uniwersalna i interdyscyplinarna, by odbiorcami jej mogli być także użytkownicy spoza branży danych przestrzennych.

Model jakości powinien być opracowany w sposób klarowny, zwięzły, stosowane definicje powinny być powtarzalne. Wpłyne to pozytywnie na odbiór dokumentu. W przeciwnym przypadku narzędzie to spowoduje duże skomplikowanie całego procesu szacowania jakości, a nie jego uczytelnienie.

Podczas przygotowywania modelu jakości niezwykle przydatne okazuje się korzystanie z dobrych praktyk. Przykładowo, w wielu oprogramowaniach typu GIS lub też systemów do prowadzenia baz danych istnieje zestaw kontroli poprawności geometrii. Jeśli jest on udokumentowany i spełnia oczekiwania, to należy go wykorzystywać, a nie na nowo definiować kontrole, analizy przestrzenne.

Należy pamiętać, że przygotowanie modelu jakości to proces iteracyjny. Podczas realizacji kolejnych etapów przygotowywania, wraca się do zagadnień, które nie zostały odpowiednio szczegółowo opisane we wcześniejszym kroku lub przeciwnie zostały opisane zbyt drobiazgowo.

Dla powodzenia wdrożenia modelu jakości, niezwykle istotne jest testowanie go po zakończeniu przygotowania wstępnej wersji. Model jakości nie powinien opierać się jedynie na teoretycznych założeniach dotyczących zbioru danych, ponieważ rzeczywistość jest o wiele bardziej skomplikowana, co odczuwają szczególnie producenci danych. Przygotowany model jakości należy przetestować na kilku niezależnych zbiorach danych, by móc stwierdzić, że będzie on użyteczny dla końcowych odbiorców. W trakcie tego etapu może okazać się, że specyfikacja danych wymaga uzupełnień, przyjęte graniczne dopuszczalne wartości wskaźników jakości nie spełniają oczekiwań użytkowników lub znacząco wpływają na podniesienie kosztów produkcji danych. Bardzo ważne jest zrównoważenie kosztów produkcji (w tym szacowania jakości danych) i korzyści z tego płynących. Im ostrzejsze kryteria jakości zostaną postawione, tym wyższe będą koszty. Należy weryfikować oczekiwania związane z jakością danych.

W związku z powyższym niewykluczone jest, że model jakości będzie ewoluował nawet po jego wdrożeniu. Wymaga to zaplanowania i przeprowadzenia, z uwzględnieniem potrzeby zachowania spójności, z poprzednią jego wersją. Przez ewolucję należy rozumieć raczej rozbudowę modelu, niż zupełną jego zmianę. Takie działanie zapewni utrzymanie powtarzalności procesu produkcji danych, możliwość porównania jakości konkretnego zbioru danych na przestrzeni czasu.



## Podsumowanie

Model jakości jest narzędziem wpisującym się w metodykę szacowania jakości danych przestrzennych, przedstawioną w normie ISO 19157. Wspiera on odbiorców danych przestrzennych w opisywaniu, szacowaniu i rozumieniu informacji o jakości tych danych.

Zastosowanie modelu jakości pozwala na organizację procesów związanych z szacowaniem jakości jako elementu produkcji i utrzymania danych. Jest to niezwykle ważne dla producentów danych, ze względu na możliwość optymalizacji procesu produkcji i zapewnienie jej powtarzalności. Pozostali odbiorcy informacji o jakości, dzięki modelowi jakości mogą lepiej poznać specyfikę danych i z taką wiedzą efektywniej je wykorzystywać.

### Literatura

- Bearé M., Henriksson R., Jakobsson A., Marttinen J., Onstein E., Lysandros Tsoulos, Williams F., Mäkelä J., Lies De Meulenaer, Persson I., Kavadas I., 2010: D 8.4 ESDIN Quality Final Report – Part A, B, C.
- Bennat H., Endrullis M., Giversen J., Groenendaal W., Hangouët J-F., Lönnberg G., Pross E., Wasström C., 2015: Guidelines for Implementing the ISO 19100 Geographic Information Quality Standards in National Mapping and Cadastral Agencies. Jakobsson A., Giversen J. (eds). Dostęp 2015 r.  
[http://www.eurogeographics.org/sites/default/files/Guidelines\\_ISO\\_19100\\_Quality.pdf](http://www.eurogeographics.org/sites/default/files/Guidelines_ISO_19100_Quality.pdf)
- Bielecka E., 2010: Zasady oceny jakości danych przestrzennych oraz ich zastosowanie do oceny jakości danych gromadzonych w TBD. *Roczniki Geomatyki* t. 8, z. 4(40): 53-66, PTIP, Warszawa.
- ISO, 2006: ISO/TS 19138 Geographic information – Data quality measures.
- PN EN ISO 19157 Informacja geograficzna – Jakość danych.
- PN-EN ISO 19115:2010 Informacja geograficzna – Metadane

### Streszczenie

*W artykule zwraca się uwagę na duże znaczenie informacji o jakości danych przestrzennych, między innymi ze względu na szerokie zastosowanie tych danych w procesach decyzyjnych. Wyróżnieni zostali w nim główni odbiorcy informacji o jakości – producenci, dysponenci i użytkownicy. Znaczenie jej jest dla każdej z tych grup inne.*

*Nakreślono także za normą ISO 19157 koncepcję jakości danych przestrzennych i umiejscowiono w niej model jakości jako istotne narzędzie w procesie szacowania jakości. Model jakości ma na celu zapewnić powszechne zrozumienie tego co oznacza jakość w kontekście specyfikacji danych i opisywać ją jak ją mierzyć.*

*W artykule przedstawiono proces budowy modelu jakości z charakterystyką poszczególnych jego etapów: inwentaryzacji wymagań, definiowania zakresu jakości danych, wyboru elementów jakości i opisu ich szacowania oraz testowania przyjętych założeń. Poza wspomnianym procesem budowy modelu jakości, nakreślono wyzwania związane z jego wdrożeniem. Dotyczą one między innymi potrzeby edukowania odbiorców informacji o jakości, zagadnienia równoważenia kosztów szacowania jakości, implementacji modelu jakości w systemach zarządzania danymi przestrzennymi.*

*Artykuł przeznaczony jest głównie dla producentów danych przestrzennych, gdyż dotyczy bezpośrednio wytwarzania poprawnych danych, spełniających oczekiwania użytkowników. Dzięki zastosowaniu modelu jakości, producenci mają szansę na uzyskanie powtarzalności kontroli, porównywalności wyników szacowania jakości, poprawę reakcji na wymagania rynku.*

*Koncepcja modelu jakości nie jest nowością. Prace związane z tym tematem były już prowadzone w ramach projektu ESDIN, w którym powstał model jakości dla danych INSPIRE – między innymi dla tematów Administrative units, Transport networks. Podobne działania zostały podjęte przez Główny Urząd Geodezji i Kartografii, a związane były z wypracowaniem modeli jakości dla wybranych polskich zbiorów danych przestrzennych, między innymi BDOT10k, PRG.*

### **Abstract**

*The article emphasizes the meaning of spatial data quality information, mostly because of the common usage of that data in decision making processes. The major recipients of quality information were defined, such as producers, administrators and users. The information meaning is different for each of those groups.*

*The concept of data quality information was presented based on ISO 19157, including the data quality model as an important tool in the quality evaluation process. The quality model was developed to assure the common understanding of the quality in the context of data specification and the quality measurements.*

*The article also introduces the process of building the quality model and describes its stages: gathering requirements, defining the data quality scope, choosing the quality elements, describing the evaluation requirements and testing the assumptions. Apart from describing that process, challenges related to its implementation were mentioned. These include educating the recipients of the quality information, balancing the costs of the quality evaluation, implementing the quality model in spatial data management systems.*

*The article is primarily intended for spatial data producers since it deals with production of correct data meeting the user requirements. By applying the quality model the producers can achieve repeatable results of control, comparable results of the quality evaluation and can speed up their reaction to user requirements.*

*The quality model concept is not new. Related research was a part of the ESDIN project; the quality model for INSPIRE data (i.e. administrative units, transport networks) was created in the frames of that project. Similar research was performed by GUGIK; it was related to creating quality models for selected Polish spatial data sets (i.e. BDOT10k – Topographic Object Database, PRG – National Register of Borders).*

mgr inż. Alicja Winiarska  
alicia.winiarska@dahliamatic.pl