



ANALIZA JAKOŚCIOWA DANYCH GEODEZYJNEJ EWIDENCJI SIECI UZBROJENIA TERENU

Monika Siejka, Marek Ślusarski

Uniwersytet Rolniczy im. H. Kollątaja w Krakowie

THE QUALITATIVE ANALYSIS DATA IN SPATIAL REGISTRATION OF UTILITY INFRASTRUCTURE

Streszczenie

Urzędowe bazy danych gromadzące dane przestrzenne powinny zawierać zbiory metadanych, służące do ich opisu. Zasadniczym elementem zbioru metadanych są cechy opisujące jakość i ważność danych przestrzennych. W pracy zaproponowano metodę oceny jakości baz danych geodezyjnej ewidencji sieci uzbrojenia terenu (GESUT) opartą na czterech kryteriach: dokładności położenia, kompletności, aktualności i spójności logicznej. Jakość danych dla cechy dokładność położenia, obliczono na podstawie atrybutu źródło pochodzenia (określone, dla każdego badanego obiektu typu GESUT). Kompletność danych obliczono na podstawie analizy obiektów typu GESUT, które istnieją na cyfrowej mapie zasadniczej, a nie zarejestrowano ich w bazie GESUT. Aktualność danych określono poprzez analizę czasów oczekiwania na wprowadzenie do bazy geodezyjnych operatów pomiarowych. Spójność logiczną oceniono analizując poziom sprzecznych relacji w danych. Wyniki oceny dla poszczególnych własności opisano za pomocą wartości punktowych w skali od 1 do 100. Całościową ocenę jakości danych baz GESUT przeprowadzono poprzez obliczenie dwóch estymatorów: wartości przeciętnej absolutnej oraz współczynnika zmienności średnich. W badanym przypadku: dwie bazy o podobnej organizacji różnią się jakością gromadzonych danych o 16 punktów procentowych.

Przedstawiona metoda to sposób na szybką i wiarygodną ocenę jakości danych GESUT na poziomie podstawowym. Zastosowanie tej metody nie wymaga angażowania dodatkowych środków, wszystkie potrzebne informacje zapisane są w systemie bazy danych.

Słowa kluczowe: GESUT, metadane, jakość informacji przestrzennej

Summary

The official databases that collect spatial data sets, should contain the sets of metadata used to describe them. The main element of the set of metadata are the features describing of the quality and validity of spatial data. The paper proposes a method for evaluating the quality of databases of spatial registration of utility infrastructure (GESUT) based on four criteria: positional accuracy, completeness, timeliness and logical consistency. The quality of property data of positional accuracy is calculated based on the attribute of source (specified for each type of test object of GESUT). Completeness of data was calculated on the basis of GESUT objects that exist on the digital basic map, and did not registered them in a GESUT. Timeliness of data was determined by analysis of time to waiting for the inserting to database geodetic measurements. Logical consistency was assessed by analyzing of level conflicting relationships in the data. The evaluation results for the various properties are described by the value of a point on a scale from 1 to 100. Overall evaluation of the quality of the data bases GESUT was performed by calculating the two estimators: the value of average absolute and factor of medium volatility. In the case at hand: two databases having similar organizations vary itself in the quality of collected data by 16 percentage points. That method is a quick and reliable assessment of the quality of data of GESUT. Use of this method does not require the involvement of additional measures, all necessary information is stored in the database system.

Key words: *spatial registration of utility infrastructure, metadata, spatial data quality*

WSTĘP

Rewolucja komputerowa drugiej połowy XX wieku doprowadziła świat do epoki informacji, w której zasoby gromadzonych danych utrzymywane są przez komputerowe systemy informatyczne. Systemy te umożliwiają proste udostępnianie danych, jak również wykonują złożone analizy w celu zaferowania przetworzonych danych. Po krótkim okresie nieświadomości, że bazy komputerowe przechowują wiarygodną informację, rozpoczęto badania nad szeroko rozumianą jakością danych informatycznych.

Według Redmana [2001] „dane są wysokiej jakości jeżeli można ich użyć w procesach operacyjnych, decyzyjnych i planowania”. Cechy danych dobrej jakości to: dostępność, zrozumiałość, spójność i poprawność, kompletność i użyteczność. Dane powinny posiadać odpowiednie metryki. Metryki jakości danych muszą charakteryzować się czytelnością, mierzalnością i łatwością użycia oraz porównywalnością wyników.

Kompendium infrastruktury danych przestrzennych: The SDI Cookbook [Nebert, 2004] wyróżnia metadane rozpoznania, które pozwalają na ocenę jakości danych zbioru oraz określenie danych zbioru pod względem wymagań użytkownika. Główne elementy standardu CSDGM (US Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata) zawierają (wg

kolejności ważności) [Longley i in., 2006]: podstawową informację o zbiorze danych, informację o jakości danych (ogólna ocena jakości danych w zbiorze), sposób uporządkowania danych przestrzennych w zbiorze i inne.

Charakteryzując jakość danych przestrzennych można użyć kilku różnych własności. Pochodzenie danych, dokładność położenia, dokładność atrybutów, spójność logiczna, kompletność, dokładność semantyczna i jakość czasowa to główne elementy jakości danych [Oort, 2005], [Devillers, 2010]

Według Gaździckiego [2008] jakość danych opisują następujące cechy: kompletność, zgodność logiczna, dokładność pozycyjna, dokładność czasowa, dokładność tematyczna oraz dokładność semantyczna i pochodzenie. Kompletność rozumiana jest jako występowanie wszystkich zamierzonych danych bez niedomiaru i nadmiaru. Zgodność logiczna to brak wewnętrznej sprzeczności w zbiorze danych. Dokładność pozycyjna dotyczy geodezyjnych dokładności – wyrażonych współrzędnymi położenia obiektów. Dokładność czasowa związana jest ze zmianami danych w czasie, a tematyczna to prawidłowość określania np. własności jakościowych. Dokładność semantyczna przedstawiana jako zbiór danych odtwarza przestrzeń rozważań (dziedzina problemu). Pochodzenie opisuje sposób i czas pozyskania danych oraz materiały źródłowe, metody i techniki.

Normy europejskie serii ISO 19100 zawierają szeroki zakres pojęć dotyczących informacji geograficznej oraz charakteryzują się bogatym aparatem pojęciowym. Metodykę opisu jakości danych zawierają normy: PN-EN-ISO 19113 „Informacja geograficzna – Podstawy opisu jakości”, PN-EN-ISO 19114 „Informacja geograficzna – Procedury oceny jakości” oraz ISO 19138 „Informacja geograficzna – Działania w zakresie jakości”. Zgodnie z zapisem norm jakość to „całość charakterystyk produktu, które zależą od jego zdolności do zaspokajania określonych i potencjalnych potrzeb”. Kompletna identyfikacja informacji o jakości powinna zawierać „nie ilościowe” oraz „ilościowe” informacje o jakości. Nie ilościowe informacje o jakości to: przeznaczenie, pochodzenie i wykorzystanie. Ilościowe to m. in.: kompletność, spójność, dokładność oraz miara, data i wynik jakości danych.

OBIEKT BADAWCZY

Geodezyjna ewidencja sieci uzbrojenia terenu (GESUT) – zgodnie z zapisem w ustawie Prawo geodezyjne – to uporządkowany zbiór danych przestrzennych i opisowych sieci uzbrojenia terenu, a także informacje o podmiotach władających siecią. Sieć uzbrojenia terenu to wszelkiego rodzaju nadziemne, naziemne i podziemne przewody i urządzenia (również budowle podziemne). Obecnie funkcjonujące bazy danych GESUT tworzone były zgodnie z zapisami Rozporządzenia z 2001 roku [Rozp. 455, 2001] oraz wytycznych zawartych w instrukcji technicznej G-7 [Instr. tech. G-7, 1998].

Baza GESUT zakładana jest w oparciu o zasób geodezyjny i dane instytucji zarządzających sieciami. Podstawowymi źródłami danych są: archiwalne operaty pomiarowe z geodezyjnej inwentaryzacji sieci uzbrojenia terenu, mapa zasadnicza, ewidencja gruntów i budynków oraz materiały branżowe.

Badania doświadczalne jakości danych GESUT wykonano na dwóch obiektach testowych. Do badań wybrano dwa miasta średniej wielkości (miasta na prawach powiatu) położone w woj. śląskim. W obydwu miastach – nazywanych w dalszej części opracowania jako obiekt A i obiekt B – GESUT założony został i jest prowadzony przy pomocy systemów informatycznych, spełniających wymogi instrukcji technicznych G-7 i K-1.

W przypadku obiektu A bazę GESUT założono głównie w oparciu o istniejący, bogaty i usystematyzowany zasób geodezyjnych operatów pomiarowych inwentaryzacji sieci uzbrojenia terenu. Na terenie obiektu A istnieją sieci, dla których nie przeprowadzono pełnej identyfikacji. Głównie są to sieci specjalnego przeznaczenia (np. wykorzystywane przez byłe obiekty przemysłowe). Lokalizacja tych sieci jest przedstawiona na mapie zasadniczej.

Na obszarze obiektu B bazę GESUT zakładano – w podobnym zakresie – w oparciu o mapę zasadniczą i geodezyjne operaty pomiarowe. Na terenie tego obiektu wykonano praktycznie pełną identyfikację wszystkich rodzajów sieci.

METODYKA BADAŃ

Badania doświadczalne przeprowadzono analizując informacje zawarte w bazach GESUT i numerycznej mapie zasadniczej dla obu obiektów testowych. Analizowano informacje opisujące obiekty typu GESUT: odcinek przewodu (liniowy fragment sieci o jednakowych cechach), obiekt punktowy (np. studzienka, zasuwa, hydrant) oraz budowla podziemna (np. tunel, parking). Ocena jakości danych GESUT przeprowadzono opierając się na czterech własnościach: dokładności położenia, kompletności, aktualności i spójności logicznej. Wyniki oceny dla poszczególnych własności opisano za pomocą wartości punktowych w skali od 1 do 100.

Jakość danych dla własności dokładność położenia obliczono na podstawie atrybutu źródło danych określonego, dla każdego badanego, pojedynczego obiektu bazy GESUT. W obydwu obiektach badawczych występuje pięć rodzajów źródeł danych. Źródło danych – określone symbolem „O” – oznacza położenie obiektu wyznaczone na podstawie pomiarów geodezyjnych wykonanych w oparciu o osnowę (tab. 1). „A” – pomiar w oparciu o lokalizator przewodów, „D” – digitalizacja mapy zasadniczej, „B” – dane branżowe, „M” – dane projektowe. Źródłom danych przypisano miary dokładności i wagi ważności. Następnie obliczono ważone wartości punktowe cechy dokładność położenia (tab. 1).

Tabela 1. Obliczenie wartości punktowych oceny jakości danych dla własności dokładność położenia
Table 1. Calculating point value of the evaluation of the quality data for the positional accuracy

Symbol źródła danych	Miara dokładności [m]	Waga ważności	Obiekt A		Obiekt B	
			Punkty	Punkty ważone	Punkty	Punkty ważone
O	0,15	1,00	69,2	69,2	41,5	41,5
A	0,25	0,36	8,7	3,1	5,1	1,8
D	0,45	0,11	19,6	2,2	49,5	5,5
B	0,60	0,06	1,8	0,1	2,5	0,2
M	0,85	0,03	0,7	0,0	1,4	0,0
Suma:			100,0	74,6	100,0	49,0

Źródło: Opracowanie własne
 Source: Own calculations

Kompletność to kryterium określające stosunek liczby zgromadzonych danych do liczby danych, które powinny być zgromadzone zgodnie z przyjętym modelem teoretycznym. Dla badanych baz kompletność danych obliczono na podstawie analizy obiektów typu GESUT, które istnieją tylko na cyfrowej mapie zasadniczej. Obiekty te nie zostały zarejestrowane w bazie GESUT głównie z powodu braku pełnej identyfikacji.

Aktualność danych zależy od charakteru zmienności danych oraz przyjętej metody ich aktualizacji. Infrastruktura techniczna – szczególnie na terenach zurbanizowanych – charakteryzuje się ciągłymi zmianami, na co wpływają procesy modernizacji i rozbudowy sieci uzbrojenia terenu. Zapewnienie wysokiego poziomu aktualności tego typu danych wymaga stosowania bieżącej aktualizacji, czyli wprowadzania zmian do bazy GESUT każdorazowo po jej zaistnieniu. W obydwu badanych przypadkach założono prowadzenie aktualizacji bieżącej. Analiza czasów oczekiwania na wprowadzenie do bazy GESUT geodezyjnych operatów pomiarowych pozwoliła na określenie wartości kryterium, aktualność danych.

Spójność logiczna – rozumiana jako brak sprzecznych relacji w danych – oceniona została bardzo wysoko ze względu na przestrzeganie istniejących w tym zakresie przepisów prawa i norm technicznych przy zakładaniu bazy GESUT. W tabeli 2 przedstawiono wartości punktowe oceny jakości danych dla czterech cech opisane za pomocą wartości punktowych w skali od 1 do 100.

Całościową ocenę jakości danych baz GESUT przeprowadzono poprzez obliczenie dwóch estymatorów: wartości przeciętnej absolutnej (S_a) oraz współczynnika zmienności średnich (A). Wartości estymatorów obliczono na podstawie oceny punktowej kryteriów jakości danych oraz ustalonych wag ważności (tab. 2). Współczynnik S_a – średnia ze wszystkich kryteriów jakości danych – pokazuje jaką część danych spełnia założone kryterium 100 punktów. Parametr A pozwala na porównanie zróżnicowania w dwóch różnych rozkładach. W tym

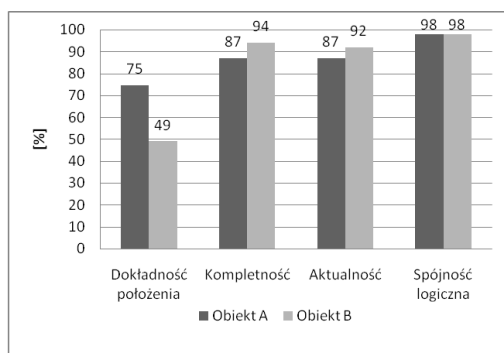
przypadku niejednorodnych pod względem jakości danych baz GESUT dla obiektów A i B. Ideowy model to taki, w którym współczynnik zmienności średnich równa się zero.

Tabela 2. Wartości punktowe oceny jakości danych i wartości obliczonych estymatorów
Table 2. Point values of data quality assessment and values of calculated estimators

	Kryteria jakości danych				Estymatory oceny	
	Dokładność położenia	Kompletność	Aktualność	Spójność logiczna	Wartość przeciętna absolutna (S_a)	Współczynnik zmienności średnich (A)
	waga = 1	waga = 1	waga = 1	waga = 0,5		
Obiekt A	75	87	87	98	85	9
Obiekt B	49	94	92	98	81	25

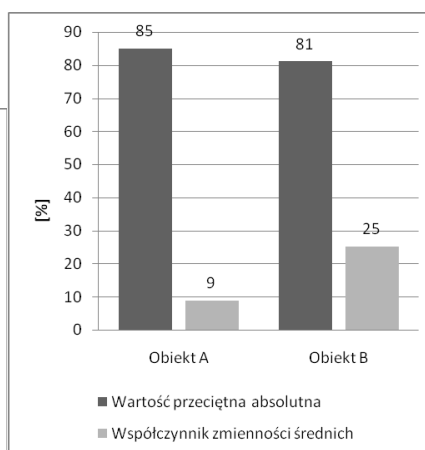
Źródło: Opracowanie własne
 Source: Own calculations

Przedstawione na rys. 1 wartości punktowe oceny jakości danych pokazują, że baza GESUT dla obiektu A charakteryzuje się zdecydowanie wyższą wartością parametru dokładność położenia (75 punktów) w stosunku do obiektu B (49 punktów). Pozostałe kryteria jakości – kompletność, aktualność i spójność logiczna – mają nieznacznie wyższe wartości dla obiektu B. Model koncepcyjny systemu GESUT zrealizowany jest w obu przypadkach w podobnym stopniu: wartość przeciętna absolutna dla obiektu A wynosi 85 punktów, dla obiektu B – 81 punktów (rys. 2).



Źródło: Opracowanie własne
 Source: Own calculations

Rysunek 1. Wartości punktowe oceny jakości danych



Źródło: Opracowanie własne
 Source: Own calculations

Figure 1. Point values of data quality assessment

Rysunek 2. Wartości estymatorów oceny

Figure 2. Value of the assessment estimators

Współczynniki zmienności średnich (wyrażone w procentach) wynoszą odpowiednio 9 i 25 (rys. 2). Można z tego wnioskować, że dla obiektu A jakość danych GESUT jest wyższa o 16 punktów procentowych od danych dla obiektu B.

WNIOSKI

Bazy danych gromadzące dane przestrzenne powinny zawierać zbiory metadanych, służące do ich opisu w sposób umożliwiający określenie zakresu przydatności danych. Kluczowym składnikiem zbioru metadanych są własności opisujące jakość i ważność danych przestrzennych.

Zaproponowany w pracy sposób oceny jakości baz danych GESUT oparto na czterech kryteriach: dokładność położenia, kompletność, aktualność i spójność logiczna. Wykonane prace doświadczalne na dwóch obiektach pokazują, że te cztery parametry umożliwiają sporządzenie precyzyjnego i wiarygodnego raportu oceny. W analizowanym przypadku, to zaobserwowanie różnic w jakości danych dla systemów urzędowych tworzonych w oparciu o te same przepisy prawne i normy techniczne. Dobrym parametrem wyniku jakości danych może być współczynnik zmienności średnich pozwalający na porównywanie zróżnicowania w różnych rozkładach. W badanym przypadku: dwa obiekty o podobnej organizacji różnią się jakością gromadzonych danych o 16 punktów procentowych.

Przedstawiona metoda to sposób na szybką i wiarygodną ocenę jakości danych GESUT na poziomie podstawowym. Zastosowanie tej metody nie wymaga angażowania dodatkowych środków, wszystkie potrzebne informacje zapisane są w systemie bazy danych.

BIBLIGRAFIA

- Devillers R. (red.) (2010). *Fundamentals of spatial data quality*. Hoboken, NJ: Wiley-ISTE, s. 178-197.
- Gaździcki J. *Leksykon geomatyczny* [online]. <http://www.ptip.org.pl> [dostęp 08.01.2013].
- Instrukcja Techniczna G-7. (1998). *Geodezyjna ewidencja sieci uzbrojenia terenu (GESUT)*. Warszawa: GUGiK.
- Longley P. i in. (2006). *GIS teoria i praktyka*. Warszawa: PWN, s. 259-260.
- Nebert D. (red.) *The Spatial Data Infrastructure Cookbook* [online]. <http://www.gsdi.org>. [dostęp 08.01.2013].
- Norma PN-EN-ISO 19113. (2009). *Informacja geograficzna – Postawy opisu jakości*. Warszawa: PKN.
- Norma PN-EN-ISO 19114. (2005). *Informacja geograficzna – Procedury oceny jakości*. Warszawa: PKN
- Norma ISO 19138. (2006). *Geographic information – Data quality measures*. Geneva: International Organization for Standardization.
- Oort P. (2005). *Spatial data quality: from description to application*, Delf: NCG, s. 13-18.
- Redman T. (2001). *Data quality. The Field Guide*. Boston: Digital Press, ss. 233.

Monika Siejka, Marek Ślusarski

Rozporządzenie Min. Rozw. Reg. i Bud. z 2 kwietnia 2001 r. w *sprawie geodezyjnej ewidencji sieci uzbrojenia terenu oraz zespołów uzgadnia dokumentacji projektowej* (Dz. U. Nr 38 poz. 455).

Dr inż. Monika Siejka,
Dr inż. Marek Ślusarski
Katedra Geodezji,
Uniwersytet Rolniczy w Krakowie
e-mail:rmwiech@cyfkr.edu.l
e-mail:rmslusar@cyfronet.pl
ul. Balicka 253a
30-198 Kraków
tel. +4812-662-45-15