

Marek Ślusarski

**DOSKONALENIE JAKOŚCI INFORMACJI
W POWIATOWYCH SYSTEMACH
GEOINFORMACYJNYCH**

***IMPROVING THE DATA QUALITY IN THE COUNTY'S
GEOINFORMATION SYSTEM***

Streszczenie

Komputerowe bazy danych przechowujące informacje opisujące przestrzeń powinny podlegać ocenie jakościowej. Wymóg ten szczególnie dotyczy urzędowych systemów geoinformacyjnych. W pracy przedstawiono propozycję ogólnej oceny jakości informacji gromadzonej w powiatowych bazach danych. Analizowano sześć grup tematycznych dotyczących: ewidencji gruntów i budynków, numerycznej mapy zasadniczej, geodezyjnej ewidencji sieci uzbrojenia terenu oraz poziomej osnowy geodezyjnej III klasy. Badania doświadczalne zrealizowano w dwóch powiatach położonych w woj. małopolskim. Ocenę jakości danych przeprowadzono w oparciu o cztery własności: dokładność, kompletność, aktualność i zgodność. Następnie obliczono wartości dwóch estymatorów ogólnej oceny baz geoinformacyjnych: wartość przeciętną absolutną i współczynnik zmienności średnich. Ogólna ocena jakości danych geoinformacyjnych w obydwu powiatach jest dobra.

Proces doskonalenia jakości informacji przedstawiano zestawiając wyniki obecnych prac z badaniami przeprowadzonymi w 2005 roku. W analizowanym okresie poprawiono ocenę ogólną jakości danych w obu powiatach o około 5 punktów procentowych. Dokładność, aktualność i zgodność danych powiatowych systemów jest dobra. Natomiast kompletność danych ciągle nie spełnia wymaganych kryteriów.

Słowa kluczowe: system informacji przestrzennej, metadane, jakość informacji przestrzennej

Summary:

Computer databases that store information describing the space should be judged qualitatively. This requirement especially applies to the official geoinformation systems. The paper presents a proposal for an overall assessment of the quality of the information collected in the county's databases. It was analyzed six thematic groups: land and buildings cadastre, base numerical map, spatial registration of utility infrastructure and horizontal control class III. Experimental studies carried out in two counties located in the Malopolska province. Data quality assessment was based on four characteristics: accuracy, completeness, timeliness and consistency. Then calculated values of the two estimators of the overall assessment of geo-information databases: the average absolute value and coefficient of variation of averages. Overall assessment of quality geospatial data in the two counties is good.

The process of improving the quality of informations presented comparing the results of current work to a study conducted in 2005. In this period fixed overall assessment of data quality in both counties by about 5 percentage points. Accuracy, timeliness and consistency of county data systems is good. While the completeness of the data still does not meet the required criteria.

Key words: *spatial information system, 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.

Na uwagę zasługuje praca Redmana [2001] porządkująca wiedzę pierwszych badaczy. „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 powinny charakteryzować się czytelnością, mierzalnością i łatwością uzyskania i porównywalnością wyników.

Systemy geoprzestrzenne opisujące część rzeczywistości związanej z przestrzenią ,przechowują informację o jakości danych w zbiorach metadanych.

Kompedium 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.

Normy europejskie serii ISO 19100 dotyczą szerokiego zakresu pojęć informacji geograficznej. Charakteryzują się bogatym aparatem pojęciowym, są formalną dokumentacją rozwoju systemów informacyjnych. Metodykę opisu jakości danych prezentuje norma PN-EN-ISO 19113 „Informacja geograficzna – Podstawy opisu jakości”. Zgodnie z zapisem normy jakość to „całość charakterystyk produktu, które zależą od jego zdolności do zaspokajania określonych i potencjalnych potrzeb”. Kompletny opis – identyfikacja informacji o jakości – powinien zawierać nieilościowe i ilościowe informacje o jakości oraz parametr meta jakości. Nieilościowe informacje o jakości to: przeznaczenie, pochodzenie i wykorzystanie. Ilościowe to m. in.: kompletność, spójność, dokładność oraz miara, data i wynik jakości danych.

Charakteryzując jakość danych przestrzennych można użyć kilku różnych własności. 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.

W roku 2005 wykonano badania empiryczne dotyczące jakości danych podstawowych baz geoinformacyjnych, gromadzonych w dwóch wydziałach geodezji starostw powiatowych [Ślusarski 2006]. Wynikiem badań było sformułowanie tzw. ogólnej oceny jakości danych, opartej na estymatorach: wartość przeciętna absolutna i współczynnik zmienności średnich. W 2010 roku przeprowadzono kolejne badania. Zastosowana metoda i obszar badawczy analogiczny jak w 2005 roku. Porównanie i analizy wyników badań poprzednich i obecnych mają pokazać za pomocą odpowiednich estymatorów proces doskonalenia jakości informacji geoprzestrzennej poziomu powiatowego w okresie ostatnich pięciu lat.

METODYKA BADAŃ

Badania przeprowadzono w dwóch powiatach, leżących w województwie małopolskim. W prezentowanym opracowaniu powiatom nadano symbole A oraz B. W powiecie pierwszym – (A) do gromadzenia, przechowywania i udostępniania informacji geoprzestrzennych wykorzystywany jest zintegrowany

system EWID 2007. W powiecie drugim – (B) wykorzystywane są produkty firmy Geobid. Narzędziem grafiki komputerowej jest program EWMAPA, a do obsługi baz przechowujących informacje opisowe stosowane są m. in. programy EWOPIS, SESUT i BANK-OSN.

Ocenie poddano reobazy, zarządzające informacjami sześciu grup tematycznych.

1. Część opisowa ewidencji gruntów.
2. Kartoteka budynków i lokali.
3. Numeryczna mapa ewidencyjna.
4. Numeryczna mapa zasadnicza.
5. Geodezyjna ewidencja sieci uzbrojenia terenu.
6. Pozioma osnowa geodezyjna III klasy.

Badania prowadzono bezpośrednio w jednostkach wydziałów geodezji starostw powiatowych. Dodatkowo, wykonano prace terenowe, polegające na weryfikacji poprawności zapisanych informacji w bazach geoinformacyjnych ze stanem faktycznym. Badania terenowe wykonano w wybranych (reprezentatywnych) obrębach ewidencyjnych.

Ocenę jakości danych systemów powiatowych przeprowadzono opierając się na czterech własnościach: dokładności, kompletności, aktualności i zgodności. W celu zbadania tych własności, a następnie określenia miary oceny systemu, atrybutom nadano charakter probabilistyczny, ustalając im wagi ważności. Dla czterech pierwszych grup tematycznych ustalono wagi równe 1, grupie piątej i szóstej przypisano wagi równe 0,8 (tab.1). Zestawione w formie tabelarycznej atrybuty z opisującymi je własnościami przedstawiono dla powiatów A i B w tab. 1.

Tabela 1. Wartości punktowe oceny jakości danych w powiatach A i B
Table 1. Point values of data quality assessment in districts A and B

Nazwa	Waga	Kryteria jakości danych							
		Dokładność		Kompletność		Aktualność		Zgodność	
		A	B	A	B	A	B	A	B
Część opisowa ewidencji gruntów	1	97	96	98	99	85	85	98	89
Kartoteka budynków i lokali	1	96	97	35	15	92	92	98	99
Numeryczna mapa ewidencyjna	1	89	89	100	100	81	89	96	96
Numeryczna mapa zasadnicza	1	90	93	15	6	77	80	97	97
Geodezyjna ewidencja sieci uzbrojenia terenu	0,8	98	96	5	7	92	90	99	99
Pozioma osnowa Geodezyjna III klasy	0,8	98	97	96	82	95	98	90	90

Źródło: opracowanie własne
Source: own calculations

Prezentowane wyniki oceny systemu geoinformacyjnego opisano za pomocą wartości punktowych w skali od 1 do 100.

Ocenę jakości danych geoprzestrzennych przeprowadzono obliczając wartość przeciętną (Wp) oraz współczynnik zmienności (λ). Wartości obliczonych estymatorów zamieszczono w tabeli 2.

Tabela 2. Estymatory oceny jakości danych w powiatach A i B
Table 2. Estimates of data quality assessment in districts A and B

Estymator	Kryteria jakości danych							
	Dokładność		Kompletność		Aktualność		Zgodność	
	A	B	A	B	A	B	A	B
Wartość przeciętna (Wp)	94	95	58	52	86	89	98	96
Współczynnik zmienności (λ)	0,04	0,03	0,71	0,83	0,11	0,12	0,02	0,04

Przedstawione w tabeli 2 wartości obliczonych estymatorów – wartość przeciętna oraz współczynnik zmienności, pozwalają przeprowadzić ocenę porównawczą jakości danych w obydwu powiatach.

Ocenę funkcjonowania całego systemu geoinformacyjnego przeprowadzono poprzez obliczenie wartości przeciętnej absolutnej (Sa) oraz współczynnika zmienności średnich (A). Estymator Sa – średnia ze wszystkich kryteriów jakości danych – pokazuje jaka część danych spełnia założone kryterium 100 punktów. Parametr A pozwala na porównanie zróżnicowania w dwóch różnych rozkładach. W tym przypadku niejednorodnych baz geoinformacyjnych w powiatach A i B. Ideowy model to taki, w którym wszystkie założone zadania są w pełni realizowane. Wówczas średnia absolutna wynosi 100, a współczynnik zmienności średnich równa się 0. Ocenę ogólną jakości danych gromadzonych w powiatach prezentuje tabela 3.

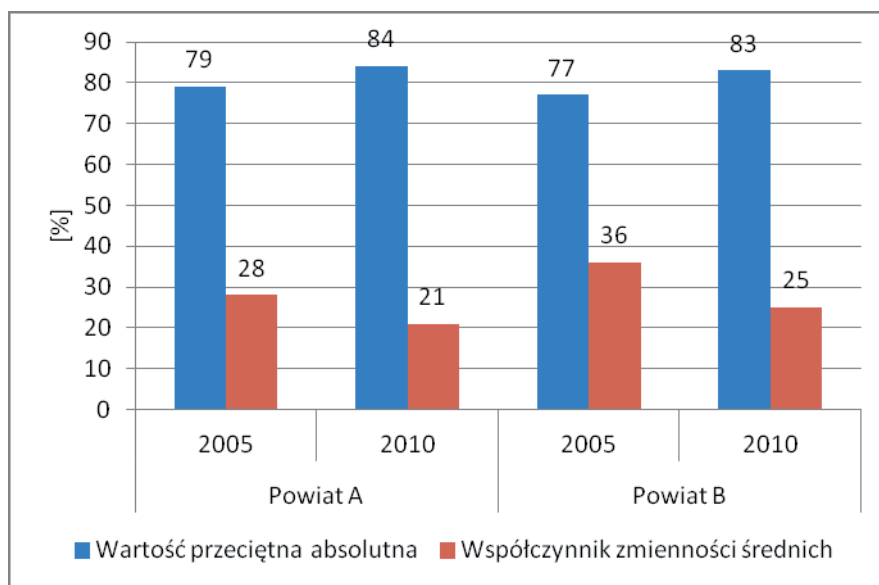
Tabela 3. Estymatory oceny ogólnej jakości danych w powiatach A i B
Table 3. Estimates of data quality overall assessment in districts A and B

Estymator	Wartości oceny ogólnej	
	Powiat A	Powiat B
Wartość przeciętna absolutna (Sa)	84	83
Współczynnik zmienności średnich (A)	0,21	0,25

Przedstawione w tabeli 3 wartości estymatorów oceny ogólnej jakości danych pokazują, że model koncepcyjny systemu geoinformacyjnego zrealizowany jest obecnie w obu przypadkach w podobnym stopniu: Sa dla powiatu A ma

wartość 84, dla powiatu B – 83. Jednakże istnieje zauważalne zróżnicowanie jakości danych w badanych powiatach. Współczynniki zmienności średnich (wyrażone w procentach) wynoszą odpowiednio 21 i 25. Z tego można wnioskować, że w powiecie A jakość danych jest o 4 punkty procentowe wyższa od danych w powiecie B.

Znając wartości estymatorów oceny jakości danych obecnych i tych z 2005 roku można przeprowadzić próbę oceny doskonalenia informacji geoprzestrzennej w okresie pięciu lat. W 2005 roku wartość przeciętna absolutna S_a dla powiatu A wynosiła 79, a dla powiatu B: 77. Współczynnik zmienności średnich miał wartość 0,28 dla powiatu A i 0,36 dla powiatu B [Ślusarski 2006]. Graficzną prezentację zmiany wartości estymatorów w roku 2005 i 2010 przedstawiono na wykresie (rys. 1).



Rysunek 1. Estymatory oceny ogólnej w roku 2005 i 2010

Figure 1. Estimates overall assessment in 2005 and 2010

Rysunek 1 pokazuje, że procesy doskonalenia jakości informacji geoprzestrzennej są widoczne. W okresie ostatnich pięciu lat wyraźnie poprawiono jakości danych w obydwu analizowanych systemach geoinformacyjnych. W każdym z powiatów po około 5 punktów procentowych wzrósł poziom zadań, które są realizowane zgodnie z obowiązującymi standardami wynikającymi z przepisów prawnych. Dokładność, aktualność i zgodność danych – w obydwu przypadkach – jest bardzo dobra ($\lambda < 0.15$). Natomiast kompletność danych nie spełnia wymaganych kryteriów.

WNIOSKI

Badania jakości danych systemów geoinformacyjnych powinny być wykonywane regularnie za pomocą odpowiednio dobranej metodyki. Przedstawione w pracy badania dotyczące jakości danych dwóch powiatowych systemów pokazują, że model koncepcyjny jest obecnie zrealizowany w obydwu przypadkach w ponad 80 procentach. W okresie ostatnich pięciu lat administracje samorządowe wykonały szereg prac bezpośrednio wpływających na podniesienie jakości danych. W obydwu powiatach opracowano m. in. numeryczne mapy ewidencyjne wykonane (w znaczącym stopniu) w oparciu o archiwalne operaty techniczne oraz pomiary terenowe.

W obydwu przypadkach należy wykonać prace mające na celu zwiększenie kompletności danych. Szczególnie brakuje opracowań dotyczących numerycznej mapy zasadniczej oraz budowy geodezyjnej ewidencji sieci uzbrojenia terenu.

BIBLIOGRAFIA

- Gaździcki J. *Leksykon geomatyczny* [online] <http://www.ptip.org.pl> [dostęp 20.12.2011]
- Longley P. i in. *GIS teoria i praktyka*, PWN Warszawa, 2006, s. 259-260.
- Nebert D. (red.) *The Spatial Data Infrastructure Cookbook* [online] <http://www.gsdi.org>. [dostęp 20.12.2011]
- Norma PN-EN-ISO 19113. *Informacja geograficzna – Postawy opisu jakości*. PKN, Warszawa, 2005.
- Redman T. *Data quality. The Field Guide*, Digital Press, Boston, 2001, ss. 233
- Ślusarski M. *Ocena porównawcza jakości danych lokalnych systemów geoinformacyjnych*. Zeszyt. Nauk. AR w Krakowie, Nr 431, Seria Geodezja Z. 22, Kraków, 2006, s. 211-216.

Dr inż. Marek Ślusarski
Katedra Geodezji
Uniwersytet Rolniczy im.H.Kołłątaja
ul. Balicka 253a
30-198 Kraków
tel.: 12 662-45-11
e-mail: rmslusar@cyf-kr.edu.pl