

SMART CITY W UJĘCIU ZORIENTOWANYM NA DANE – POLSKA W BAZIE EUROSTAT

Adam SOJDA¹, Tomasz OWCZAREK², Maciej WOLNY³

¹Wydział Organizacji i Zarządzania, Politechnika Śląska, Zabrze; adam.sojda@polsl.pl

²Wydział Organizacji i Zarządzania, Politechnika Śląska, Zabrze; tomasz.owczarek@polsl.pl

³Wydział Organizacji i Zarządzania, Politechnika Śląska, Zabrze; maciej.wolny@polsl.pl

Streszczenie: W artykule przedstawiono dostępność danych statystycznych w bazie Eurostat z zakresu Smart City odnoszących się do miast z Polski. Ocena miast w kontekście Smart City jest możliwa poprzez konstrukcję odpowiednich rankingów. Do ich konstrukcji konieczna jest znajomość zmiennych opisujących aspekty funkcjonowania miasta. Zbadano, jak kształtuje się liczba miast oraz liczba zmiennych przekazywanych z Polski do europejskiej bazy danych. Starano się odpowiedzieć, czy zawartość tej bazy danych pozwoli na budowę rankingu dla miast polskich oraz europejskich.

Słowa kluczowe: Smart City, Eurostat.

SMART CITY IN A DATA-ORIENTED APPROACH – POLAND IN THE EUROSTAT DATABASE

Abstract: The article presents the availability of statistical data in the Eurostat database on Smart City, related to cities from Poland. The assessment of cities in the context of Smart City is possible through the construction of relevant rankings. Their construction is necessary to know the variables describing the aspects of the operation of the city. The number of cities and the number of variables transferred from Poland to the European database has been examined. The question we tried to answer was whether the contents of this database will allow for the construction of a ranking for Polish and European cities.

Keywords: Smart City, Eurostat.

1. Wprowadzanie

Smart City to pojęcie wieloaspektowe, definiowane w różnoraki sposób (Albino et al., 2015). Idea Smart City skupia się w dwóch głównych nurtach. Pierwszy związany jest z wykorzystaniem ICT i nowoczesnymi technologiami dla miasta – głównie rozwiązania techniczne. Drugi nurt wskazuje na rolę kapitału ludzkiego w rozwoju inteligentnego miasta. Można wskazać dwa główne aspekty definicji. Pierwszy aspekt wskazuje, że Smart City to koncepcja rozwoju miast w określonych obszarach - najczęściej: gospodarka, ludzie, zarządzanie, mobilność, środowisko, jakość życia (Giffinger et al., 2007). Drugi dotyczy właściwości: hierarchiczność oraz wieloatrybutowość – dekompozycja pojęcia Smart City na obszary, podobszary, czynniki i wskaźniki (Bosch et al., 2016). Z jednej strony można analizować pojedyncze obszary czy aspekty, jednak dopiero podejście holistyczne kreuje Smart City (Jonek-Kowalska et al., 2018)

Pomiar i ocena Smart City za pomocą wskaźników implikuje podejście zorientowane na dane. Z jednej strony zestawy wskaźników i miar Smart City w różnych systemach definiują Smart City, z drugiej strony wskazują kierunki rozwoju miast (Giffinger et al., 2007). Można też zauważyć, że Smart City integruje i wzbogaca koncepcje rozwoju miast – „smart sustainable cities” (Ahvenniemi et al., 2017; Kramers et al., 2014).

Prognozowanie rozwoju miast w ramach pojęcia Smart City jest relatywnie proste w odniesieniu do pojedynczych wskaźników. W sytuacji, gdy znane są szeregi czasowe ich wartości, można wykorzystać całe spektrum znanych metod analizy szeregów czasowych. Główny problem leży wtedy w dużej liczbie zmiennych (wskaźników), dostępie do danych oraz jakości danych. Ważna w tym aspekcie jest również automatyzacja zbierania i przetwarzania danych, która wymaga opracowania i wykorzystania odpowiednich skryptów oraz narzędzi informatycznych. Jednym z kluczowych zagadnień Smart City jest problem dostępności danych (ISO, 2014): źródła danych, open-data, jakość danych, problem braków danych, automatyzacja kolekcji danych i monitoring zmian.

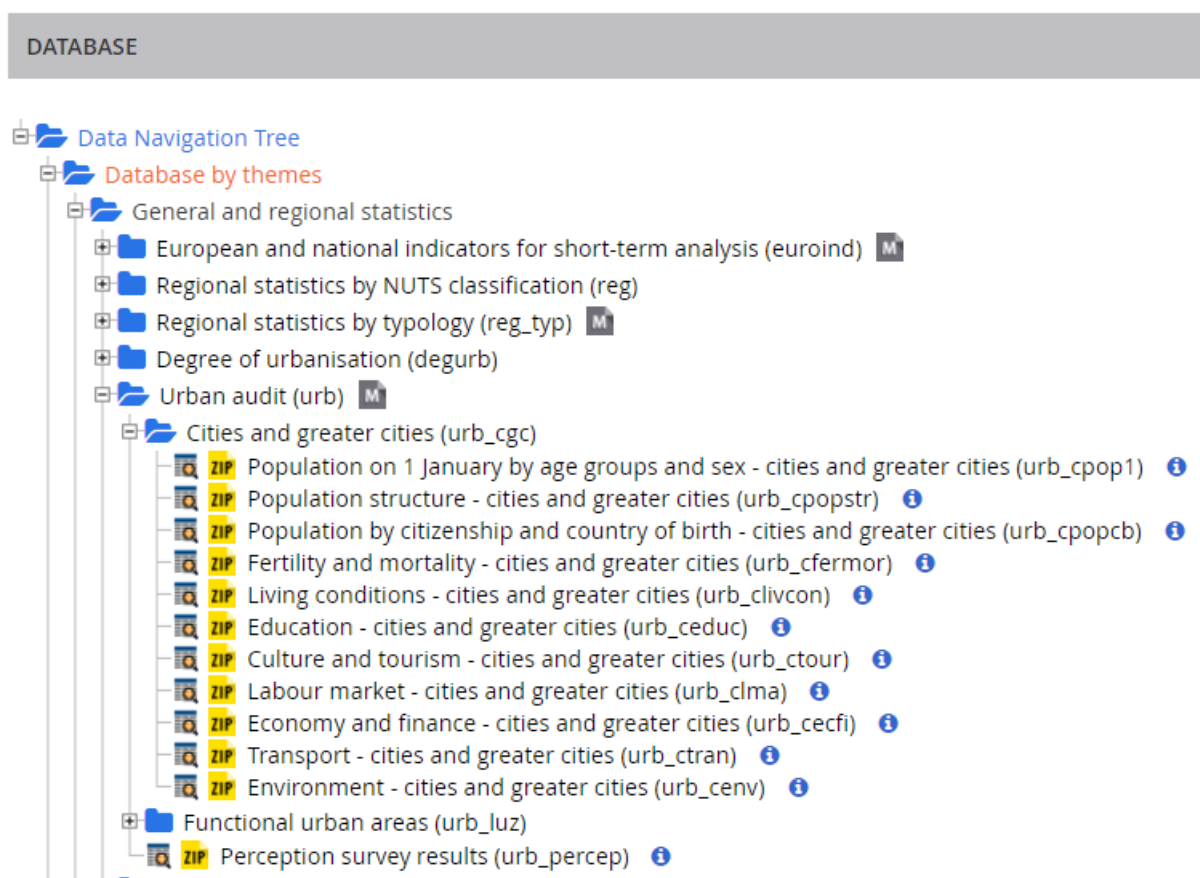
Jedną z ogólnodostępnych baz danych jest baza Eurostat, w której gromadzone są dane dotyczące krajów europejskich, w tym również w zakresie danych społeczno-gospodarczych regionów miejskich w różnych zakresach – Urban Audit (Eurostat, 2018). Dane te mogą być wykorzystane do analizy na potrzeby oceny Smart City (Szczech-Pietkiewicz, 2015).

Celem pracy zatem jest z jednej strony prezentacja narzędzia i skryptów, które umożliwiają automatyczne pobranie oraz aktualizację danych z bazy Eurostat. Z drugiej strony artykuł ma na celu przedstawienie problemów w dostępności danych.

2. Dostęp do danych Eurostatu

Eurostat (Europejski Urząd Statystyczny) jest urzędem Komisji Europejskiej, który zajmuje się sporządzaniem analiz i prognoz dotyczących Europy¹, bazując na danych statystycznych. Urząd udostępnia również bazy danych.

W szerokim spektrum zainteresowań znajduje się obszar poświęcony miastom i obszarom miejskim (Urban Audit). Na rysunku 1 pokazano drzewo dostępu do tych danych przez stronę www.



Rysunek 1. Zrzut ekranu strony internetowej Eurostatu. Eurostat (2018). <https://ec.europa.eu/eurostat/data/database>

Dostęp do danych jest możliwy z poziomu serwisu www Eurostatu (Eurostat, 2018) i umożliwia przegląd danych online lub pobranie zbiorów w jednym z dostępnych formatów: Excel, SPSS, CSV, PC-Axis, TSV, PDF oraz HTML. Możliwa jest również automatyzacja procesu pobierania i analizy danych bezpośrednio z serwera baz Eurostatu za pomocą programu R. Umożliwia to pakiet „eurostat” (Lahti et al., 2017). Na rysunku 2 przedstawiono fragment skryptu R wykorzystany do pozyskania danych z Urban Audit.

¹ <https://ec.europa.eu/eurostat/about/overview>

```
db_name <- c('urb_cpop1', 'urb_cpopstr', 'urb_cpopcb', 'urb_cfermor', 'urb_clivcon',
            'urb_ceduc', 'urb_ctour', 'urb_clma', 'urb_cecfi', 'urb_ctrans', 'urb_cenv')

my.base <- function(x) { db_f <- get_eurostat(x, time_format = "raw")
  db_f$base <- c(x)
  db_f }

db_all <- my.base(db_name[1])

for (ii in 2:length(db_name)){
  db_all <- rbind(db_all, my.base(db_name[ii]))}

db_all$name <- label_eurostat(db_all$cities, dic = "cities", lang = "en", fix_duplicated = T)
db_all$idcountry <- substr(db_all$cities, 1, 2)
db_all$c <- label_eurostat(db_all$idcountry, dic = "cities", lang = "en", fix_duplicated = T)
db_all$v <- label_eurostat(db_all$indic_ur, dic = "indic_ur", lang = "en", fix_duplicated = T)
```

Rysunek 2. Fragment kodu skryptu języka R służący do pobierania danych z wykorzystaniem pakietu „eurostat”.

Zadaniem skryptu jest pobranie danych z odpowiednich tabel („urb_cpop1”, ... , „urb_cenv”), połączeniu ich w jedną ramkę danych, dodawaniu etykiet dotyczących miast, krajów oraz zmiennych. Tak przygotowana ramka danych stanowi dane wejściowe do analiz.

3. Wyniki

Przedmiotem badania była dostępność danych dotyczących polskich miast w bazie Eurostat. Na początku zbadano jak kształtuje się liczba miast z Polski na tle pozostałych krajów europejskich. Zakres badania obejmował lata 2010-2017 i dotyczył wszystkich krajów, które w tym okresie znalazły się w bazie.

Analiza liczby miast pozwoliła na ustalenie trzech grup krajów:

- o co najmniej 50 miastach w bazie danych: Wielka Brytania, Hiszpania, Niemcy, Francja, Włochy, Polska,
- zawierających od 10 do 50 miast: Holandia, Rumunia, Portugalia, Węgry, Bułgaria, Czechy, Szwajcaria, Szwecja, Belgia, Finlandia, Grecja,
- zawierających poniżej 10 miast: Słowacja, Chorwacja, Austria, Irlandia, Litwa, Norwegia, Dania, Łotwa, Estonia, Cypr, Słowenia, Islandia, Luksemburg, Malta.

W dalszej części przedstawiono wyniki analizy odnośnie liczby miast z pierwszej grupy krajów w rozbiciu na lata 2010-2017.

Najwięcej miast jest z Wielkiej Brytanii i jest ich średnio 2,5 razy więcej niż miast z Polski. Druga w rankingu Hiszpania ma dwa razy więcej miast niż Polska. Znajdujący się na trzecim miejscu Niemcy mają w bazie prawie dwa razy tyle miast niż Polska. Francja ma 1,7 razy więcej niż Polska, a zajmujący miejsce przed Polską Włochy tylko o 36% więcej miast.

Tabela 1.*Liczba miast w bazie dla wybranych krajów*

	2010	2011	2012	2013	2014	2015	2016	2017
WIELKA BRYTANIA	172	172	172	172	172	188	189	189
HISZPANIA	111	111	111	111	134	143	143	-
NIEMCY	125	125	125	125	125	127	127	119
FRANCJA	115	116	116	116	116	116	116	115
WŁOCHY	94	94	94	94	94	94	94	-
POLSKA	69	69	69	69	69	-	-	-

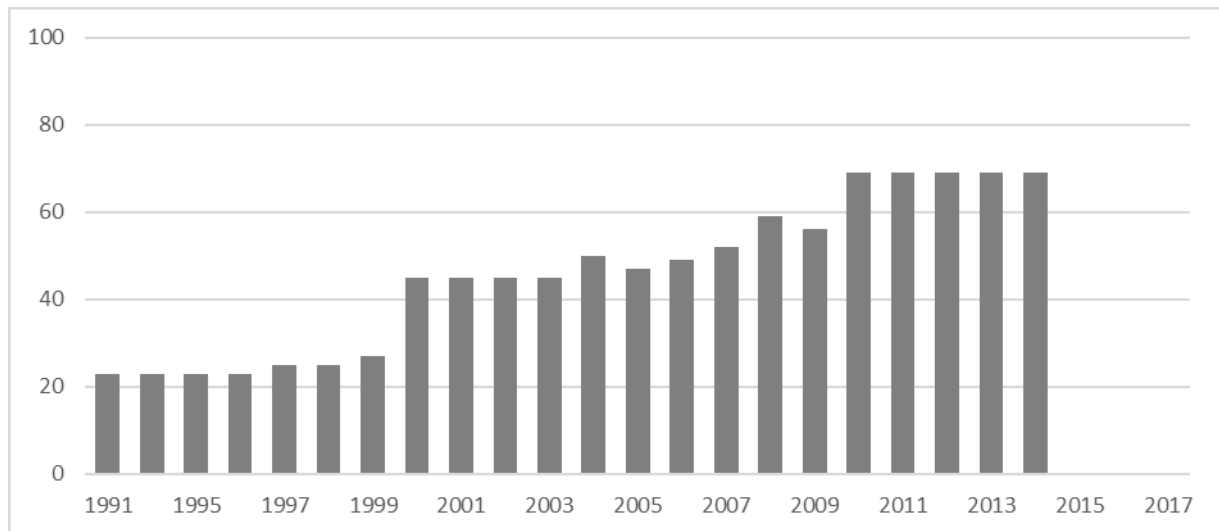
Ocena dostępności danych nie może odbywać się tylko na podstawie liczby zgłoszonych miast. Ważna jest liczba wskaźników, dla których pokazano dane z poszczególnych miast. Jak się okazuje liczba wskaźników nie jest jednakowa, i może dochodzić do 183 wskaźników. W tabeli poniżej zaprezentowano zakres zmienności liczby wskaźników (minimalna – maksymalna liczba wskaźników) dla miast w wybranych krajach w poszczególnych latach.

Tabela 2.*Zakres zmienności ilości wskaźników dla miast w wybranych krajach*

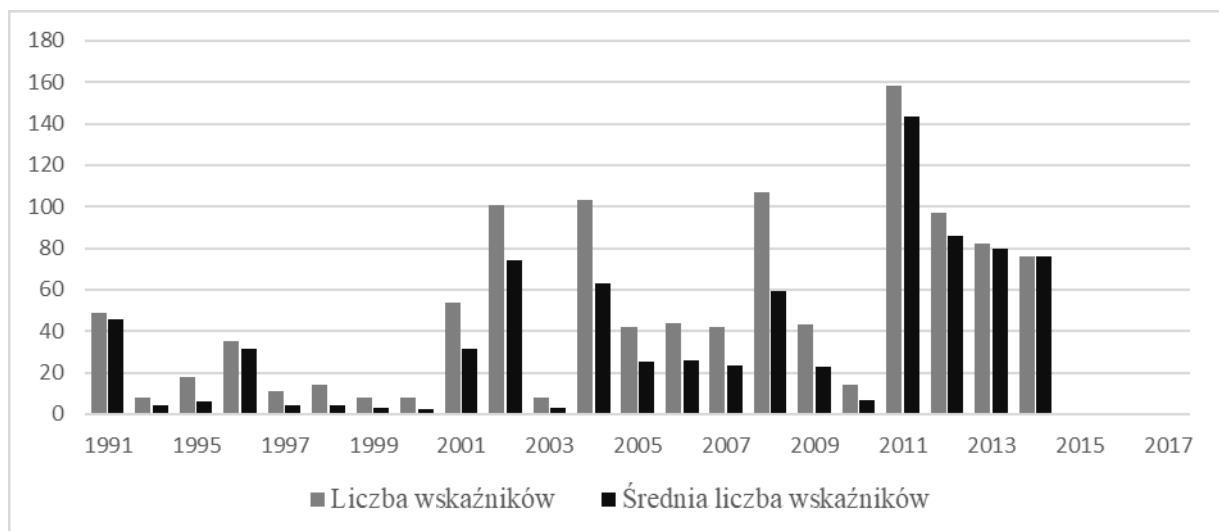
	2010	2011	2012	2013	2014	2015	2016	2017
WIELKA BRYTANIA	63-89	128-177	67-116	65-113	67-109	7-155	66-144	27-56
HISZPANIA	85-119	128-174	85-115	121-153	8-154	95-145	122-149	---
NIEMCY	167-197	168-197	173-200	131-162	126-154	2-140	6-122	1-3
FRANCJA	1-153	1-182	1-184	1-180	117-174	11-88	10-16	1-2
WŁOCHY	18-95	40-156	84-107	84-98	75-93	72-95	71-78	---
POLSKA	1-12	106-157	68-97	69-82	69-76	---	---	---

Jak można zauważyć Polska w bazie Eurostat w porównaniu z kluczowymi krajami Unii Europejskiej jest reprezentowana w stopniu odpowiadającym jej miejscu w tym gremium. Pewne obawy związane są z ostatnimi latami, gdzie dane przestały być przekazywane.

Informacje przedstawione w tabeli 2 wskazują, że nie tylko liczba krajów jest ważna, ale również liczba wskaźników oraz ich wartość przeciętna charakteryzująca miasta z danego państwa. W przypadku Polski zbadano również i te aspekty.



Rysunek 3. Liczba miast z Polski w bazie Eurostat.



Rysunek 4. Liczba wskaźników opisujących polskie miasta wraz ze średnią liczbą wskaźników przypadającą na pojedyncze miasto.

Dane przechowywane są w osobnych zbiorach, każdy z nich odnosi się do innego aspektu funkcjonowania miasta. W rozbiciu na bazy wykonano analizę pokazującą liczbę zmiennych, liczbę miast jak i również procent pokrycia zbioru danymi. Poniżej przedstawiono, jak kształtuje się latach 2010 – 2014 liczba zmiennych w poszczególnych bazach danych oraz pokrycie tymi danymi bazy.

Tabela 3.

Udział ilościowy miast oraz wskaźników odnoszących się do polskich miast w poszczególnych bazach.

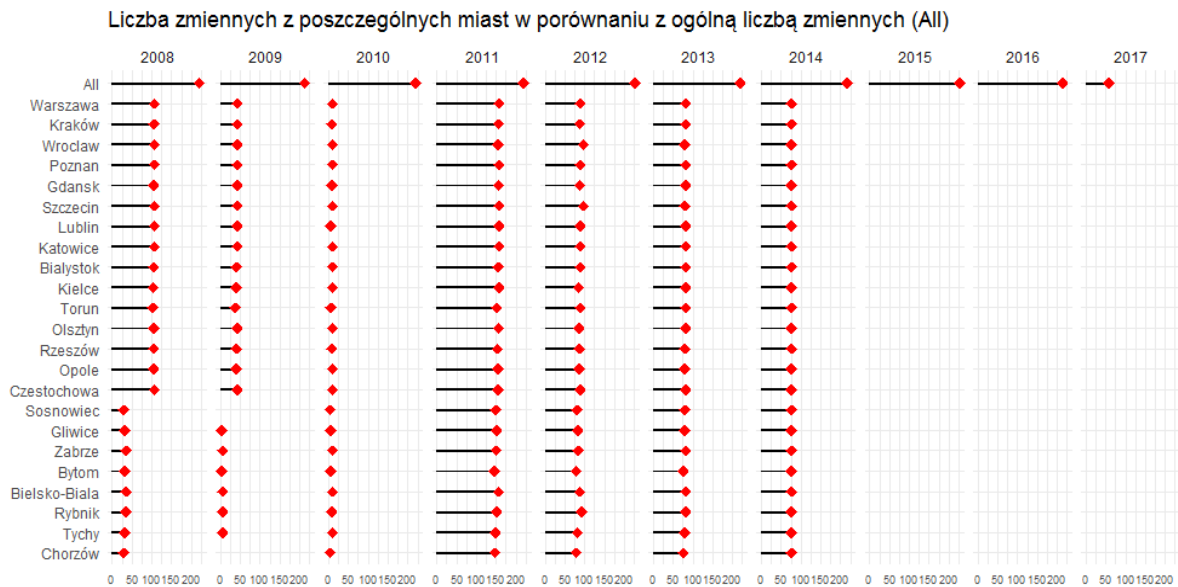
	Liczba miast z Polski					Liczba wskaźników z polskich miast				
	2010	2011	2012	2013	2014	2010	2011	2012	2013	2014
urb_cecfi	--	69	--	--	--	--	1	--	--	--
urb_ceduc	--	69	69	69	69	--	11	4	4	4
urb_cenv	69	69	69	69	69	11	13	19	7	1
urb_cfermor	--	69	69	69	69	--	12	8	8	8
urb_clivcon	68	68	68	68	68	3	20	3	1	1
urb_clma	--	69	69	69	69	--	16	3	3	3
urb_cpop1	--	69	69	69	69	--	36	36	36	36
urb_cpopcb	--	69	--	--	--	--	14	--	--	--
urb_cpopstr	--	69	69	69	69	--	17	17	17	17
urb_ctour	--	68	68	68	68	--	11	4	4	4
urb_ctrans	--	69	69	69	69	--	7	3	2	2

Tabela 4.

Liczba wskaźników dostępna w poszczególnych bazach danych wraz z procentowym pokryciem ich przez polskie wskaźniki.

	Liczba wskaźników w bazie ogółem					% pokrycia bazy polskimi wskaźnikami				
	2010	2011	2012	2013	2014	2010	2011	2012	2013	2014
urb_cecfi	2	2	2	2	2	--	50%	--	--	--
urb_ceduc	12	14	12	11	12	--	79%	33%	36%	33%
urb_cenv	16	20	28	16	22	69%	65%	68%	44%	5%
urb_cfermor	12	12	12	12	12	--	100%	67%	67%	67%
urb_clivcon	37	37	37	37	36	8%	54%	8%	3%	3%
urb_clma	38	38	38	38	38	--	42%	8%	8%	8%
urb_cpop1	36	36	36	36	36	--	100%	100%	100%	100%
urb_cpopcb	14	14	14	14	14	--	100%	--	--	--
urb_cpopstr	18	18	18	18	18	--	94%	94%	94%	94%
urb_ctour	11	11	11	11	11	--	100%	36%	36%	36%
urb_ctrans	11	17	17	17	17	--	41%	18%	12%	12%

Zbadano również jak kształtuje się relacja liczby zmiennych z poszczególnych miast w stosunku do wszystkich zmiennych znajdujących się w bazie. Wyniki analizy przedstawiono na rysunku 5.



Rysunek 5. Liczba zmiennych z wybranych miast polskich na tle ogólnej dostępnej liczby zmiennych.

4. Interpretacja wyników

Analiza danych pokazała, że wielkość kraju wpływa znacząco na liczbę miast, jakie kraj zgłasza do bazy danych. Liczba miast reprezentowanych w bazie odpowiada miejscu Polski na mapie Europy (tabela 1). Liczba wskaźników, które opisują polskie miasta też wyznacza Polsce podobne miejsce (tabela 2). Jednakże, nie chodzi tu tylko o przekazanie informacji do bazy proporcjonalnie do wielkości kraju. Tylko liczba miast może być w ten sposób uzasadniona. Natomiast przekazywanie informacji o miastach powinno odbywać się tych samych lub zbliżonych zasadach. Niezależnie od miasta informacja powinna dotyczyć pełnej liczby wskaźników. Nawet jeśli miasto nie posiada bazy hotelowej, to powinno być to odnotowane w bazie nie na zasadzie braku danych, ale konkretnej wartości 0.

Pod względem liczby miast, jakie zostały zgłoszone z Polski do bazy, można zauważyć, że od roku 1991 liczba ta stale rosła, aż do roku 2010 (rysunek 3). W roku 2010 i w latach późniejszych liczba ta zatrzymała się na 69 miastach i wydaje się adekwatna do wielkości i charakteru miast w Polsce. Należy wyrazić w tym miejscu zdziwienie, dlaczego dane urywają się po roku 2014. Od 2015 roku nie zostały opublikowane żadne dane dla Polski w odniesieniu to tego obszaru badawczego. O ile liczba miast w bazie rosła, by ustabilizować się na poziomie 69, to zupełnie inaczej zachowywała się liczba wskaźników (rysunek 4) Liczba ta od 48 w 1991 spadała do minimalnie 8 i maksymalnie 158 w roku 2011. Następnie liczba ta spadła do poziomu 76 zmiennych w 2014 roku. Można zauważyć, w latach 2012 – 2014 liczba wskaźników była stała i wszystkie miasta przekazywały informacje w tych obszarach.

Kiedy jednak porównany dane z Polski w stosunku do wszystkich zmiennych w bazie danych (tabela 3 i tabela 4), to sytuacja nie jest tak zadawalająca. Tylko w bazie dotyczącej populacji miast przekazujemy wartości odnośnie do wszystkich zmiennych. Wysoki współczynnik pokrycia znajduje się również w bazie odnoszącej się do struktury populacji. W pozostałych bazach sytuacja jest gorsza. Na poziomie pokrycia 67% jest baza (urb_cfermor) odnosząca się do urodzeń jak i śmiertelności w miastach. Informacje odnośnie środowiska (urb_cenv) stale spada od 69% do 5% w 2014. Na poziomie 36 % są bazy odnoszące się do edukacji (urb_ceduc) oraz turystyki (urb_ctour). W pozostałych bazach poziom pokrycia nie przekracza 12%. Można śmiało stwierdzić, że na tle wszystkich dostępnych zmiennych dane z Polski nie prezentują się już tak okazale.

Sprawdzono również, jak kształtuje się liczba wskaźników z poszczególnych miast (rysunek 5). Analiza potwierdziła, że liczba przekazywanych zmiennych z dużych (wojewódzkich) miast jest porównywalna z liczbą zmiennych miast mniejszych.

5. Podsumowanie

Problem dostępności danych (pokrycia) jest charakterystyczny i nie dotyczy wyłącznie danych Eurostatu. Podobne zjawisko można obserwować dla danych związanych z normą ISO37120 zarówno w zakresie danych dodatkowych, jak i podstawowych².

Problem z dostępnością danych można traktować w ujęciu regionalnym – wtedy dostarczone dane są spójne, oraz dostarczona jest podobna liczba wskaźników. W ujęciu globalnym – dostarczane wskaźniki przez Polskę są niepełne i nie pozwalają na porównanie jak i konstrukcje europejskich rankingów.

Pewną propozycją rozwiązania problemu braku dostępności danych może być analiza dla różnych poziomów pokrycia danymi. Przyjmując odpowiednią strategię odnośnie do braków danych, można uzupełniać te braki dla miast i uwzględniać miasta w analizach na wyższym stopniu pokrycia danymi.

Podejście pesymistyczne wskazuje na uzupełnienie braków wartościami najgorszymi dla danego wskaźnika. Wtedy z punktu widzenia budowy rankingu brak danych będzie mniej preferowany, o ile rzeczywista wartość wskaźnika nie będzie najmniejsza w stosunku do wszystkich pozostałych. W celu ograniczenia występowania takich sytuacji można uwzględniać w systemie kary za brak danych.

²World Council of City Data (<http://www.dataforcities.org>) certyfikuje od 30-40 wskaźników podstawowych – poziom aspirujący, przy tym dla tego poziomu mogą to być różne wskaźniki. Zatem miasta dla poziomu aspirującego mogą gromadzić dane z różniących się zakresów.

Problem braku dostępności danych może być również rozwiązany na drodze urzędowej poprzez wydanie odpowiednich rozporządzeń obligujących państwa do dostarczenia danych odnośnie swoich miast.

Inny obszar badawczy dostarczający pośrednio rozwiązania przedstawionego problemu dotyczy badania współzależności, a precyzyjniej, współmienności cech definiujących Smart City. Potencjalne grupy wskaźników w szerokim spektrum oceny Smart City wykazujące współmienność można zastąpić pojedynczym wskaźnikiem z grupy. Wskaźnik ten powinien charakteryzować się największym pokryciem danych w grupie badanych miast. Można również poszukiwać podobieństw między miastami, badając na przykład geograficzne podobieństwa rozkładów wartości wskaźników oceny Smart City.

Bibliografia

1. Giffinger, R., Fertner, C., Kramar, H., Kramar, H., Kalasek, R., Pichler-Milanovic, N., Meijers, E. (2007). *Smart Cities. Ranking of European medium-sized cities*. Centre for Regional Science, Vienna University of Technology, http://www.smart-cities.eu/download/smart_cities_final_report.pdf (31.08.2018).
2. Ahvenniemi, H., Huovila, A., Pinto-Seppä, I., Airaksinen, M. (2017). What are the differences between sustainable and smart cities?. *Cities*, 60, pp. 234-245.
3. Kramers, A., Höjer, M., Lövehagen, N., Wang, J. (2014), Smart sustainable cities – Exploring ICT solutions for reduced energy use in cities. *Environmental Modelling & Software*, 56, 52-62. <http://dx.doi.org/10.1016/j.envsoft.2013.12.019>
4. ISO (International Standardization Organization) (2014). *ISO 37120:2014 Sustainable development of communities — Indicators for city services and quality of life*.
5. Jonek-Kowalska, I., Kaźmierczak, J., Kramarz, M., Hilarowicz, A., Wolny, M. (2018). *Introduction to the research project “Smart City: a holistic approach”*, SGEM.
6. Bosch, P., Jongeneel, S., Neumann H.M., Branislav I., Huovila, A. (2016). *Deliverable 3.3 Recommendations for a Smart City index*, CITYkeys project, <http://nws.euocities.eu/MediaShell/media/CITYkeys%20D3.3%20-%20Recommendations%20for%20the%20Smart%20City%20Index.pdf> (31.08.2018).
7. Albino, V., Berardi, U., Dangelico, R. (2015). Smart cities: Definitions, dimensions, performance and initiatives. *Journal of Urban Technology*, 22(1), pp.3-21, <http://dx.doi.org/10.1080/10630732.2014.942092>
8. Szczech-Pietkiewicz, E. (2015). Smart City – próba definicji i pomiaru. *Prace naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Gospodarka lokalna w teorii i praktyce*, 391, doi: 10.15611/PN.2015.391.07
9. Eurostat. (2018). <https://ec.europa.eu/eurostat/data/database> (31.08.2018)
10. Lahti, L., Huovari, J., Kainu, M., Biecek, P. (2017). Retrieval and analysis of Eurostat open data with the eurostat package. *The R Journal*, 9(1), 385-392.