

Założenia projektowe i obszary zastosowań systemu GIS Kultura

Raport kończący projekt budowy
systemu geolokalizacji
infrastruktury żywej kultury

DR MARIUSZ PIOTROWSKI

24 kwietnia 2017

Redakcja i korekta:
prof. dr hab. Barbara Fatyga
dr Bogna Kietlińska

Współpraca:
mgr Piotr Michalski

skład komputerowy: Mariusz Piotrowski

Raport opublikowany na licencji Creative Common BY & SA.

Pierwsza wersja: grudzień 2016

Spis treści

1	Przedmowa	1
2	Założenia systemu GIS Kultura	3
3	Badanie dostępności	8
4	Dane wejściowe	15
5	Charakterystyka źródeł danych	28
6	Funkcjonalność systemu.	41
6.1	Analizy implicite przestrzenne.	41
6.1.1	Wydatki na kulturę ze względu na typ historyczny. Zestawienia.	41
6.1.2	Wydatki na kulturę. Wizualizacja danych.	46
6.2	Analizy explicite przestrzenne.	55
6.2.1	Obiekty wokół punktu. Jeszcze raz o wydatkach gmin.	55
6.2.2	Obiekty przy drodze.	59
6.2.3	Obiekty przy trasie komunikacji publicznej.	64

<i>SPIS TREŚCI</i>	iii
6.2.4 Zasięg dostępności obiektu. Metoda izochronu.	67
7 Zakończenie	71

Spis rysunków

4.1	Schemat procesu projektowego GIS Kultura	15
4.2	Okno crawlera.	17
4.3	Źródła danych w projekcie GIS Kultura	19
4.4	Schemat przetwarzania danych z projektu openstreetmap. . .	21
4.5	Gminy wg typu historycznego (stan na 1.01.2016).	25
5.1	Obiekty żywej kultury wg funkcji regulującej (n=507 691). . .	29
5.2	Obiekty wąsko rozumianej kultury w bazie GIS Kultura (n=507 691).	30
5.3	Indeksy obiektów kultury wg GUS i obiekty żywej kultury wg indeksów upowszechniania kultury i kultury „wysokiej” w województwach.	32
5.4	Obiekty żywej kultury wg indeksów (n=834 883).	33
5.5	Obiekty z KRS, których funkcji nie określono ze względu na typ prawny (n=151 040).	35
5.6	Obiekty z OSM, których nie zostały przypisane do indeksów obiektów żywej kultury ze względu na charakterystykę tagów (n=176 152).	39

6.1	Wydatki na kulturę ze względu na typ historyczny w gminach.	45
6.2	Użycie metody kwantyli do zaprezentowania wydatków gmin na kulturę, przy użyciu oprogramowania QGIS.	47
6.3	Nałożenie zakresu granic podziałów historycznych na warstwy z mapą rozkładów wydatków.	48
6.4	Gminy, które wykazały dodatnie saldo w pozycji 921, metoda klasyfikacji - kwantyle.	50
6.5	Migracje ludności między gminami w Polsce pomiędzy rokiem 2005 a 2015.Przedziały tworzone metodą równych podziałów.	51
6.6	Migracje ludności między gminami w Polsce pomiędzy rokiem 2005 a 2015.Przedziały tworzone metodą kwantylową.	51
6.7	Gminy wg typu metropolitalnego.	53
6.8	Wydatki na kulturę na mieszkańca gmin w okolicach Wrocławia.	57
6.9	Wydatki na kulturę w gminach w okolicach Wrocławia i Warszawy.	58
6.10	Obiekty w odległości 50 m od najszybszej trasy Warszawa-Gdańsk (N=544).	60
6.11	Obiekty żywej kultury na trasie Warszawa-Gdańsk (n=544).	62
6.12	Obiekty na trasie Warszawa-Gdańsk, bez przypisanego indeksu żywej kultury (n=107).	63
6.13	Obiekty żywej kultury na trasie linii autobusowej 116, wg typów indeksów. Legenda sortowana alfabetycznie.	66
6.14	Obiekty żywej kultury w dostępności 12 minut od Domu Kultury (Sierpiec).	69
6.15	Informacje o obiektach żywej kultury w dostępności 12 minut od Domu Kultury.	70

Spis tabel

3.1	Lista indeksów obiektów żywej kultury.	11
5.1	Typ Obiektu Żywej Kultury ze względu na typ obiektu KRS - 10 najbardziej licznych.	36
5.2	Typ Obiektu Żywej Kultury ze względu na typ obiektu KRS wg. województw - 10 z największą liczebnością.	37
5.3	Częstości występowania obiektów z bazy OSM ze względu na typ indeksu obiektu żywej kultury – 10 najczęściej występu- jących.	40
6.1	Suma wydatków na kulturę ze względu na typ historyczny (dane na rok 2015).	42
6.2	Wydatki na kulturę ze względu na typ historyczny (dane na rok 2015), bez M. st. Warszawa.	42
6.3	Wydatki na kulturę ze względu na typ historyczny (dane na rok 2015), wg typów gmin.	43
6.4	Wydatki na kulturę na mieszkańca ze względu na typ histo- ryczny (dane na rok 2015).	43
6.5	Gminy, które wykazały dodatnie saldo w pozycji 921, 5 gmin z najwyższymi kwotami.	49

6.6	Wydatki sumaryczne gmin z typu A1 najbliższego ośrodki metropolitalnemu i B2 znajdującego się w najdalszej strefie wpływu.	54
6.7	Przykładowe 9 obiektów znajdujących się w odległości 50 m od najszybszej trasy Warszawa-Gdańsk.	61
6.8	Linie autobusowe, przy trasie których znajduje się największa liczba obiektów żywej kultury	65
6.9	Przykładowe obiekty żywej kultury przy trasie linii autobusowej 116, z uwzględnieniem dostępu dla osób poruszających się na wózkach inwalidzkich.	67

Rozdział 1

Przedmowa

Struktura raportu

Raport ma charakter prezentacji metodologii pracy z danymi przestrzennymi o kulturze. Prezentacja wyników pracy zostanie poprzedzona rozdziałami opisującymi założenia teorii żywej kultury, w kontekście badania dostępności, oraz pogłębioną charakterystyką użytych danych.

Strona internetowa

Strona ozkultura.pl¹ stanowi serce prezentowanego systemu. Szczególnie podczas lektury raportu warto zapoznać się z niektórymi podstronami:

- Sekcja Słownik - zawiera definicje pojęć używanych w projekcie.
- Sekcja GIS Kultura - oferuje dostęp do geoportalu.
- Sekcja Teksty OZK-SB - prezentuje teksty metodologiczne dotyczące badań kultury.

¹<http://ozkultura.pl>

Narzędzia

Projekt GIS Kultura bazuje na oprogramowaniu o otwartych źródłach. Narzędzia udostępnione na jednej z otwartych licencji zostały użyte na każdym z etapów prac, począwszy od przetwarzania i przechowywania danych, przez analizy, a kończąc na tym raporcie:

- Dane projektu są przechowywane w bazie Postgresql², rozszerzonej o dodatki: Postgis³ i Pgrouting⁴.
- Wizualizacja danych przestrzennych została przeprowadzona w programie QGIS.⁵
- Pozostałe wizualizacje zostały wykonane przy użyciu pakietu ggplot2 programu R.⁶
- Skład tekstu został wykonany w systemie L^AT_EX.

²<https://www.postgresql.org/>

³<http://postgis.net/>

⁴<http://pgrouting.org/>

⁵<http://www.qgis.org/pl/site/>

⁶<https://cran.r-project.org/>

Rozdział 2

Założenia systemu GIS Kultura

W ramach trzyletniego projektu Fundacji Obserwatorium Żywej Kultury - Sieć Badawcza realizowanego wraz z Wydziałem Geodezji i Kartografii Politechniki Warszawskiej, zbudowany został system analizy i wizualizacji danych „GIS Kultura”. System ten powstał przy wsparciu Ministerstwa Kultury i Dziedzictwa Narodowego, w ramach programu „Obserwatorium kultury”. Niniejszy raport stanowi podsumowanie podjętych działań oraz prezentuje system zbierania danych i analiz stanu kultury w Polsce.

Obecnie istniejące systemy wizualizacji danych organizują informacje dotyczące różnych poziomów administracyjnych. Podstawowe źródło danych o kulturze - Bank Danych Lokalnych GUS - pozwala na zapoznanie się z danymi do poziomu gminy¹. Częściej jednak dane są agregowane i udostępniane tylko do poziomu powiatu, a nierzadko jedynie do poziomu województwa. Z potrzeby ujednolicania statystyk europejskich dane GUS są także organizowane w jednostkach statystycznych, takich jak regiony (NTS-1) i podregiony (NTS-3). Zarazem trzeba tu zaznaczyć, że sprawozdawczość kulturalna na poziomie europejskim jest szczątkowa. Mimo wszystkich wad i braków w sposobie zbierania danych o kulturze, statystyki GUS dotyczą-

¹<https://bdl.stat.gov.pl/BDL/start>

ce kultury pozostają jedynym źródłem danych dla całego kraju. Dane te są zbierane z formularzy sprawozdawczości kulturalnej, oznaczonych numerami K-01 do K-10. Problematyka samego sposobu gromadzenia danych została obszernie opisana w raporcie podsumowującym projekt „Poprawa jakości gromadzenia danych o publicznych i niepublicznych instytucjach kultury w Polsce”.²

Jak pokazały analizy programu MKiDN „Obserwatorium Kultury”, istnieje niezwykle silna potrzeba zdobywania informacji o świecie kultury.³ Jednak zagadnienie krajobrazu rzeczywistości społecznej kształtowanego przez granice administracyjne, jest drugorzędne wobec problematyki zwykle podejmowanej w badaniach kultury, skupionych przede wszystkim na uczestnictwie. Z tego punktu widzenia takie tematy, jak analizy polityk kulturalnych, w których diagnozy dotyczą m.in. finansów jednostek samorządu terytorialnego, są jednym z pierwszych kroków do określenia stanu kultury w mieście A czy wsi B. W różnych proponowanych metodologiach badania kultury wyróżnić można takie obszary zagadnień jak: analiza praktyk kulturalnych, w tym praktyk kadr kultury i ich kompetencji, badania nad uczestnikami kultury, badania wydarzeń kulturalnych czy wreszcie infrastruktury kultury. I właśnie ten ostatni element - czyli ukazanie materialnych (infrastrukturalnych) warunków aktywności kulturalnej - stanowiło podstawowy cel budowy systemu gromadzenia i analizy danych o kulturze.

Wizualizacja rozmieszczenia instytucji kultury i sposoby prezentowania takich danych w badaniach są możliwe np. dzięki serwisowi mojapolska.pl, a od 2015 roku w ramach geoportalu statystycznego GUS. Niemniej jednak, dane tam prezentowane są agregowane do poziomu gmin, co znacznie ogranicza możliwości wykonywania analiz relacji przestrzennych. Najczęściej bowiem dane takie są już w określony sposób zinterpretowane, chociażby przez użycie jednego, arbitralnie określonego, sposobu ich klasyfikacji. Są

²Nowa sprawozdawczość instytucji kultury:http://badania-w-kulturze.mik.krakow.pl/files/nowa_sprawozdawczosc_BWK_raport_2015_9.pdf

³Przegląd najciekawszych raportów o kulturze Krzysztofa Olechnickiego i Tomasza Szlendaka został opublikowanych w 100 numerze NN6T, jest to wycinek Raportu o raportach. Wielowymiarowej i wielofunkcyjnej oceny trafności, recepcji i użyteczności raportów o stanie kultury. Pełna wersja, jednakże nie doczekała się publikacji w internecie, por.: http://www.beczmiara.pl/1158,raport_o_raportach_o_kulturze.html

to więc narzędzia prezentujące pewną koncepcję teoretyczną, która zarazem (na ogół) nie jest wyartykułowana wprost. Takie systemy umożliwiają wykonywanie poprawnych analiz ich bardziej zaawansowanym użytkownikom, dostrzegającym ten problem, a mniej zaawansowanych narażając na rozmaite pochopne generalizacje. Zaprojektowane interfejsy dostępne (geoportal.gus, mojanapolis, także dostęp do danych bdl poprzez serwis mojepanstwo.pl) stanowią zamknięte środowiska organizujące dane o różnym stopniu *przyjrzalności* dla użytkownika w sposób względnie trwałe. Natomiast - szczególnie w przypadku kultury, kiedy badane zjawiska mają charakter bardzo dynamiczny - metody analiz, organizacji danych i przeorganizowania tych informacji powinny być równie dynamiczne. Jest to możliwe wyłącznie dzięki możliwości operowania na danych źródłowych, pierwotnych. Budowa interfejsów koresponduje z budową teorii, która w tym przypadku, jest procesem *kroczącym za badaczem* i zwykle nie jest na tyle gotowa by mogła być operacjonalizowana jako interfejs właśnie. Opisany stan rzeczy można uznać za dowód eksploracyjnego charakteru tego rodzaju działań.

Przestrzeń, obok czasu, stanowi podstawowy wymiar analiz kultury. W badaniach nad polityką kulturalną podstawowym pytaniem jest: gdzie - w jakich miejscach, na jakich terytoriach, w jakich środowiskach (*milieux*) - konkretna polityka jest realizowana. Przyjmujemy, że przestrzeń jest jedną ze zmiennych wyjaśniających, która określa różnice w kształcie tych polityk. Fundacja Obserwatorium Żywej Kultury - Sieć Badawcza od początku swojego działania zwracała uwagę na te różnice, uwzględniając w projektach i analizach wymiar przestrzenny. Wypracowanie na podstawie teorii żywej kultury ⁴ własnego podejścia badawczego, skutkuje zainteresowaniem szerokim środowiskiem życia człowieka, grup społecznych i instytucji. Środowisko to, z jednej strony, kształtowane jest przez dynamiczne procesy; z drugiej zaś, samo te procesy kształtuje. Przestrzeni materialnej nadawany jest kształt przez praktyki kulturowe, które się w niej *materializują*. Jeśli są one względnie trwałe, wówczas mogą manifestować się w formie konkretnej infrastruktury; gdy ich natura jest efemeryczna, śladem może pozostać tylko wydarzenie zapisane w pamięci uczestników. Dlatego też jednym z pierw-

⁴Barbara Fatyga, Żywa kultura, definicja autorska dla OŻK, (w:) Wielozródłowy Słownik Kultury OŻK, 2013, <http://ozkultura.pl/wpisy/86>

szych założeń budowy systemu analitycznego „GIS Kultura” było skupienie się na względnie trwałych praktykach kulturowych, które skutkowały pojawieniem się materialnych wytworów w przestrzeni fizycznej. Z tego względu założyliśmy, iż **umieszczenie tych wytworów (obiektów) na mapie, pozwoli odpowiedzieć na pytania: gdzie materializują się praktyki kulturowe oraz w jaki sposób prezentują się przestrzenne wzory kultury.**

Oprócz pytania analitycznego, dotyczącego wzorów kultury, pojawiło się **pytanie o dostępność zasobów dla konkretnych działań społecznych, a w szczególności działań z obszaru edukacji kulturalnej.** Związek między tymi pytaniami obrazowało, z kolei, założenie, że możliwe jest określenie zarówno nieoczywistych zasobów pomocnych, jak i stojących na przeszkodzie tejże edukacji. W projekcie „GIS Kultura” ważne było także **założenie o potrzebie wyjścia poza dwuwartościową kategorię wskazującą daną instytucję jako istniejącą bądź nieistniejącą (*jest dom kultury/nie ma domu kultury*).** Samo określenie liczby instytucji stanowi co najwyżej przyczynek do typologizacji obszarów przestrzennych, a z racji tego, że powszechnie wykorzystywane dane GUS agregowane są do poziomu gminy, można mówić tu o pewnego rodzaju typologizacjach gmin w Polsce, np. przez parametr infrastrukturalny.

Zdarza się, że na pytanie czy na terenie gminy jest dostępna biblioteka lub centrum kultury, odpowiada się, określając, czy obiekt ten znajduje się w granicach danej gminy, województwa.⁵ Na tej podstawie trudno jest np. stwierdzić czy korzystają z niego tylko mieszkańcy danej gminy czy także gmin sąsiednich (usytuowanie takiego obiektu może być przecież korzystniejsze dla sąsiadów). Inną kwestią jest czy obiekt działa, jakiego rodzaju ofertę kulturalną posiada, czy jest dostosowany dla osób o specjalnych potrzebach (dla osób z niepełnosprawnością ruchową lub dla rodziców z dzieckiem). Chociaż część tych danych jest dostępna w BDL, to ze względu na tajemnicę statystyczną nie są one odpowiednio udostępniane. Np.: jeśli nawet w gminie istnieje więcej niż jeden obiekt określonego rodzaju to nie sposób do-

⁵Problematykę tę podjęto w publikacji Zespołu ds. Statystyki Kultury NCK w raporcie: Statystyka kultury w Polsce i Europie. Aktualne zagadnienia http://nck.pl/media/attachments/317207/Statystyka%20kultury%20raport_oWV5leL.pdf,

wiedzieć się, który z nich jest dostosowany dla osób z niepełnosprawnością, a który nie. Oprócz tego zdarzają się tutaj błędy mające różne przyczyny. Mogą być to błędy w sprawozdawczości, które popełniają osoby wypełniające formularze lub przygotowujące zestawienia. Przy tej skali przedsięwzięcia (2478 gmin opisywanych przez setki a nawet tysiące konkretnych danych) można więc mówić o pewnego rodzaju przybliżeniach, szacunkach, a nie o danych precyzyjnych. Z wielu powodów, których tu nie będę opisywał, trudno też oszacować skalę błędów.

Dla osób badających kulturę, dane GUS nie są wystarczające. Istnieje bowiem potrzeba bardziej precyzyjnych, w tym lepiej ukontekstowanych, informacji. Inna sprawa, że badacze sami rzadko starają się budować konteksty dla danych, którymi się zwyczajowo zajmują. Ponadto kultura sama w sobie jest trudna do zmierzenia, a w sytuacji, w której jedno z najprostszych narzędzi pomiarowych nie działa jak należy, dopracowywanie metodologii jest znacznie ograniczone. Dlatego też projekt „GIS Kultura” - zanim odpowie na pytanie, czy kultura jest dostępna - musi być załącznikiem alternatywnego systemu zbierania danych o kulturze. Projekt miał więc na celu stworzenie prototypowego narzędzia, które na podstawie jak najszerszych źródeł informacji, pozwoli na analizy stanu kultury w Polsce, poczynając od materialnego zaplecza – czyli infrastruktury.

Dalsza część raportu jest sprawozdaniem ze wszystkich zrealizowanych w projekcie etapów prac. Są więc w nim części dotyczące: założeń operacyjnych, etapu gromadzenia i klasyfikacji danych wykorzystanych w zharmonizowanej bazie danych, etapu prac technologicznych wykonanych przez partnera projektu oraz przykłady zastosowań, które są możliwe do wykonania na prototypie narzędzia. .

Efekt prac dostępny jest w trojaki sposób - w zależności od potrzeb potencjalnych użytkowników. Po pierwsze, jest to geoportal kultury dostępny pod adresem (<http://ozkultura.pl/node/30>); po drugie, serwis WFS dostępny pod adresem (http://zks.gik.pw.edu.pl/ozk_portal_wfs/) oraz po trzecie, zrzut bazy danych z ostatniego etapu prac, dystrybuowany na życzenie.

Rozdział 3

Badanie dostępności

Operacjonalizacja założeń i zasięgu systemu

Zbudowanie systemu analiz przestrzeni kulturalnej wymagało w pierwszym etapie określenia zakresu i „głębokości” zbierania danych. Kwestia zakresu wynika z teoretycznej koncepcji żywej kultury, zaś „głębokość” czy też szczegółowość zbieranych informacji o poszczególnych obiektach ma odpowiedzieć na pytanie: czy kultura jest dostępna? Warto tu zaznaczyć, że nie chodzi o wizję kultury pasywnej, wyłącznie odbieranej, dziejącej się tylko w pewnych typach instytucji. Poza tym, gdy mówi się o systemie kultury, można oczekiwać istnienia pewnego rodzaju homeostazy, w którym funkcje będą przenoszone między instytucjami społecznymi. Zmiana w takim rozumieniu jest częścią działań w obszarze rekonfiguracji systemu. To działania ludzkie, interakcje zorganizowane i nieformalne, sprawiają, że system ten ulega przekształceniom. By uchwycić dostępność kultury trzeba więc jak najszybciej zdefiniować miejsca, w których organizowana jest aktywność społeczna. Trzeba zarazem mieć na uwadze, że nie każdy obiekt infrastruktury żywej kultury pełni taką samą rolę w organizowaniu życia społecznego, część służy przede wszystkim realizacji potrzeb podstawowych, część odpowiada za kwestie związane z jakością życia jednostkowego, reszta - za wytwarzanie i podtrzymywanie więzi społecznych. Takie ujęcie pokazuje, że system wychodzi poza wąskie monitowanie **sektora kultury**, czy też badanie *uczestnictwa*

w życiu kulturalnym. Ponieważ zajmujemy się kulturowym wymiarem życia społecznego musieliśmy w pierwszym etapie określić instytucje towarzyszące człowiekowi, poczynając od jego narodzin, aż do śmierci.

Z jednej strony, istnieje sprzeciw wobec tak szerokiego ujmowania świata praktyk kulturowych, gdyż w potocznym rozumieniu kultura jest czymś odświętnym, spoza porządku dnia codziennego¹. W takim myśleniu o kulturze, ujawnia się pewna wewnętrzna sprzeczność. Sprzeczność tę można zobrazować na przykładzie bibliotek. Jest to typ instytucji utożsamiany z wąsko rozumianymi praktykami kulturowymi. Jeśli jednak biblioteka przynależy do tego typu obszaru, to czy jej funkcjonowanie nie powinno pokrywać się z tym, co jest odświętne, spoza dnia codziennego? W rzeczywistości jednak ile bibliotek w skali kraju prowadzi swoją działalność np. w niedzielę? Przy okazji pojawia się kolejne, empirycznie interesujące pytanie - ile kultury dostępnej jest od święta albo – nie przymierzając – w tzw. czasie wolnym? Wąskie myślenie o tym, czym jest kultura nie odpowiada temu, co się dzieje w rzeczywistości społecznej. Bo to właśnie praktyki kulturowe, polisemiczne, zmienne, nadają sens myśleniu o tym, czym w rzeczywistości kultura jest. W tym rozumieniu rozdzielanie sfer życia na kulturowe i pozakulturowe doprowadza do przyjęcia paradoksalnego założenia o istnieniu kultury uczestnictwa i nieuczestnictwa. Takie szerokie podejście można znaleźć w pracach choćby Antoniny Kłoskowskiej.²

W naszych badaniach kultury szeroki zakres ujmowanych zjawisk przypomina też o złożoności systemu społecznego i relacjach między jego elementami. Świat społeczny zwykle rekonstruowany jest poprzez poszczególne fragmenty, empirycznie uchwytnie. Natomiast dopiero zrozumienie relacji między nimi pozwala na skuteczne projektowanie działań.

Tworzenie narzędzi w ramach systemu „GIS Kultura” poprzedzały założenia o niedyskryminowaniu żadnych źródeł informacji o obiektach za-

¹Rafał Drozdowski, Barbara Fatyga, Mirosław Filiciak, Marek Krajewski, Tomasz Szlendak, Zakończenie, (w:) *tychże, Praktyki kulturalne Polaków*, Toruń, :Wyd. UMK, 2014, zob. też: http://ozkultura.pl/sites/default/files/strona-archiwum/PRAKTYKI_KULTURALNE_POLAKOW.pdf

²Antonina Kłoskowska, *Kultura masowa. Krytyka i obrona*, Warszawa: PWN, 1980, s.40

kwalifikowanych przez nas jako obiekty kultury. Celem było zebranie danych o tych obiektach żywej kultury, definiowanych w szeroki sposób i później-sze sklasyfikowanie ich ze względu na funkcję kulturową jaką pełnią. W tym sensie każdy obiekt na mapie jest obiektem kultury, każdy też wyposażony jest w bogate sieci aksjosemiotyczne (wartości i znaczenia).

Prace nad metodologią klasyfikacji obiektów rozpoczęły się przed realizacją właściwego projektu „GIS Kultura”, w 2013 roku. Wówczas to, na bazie inwentarza obiektów zebranych ze stron internetowych gmin, zostało stworzonych 28 typów obiektów infrastruktury.³ W trakcie gromadzenia danych do typologii tej został dodany 29 typ, czyli infrastruktura transportowa. W celu ujednoznacznienia typów zostały doprecyzowane ich nazwy, ale zasadniczy sposób klasyfikacji nie uległ zmianie. Dodatkowo, w trakcie prac nad interfejsem przedstawiania danych, zostały stworzone kategorie grupujące te obiekty. Wynikają one z założenia o funkcjonalności kultury wobec życia społecznego, co tu znaczy tyle, iż każda instytucja pełni określoną społecznie rolę. Można rzec, że kultura jako system pełni funkcję regulatywną dla wszystkich wymiarów i poziomów ludzkiego życia: od biologicznego po społeczny. Zinwentaryzowane przez nas obiekty zostały pogrupowane w indeksy właśnie ze względu na bardziej szczegółowe funkcje: regulujące potrzeby podstawowe, regulujące jakość życia, regulujące więzi społeczne. Listę 29 typów i przynależności do jednego z trzech typów regulujących przedstawia Tabela 3.1.

Zorganizowany system klasyfikacji informacji o żywej kulturze mierzy się także z problematyką badania dostępności. Dostępność stanowi jedno z zagadnień, które w sensie empirycznym jest poruszane w różnych dziedzinach nauki: pojawia się ono zarówno w geografii społecznej, ekonomii, jak i w socjologii i antropologii. W projekcie „GIS Kultura” kategoria ta zdefiniowana została jako **możliwość zaistnienia relacji między więcej niż jednym elementem zbioru**. Jest to definicja bardzo ogólna i jak na humanistyczne gusta - minimalistyczna, ale warto zaznaczyć, że wyraźnie odcina się od mierzenia realnych relacji między elementami zbioru. Dostępność transportowa, czy wężiej - komunikacyjna, w tym rozumieniu będzie

³O tym etapie więcej w Aneksie 1 i 2 do niniejszego raportu.

Tabela 3.1: Lista indeksów obiektów żywej kultury.

Nr	Nazwa indeksu infrastruktury	Funkcja regulująca
1	Edukacji i nauki	jakości życia
2	Zbiorowych spotkań okazjonalnych	więzi społecznych
3	Kultury "wysokiej"	jakości życia
4	Upowszechniania kultury	jakości życia
5	Aktywności fizycznej	potrzeb podstawowych
6	Kultury religijnej	jakości życia
7	Gastronomii	potrzeb podstawowych
8	Turystyki	jakości życia
9	Handlowa	potrzeb podstawowych
10	Produkcji i dystrybucji informacji	jakości życia
11	Usług kulturalnych	jakości życia
12	Oswojonej natury	potrzeb podstawowych
13	Upiększania/dbania o ciało	jakości życia
14	Magii/medycyny para- i niekonwencjonalnej	jakości życia
15	Organizacji życia zbiorowego	więzi społecznych
16	Zdrowia	potrzeb podstawowych
17	Rzemiosła i kultury ludowej	jakości życia
18	Instytucji finansowych	więzi społecznych
19	Handlu kulturą	jakości życia
20	Przemysłów kultury	jakości życia
21	Nieformalna/subkulturowa	więzi społecznych
22	Administrowania kulturą	więzi społecznych
23	Informacyjno-promocyjna	jakości życia
24	Obsługi społecznego funkcjonowania jednostek i grup	potrzeb podstawowych
25	NGO's	więzi społecznych
26	Infrastruktura pomocy społecznej	więzi społecznych
27	Produkcyjna	potrzeb podstawowych
28	Usług produkcyjnych	potrzeb podstawowych
29	Transportowa	regulacji jakości życia

Źródło: Mariusz Piotrowski, opracowanie własne, na podstawie tekstów

metodologicznych Barbary Fatygi, Elizy Gryszko i Aleksandry

Orkan-Łęckiej dostępnych pod adresami:

<http://ozkultura.pl/node/1650>; <http://ozkultura.pl/node/1800>;

<http://ozkultura.pl/node/955>.

określała relacje między domem kultury a odległością do najbliższego przystanku komunikacji publicznej, nie zaś liczbę osób korzystających z transportu publicznego w celu dotarcia do domu kultury. Aby móc odpowiedzieć na pytanie o dostępność sformułowane ze wskazanego punktu widzenia należałoby zbudować całkowicie nowy system zbierania danych, wystandaryzowany pod kątem wielu różnych podmiotów. W tym sensie projekt „GIS Kultura” pozwala na badanie dostępności, nie zaś „ruchliwości”, czyli faktycznego przemieszczania się ludzi.⁴ Budowa systemu stała się okazją, żeby wyjść poza badanie dostępności przestrzennej, czyli relacji określających odległości między obiektami bądź to w linii prostej, bądź wg istniejącej siatki ulic i ścieżek pieszych. System został zaprojektowany w taki sposób, żeby zbierać informacje także o dostępności informacyjnej, dostępności dla grup osób o specjalnych potrzebach czy wielkości obiektu ze względu na pełnioną funkcję - np. w teatrach byłyby to liczba miejsc siedzących na sali. Warto cały czas pamiętać, że *gros* tych danych jest zbieranych w ramach sprawozdawczości GUS i jeśli tylko zmieniłby się status tych informacji w ramach statystyk publicznych (albo – *nomen omen* - sposób ich udostępniania), mogłyby być one zintegrowane w ramach tego systemu.

Wymienione powyżej wymiary dostępności mają być elementami równania, które w konsekwencji odpowie na pytanie - jaka jest dostępność kultury. W słowniku portalu ozkultura.pl „dostęp/dostępność kultury” została zdefiniowana następująco: „dostęp/dostępność kultury to: a) na poziomie jednostki - możliwy do zmierzenia, (choć nie jest to łatwe), zakres wykorzystania przez nią zasobu kulturowego, pozostającego w dyspozycji grup i społeczności, których jednostka ta jest członkiem; b) na poziomie grupy oznacza - także mierzalne i także niełatwe - wykorzystanie zasobu kulturowego będącego w zasięgu jej możliwości i potrzeb. Dostęp do kultury uwarunkowany jest wieloma czynnikami: biologicznymi (np.: indywidualnymi cechami fizycznymi, niepełnosprawnością, itd.), demograficznymi (np.: takimi jak wiek, płeć), społecznymi (np.: przynależnością do warstwy społecznej, pełnionymi rolami zawodowymi), geograficznymi (np.: zróżnicowa-

⁴Zbigniew Taylor, *Przestrzenna dostępność miejsc zatrudnienia, kształcenia i usług a codzienna ruchliwość ludności wiejskiej*, (w:) „Prace Geograficzne IGI PAN”, nr 171, Wrocław: Wyd. Continuo, 1999, s. 239.

niami terytorialnymi, miejscem zamieszkania), ekonomicznymi (np.: poziomem dochodów), kulturowymi (np.: wykształceniem, poziomem kompetencji kulturowych) oraz emocjonalnymi. Z drugiej strony - należy pamiętać, że wszystkie wymienione czynniki mogą wpływać na fakt, iż jednostki lub grupy będą wykorzystywały dostęp do kultury w sposób selektywny: kierując się własnymi potrzebami kulturalnymi, upodobaniami, nawykami, względami towarzyskimi czy nawet nastrojem w danej chwili. (W ten sposób tworzą one jednak własne wersje subkulturowe, kultury rozumianej całościowo i/lub oficjalnie). O tym fakcie często zapominają twórcy oferty kulturalnej, zwłaszcza ci, którzy uważają, że to, co dają odbiorcom musi być skonsumowane”.⁵ Wymienione tutaj elementy różnicujące korzystanie z oferty instytucji, zarówno na poziomie jednostkowym, jak i grupowym, dla potrzeb opisywanego projektu muszą zostać przesunięte z obszaru kapitału kulturowego pozostającego w dyspozycji jednostek i grup, na poziom zasobu będącego w dyspozycji instytucji. Zamiast więc badać: *w jakim stopniu osoby z niepełnosprawnością ruchową korzystają z oferty instytucjonalnej, chcielibyśmy najpierw określić: czy sama instytucja stwarza takim osobom warunki dla korzystania ze swojej oferty*. Analogicznie - dostępność informacyjna więcej mówi o istnieniu kanałów komunikacyjnych (adresie email, stronie www, profilu w serwisie społecznościowym), niż o liczbie maili wysłanych do instytucji, o liczbie wejść na jej stronę, czy liczbie osób śledzących/lubiących dany profil w serwisie społecznościowym. Dostępność miejsc siedzących na sali także nie odnosi się do wypełnienia tej sali przez widownię w trakcie organizowanych wydarzeń. Dopiero zebranie wiedzy o istniejącym zasobie i poziomie jego wykorzystywania. może przyczynić się do sformułowania przynajmniej wstępnych wniosków o stanie kultury na danym terenie. System „GIS Kultura” ma za zadanie określić ten zasób, co w dalszej kolejności ma ułatwić zbieranie informacji o korzystaniu z kultury i stworzyć jeden z ważnych kontekstów dla innych badań; w tym uczestnictwa. Aby osiągnąć ten cel należałoby stworzyć system zbierający informacje również o ofercie kulturalnej, w takim samym zakresie jak zostały zebrane i są zbierane informacje o zasobach infrastrukturalnych. Niemniej jednak budowa takiego systemu musiałaby uwzględniać także inny sposób myślenia o kulturze,

⁵ Barbara Fatyga, Dostępność kultury, definicja autorska dla OŻK-SB,(w:) Wieloźródłowy Słownik Kultury OŻK, 2013, <http://ozkultura.pl/wpisy/113>

w którym praktyki kulturalne realizowane nie tylko przez sformalizowane struktury. System musiałby mieć też możliwość uchwycenia działań nieformalnych, spontanicznych, jednorazowych. Badanie istniejącej infrastruktury żywej kultury jest więc częścią szeroko zakrojonego projektu badawczego, który nazwaliśmy „Mapą wiedzy i niewiedzy o polskiej kulturze”.⁶ W kolejnym rozdziale zostało opisane, jakiego typu informacje zostały wykorzystane w projekcie „GIS Kultura”, wraz z ich oceną jakościową.

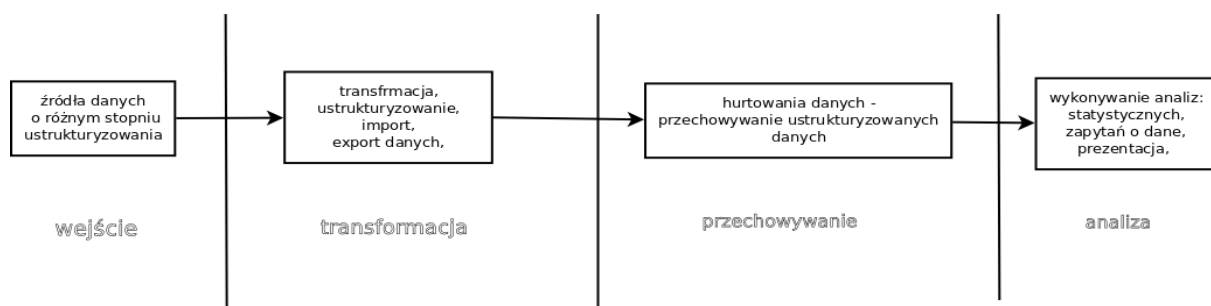
⁶ Autorką samej nazwy jest członkini Zespołu Węzła Centralnego FOŻK-SB - dr hab. Magdalena Dudkiewicz.

Rozdział 4

Dane wejściowe

Opis modelu systemu informacji przestrzennej o infrastrukturze kultury można przedstawić w formie czteroelementowego schematu. Zostały w nim uwzględnione źródła danych, narzędzia do przetwarzania i przechowywania danych (osobno) oraz proces analizy. Proces ten prezentuje Rysunek 4.1. Jego elementy są obowiązujące dla każdego projektu badawczego, w którym na dużą skalę przetwarzane są dane czy to liczbowe, czy tekstowe. Rozdział ten dotyczy etapu pozyskiwania danych i prezentuje także pewne podstawowe informacje charakteryzujące główne źródła danych.

Rysunek 4.1: Schemat procesu projektowego GIS Kultura



Źródło: Mariusz Piotrowski, opracowanie własne

Pierwotnym założeniem projektu było skupienie się na zebraniu szerokich informacji o obiektach w województwach mazowieckim i warmińsko-mazurskim, oraz w ograniczonym stopniu (do poziomu jednego indeksu ży-

wej kultury) zgromadzenie informacji z całego kraju. W tym celu podjęto próbę stworzenia crawlera ¹, który używany przez korespondentów lokalnych pozwalałaby zbierać informację o infrastrukturze. Równolegle zostały podjęte próby zebrania danych z serwisów agregujących szeroki zakres informacji adresowych oraz z Bazy Danych o Obiektach Topograficznych 10k (BDOT10k). Rozważano zatem trzy kierunki podjęcia działań dotyczących pozyskania informacji. Pierwszy miał na celu zebranie podstawowej informacji ze stron jednostek samorządu terytorialnego. Drugi miał na celu zebranie i przekategoryzowanie danych zgromadzonych przez komercyjne serwisy internetowe. Trzeci natomiast – miał prowadzić do pozyskania dostępu do danych publicznych.

Problem z pierwszym rozwiązaniem uwidocznił się na etapie projektowania crawlera. Wynikał on z braku standaryzacji informacji umieszczanej na stronach gminnych. W wielu przypadkach strony Biuletynu Informacji Publicznej (BIP) prezentowały niewielki wycinek informacji w stosunku do tego, co znajdowało się na stronach gminnych. Strony BIP mają jednolitą strukturę organizowania informacji, zaś strony gminne są tworzone w różnorodnych technologiach. To, z kolei, uniemożliwiało ekstrakcję ustrukturyzowanych informacji. Testowane strony gminne były przygotowywane w różnych standardach i dodatkowo część posiadała zabezpieczenia przed *crawlingiem*. Przy zakładanej pierwotnie w projekcie liczbie gmin do zbadania (413 gmin oraz dodatkowo powiatów i witryn wojewódzkich), należałoby zbudować podobną liczbę *crawlerów*. Zbudowany prototyp narzędzia nie sprawdził się w pozyskiwaniu informacji na taką skalę². Interfejs *crawlera* prezentuje Rysunek 4.2.

Przy tak zarysowanym problemie badawczym, czyli zebraniu szerokiego typu informacji o obiektach, problem techniczny przerodził się w problem budowy wyszukiwarki internetowej oraz systemu klasyfikacji obiektów. Tak zbudowany automatyczny system wyszukiwania dla każdego obiektu powinien precyzyjnie gromadzić informację o 1) nazwie, 2) funkcji, dzięki

¹Program zbierający informacje o strukturze, stronach i treściach znajdujących się w internecie, źródło: https://pl.wikipedia.org/wiki/Robot_internetowy

² Narzędzie to przydało się natomiast do pilotażu projektu, w tym do pierwszej inwentaryzacji danych, koniecznej przed etapem tworzenia indeksów infrastruktury.

której możliwe byłoby przyporządkowanie obiektowi indeksu żywej kultury i 3) lokalizacji, adresie, z nazwą ulicy i numerem budynku, ale także z jednoznaczną nazwą miejscowości. Droga ta została porzucona jako zbyt kosztowna i czasochłonna. Jednak w przyszłości do tego typu zadań pomocne mogą być rozwiązania z zakresu machine learning³. Technologie te stają się z dnia na dzień coraz bardziej dostępne.

Rysunek 4.2: Okno crawlera.

Źródło: Michał Kaszlej, twórca crawlera

Problem z drugim rozwiązaniem, czyli z zebraniem i przekatego-

³Uczenie maszynowe - Machine learning albo samouczenie się maszyn, systemy uczące się (ang. machine learning) – dziedzina wchodząca w skład nauk zajmujących się problematyką SI (patrz sztuczna inteligencja). Jest to nauka interdyscyplinarna ze szczególnym uwzględnieniem takich dziedzin jak informatyka, robotyka i statystyka. Głównym celem jest praktyczne zastosowanie dokonań w dziedzinie sztucznej inteligencji do stworzenia automatycznego systemu potrafiącego doskonalić się przy pomocy zgromadzonego doświadczenia (czyli danych) i nabywania na tej podstawie nowej wiedzy. Uczenie maszynowe jest konsekwencją rozwoju idei sztucznej inteligencji i metod jej wdrażania praktycznego. źródło:https://pl.wikipedia.org/wiki/Uczenie_maszynowe

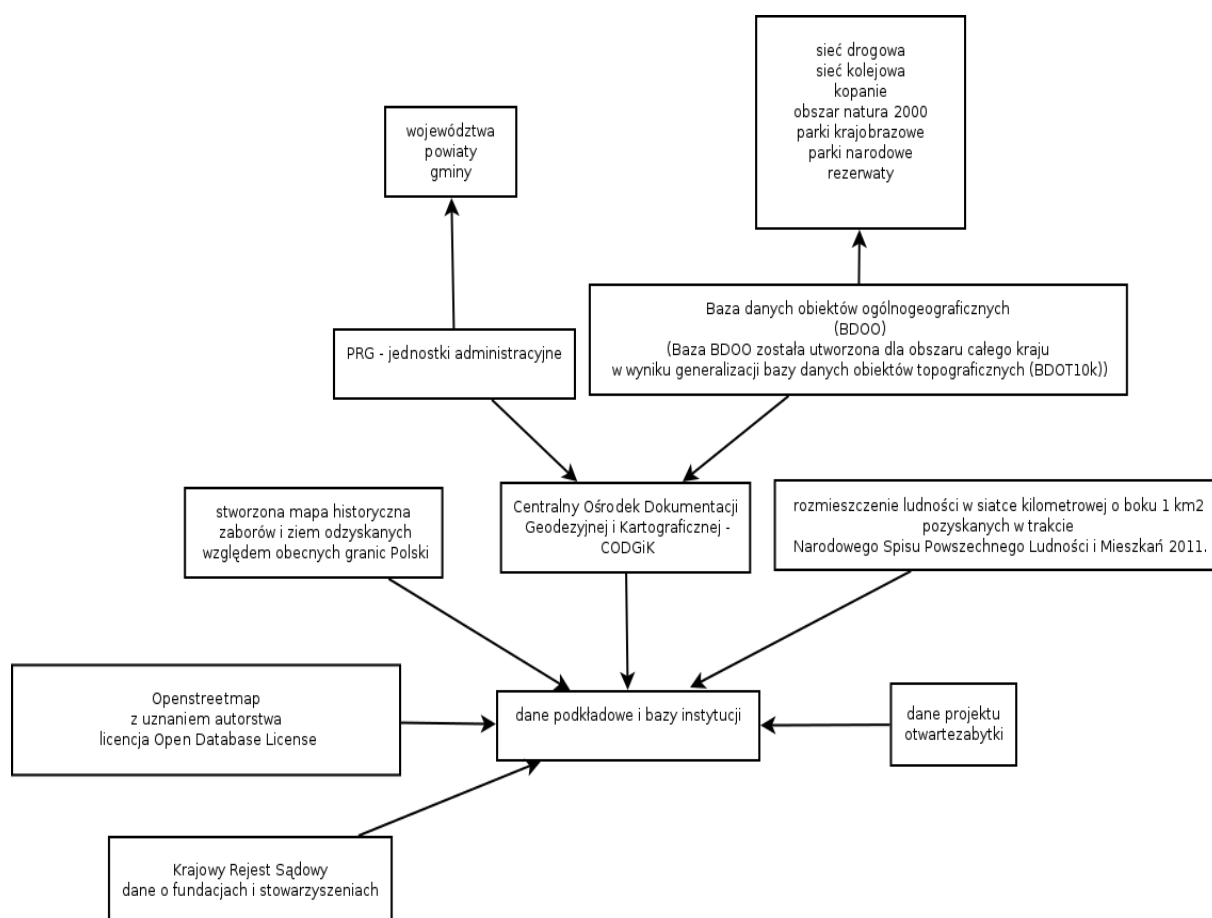
ryzowaniem danych zgromadzonych przez komercyjne serwisy internetowe, miał charakter nie tyle technologiczny, ile prawny i dotyczył możliwości przetwarzania takich informacji. Koszty zdobycia praw do zgromadzonych danych przekraczały środki przeznaczone w projekcie na pozyskanie informacji. Rozważane były dwa źródła informacji: dane zebrane i udostępnione w serwisie Google Maps oraz dane Panoramy Firm. Nieprecyzyjny status danych, sposób ich zbierania przez te serwisy, z trudną do określenia perspektywą aktualizacji wyeliminował to rozwiązanie, jako bazowe pod kątem budowanego systemu.

Jednocześnie, wraz z zespołem prof. Roberta Olszewskiego z Wydziału Geodezji i Kartografii Politechniki Warszawskiej, od samego początku podjęliśmy działania zmierzające do wykorzystania danych zharmonizowanych w Bazie Danych o Obiektach Topograficznych 10k (BDOT10k), wypracowanej w ramach dyrektywy UE - Inspire. Otrzymanie tej bazy dla celów naukowych okazało się jednak niemożliwe. W odpowiedzi na skierowane do Głównego Geodety Kraju pismo, w którym prof. Olszewski zwracał się o dostęp do bazy, otrzymaliśmy negatywną odpowiedź sformułowaną przez prawnika. Tę niekorzystną dla nas decyzję argumentowano następująco: *dla dysponentów bazy nie jest jasne, czy prof. PW Robert Olszewski zwracał się o jej udostępnienie w imieniu własnym (jako Robert Olszewski), czy też jako przedstawiciel świata nauki (prof. Politechniki Warszawskiej)*. Realizowanie projektu w ramach programu MKiDN okazało się niewystarczającym powodem. Jednocześnie środki przeznaczone na projekt, nie pozwalały na zakup tej bazy. W efekcie okazało się, że baza BDOT10k jest tylko teoretycznie bazą publiczną, a dostęp do niej jest ograniczany na wiele różnych sposobów.

Faktycznym sercem „GIS Kultury” stały się dwa źródła danych, zbierające różne informacje. Oba o zasięgu ogólnokrajowym. Pierwsze to serwis openstreetmap.org (OSM), drugim zaś - dane o działalności gospodarczej zgromadzone w Krajowym Rejestrze Sądowym (KRS). Dodatkowo informacje te są na bieżąco poszerzane o inne źródła. Spis źródeł danych prezentuje Rysunek 4.3.

Dane o działalności gospodarczej gromadzi także baza Centralnej

Rysunek 4.3: Źródła danych w projekcie GIS Kultura



Źródło: Mariusz Piotrowski, opracowanie własne

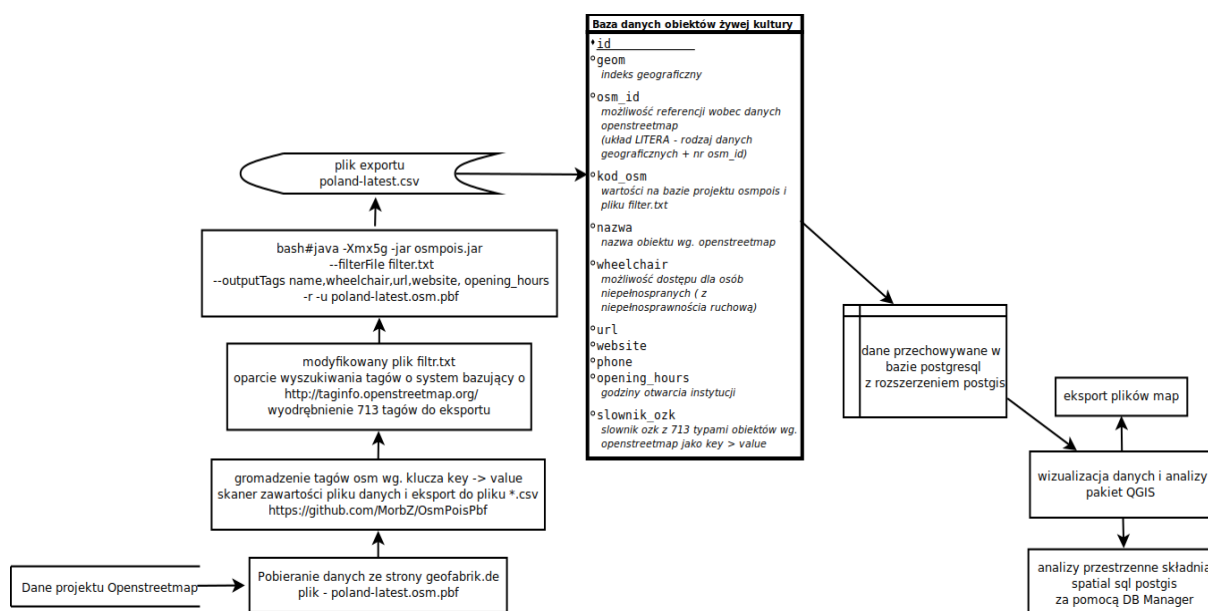
Ewidencji i Informacja o Działalności Gospodarczej. Została ona przez nas pozyskana, lecz - jak się okazało- nie spełnia zasady komplementarności danych. Znajdują się w niej informacje głównie o nazwie i adresie, pod którym zarejestrowano działalność. Brakuje natomiast możliwości pozyskania informacji o charakterze prowadzonej działalności.

Podstawowy problem z przełożeniem danych projektu OSM na kategoryzację obiektów żywej kultury wynika ze sposobu organizacji zasobów w dwóch projektach. OSM bazuje na tagach przypisywanych do obiektu przestrzennego, którym może być punkt, linia, powierzchnia czy też relacja przestrzenna. Złożoność problemu można zaprezentować na przykładzie lasu. Może być on punktem, czyli informacją przybliżoną, lub też może być przedstawiony jako pełna informacja o jego granicach (wówczas będzie to powierzchnia). Ścieżki w lesie będą liniami; jeśli las stanowi część rezerwatu, taka informacja będzie zapisana w formie relacji przestrzennej. Formuła tagów w projekcie OSM jest organizowana przy pomocy dość swobodnie określanych relacji klucz $\rightarrow$ wartość (*key* $\rightarrow$ *value*). Z jednej strony, lista kluczy i wartości jest negocjowana w drodze konsensusu społeczności mapującej; z drugiej jednak, sam proces wpisywania angażuje rzesze użytkowników.⁴ W konsekwencji można wyróżnić obszerną listę możliwych błędów charakterystycznych dla projektów społecznościowych, nie wspominając już o aktach wandalizmu, które wpływają na zmianę stanu infrastruktury. Z punktu widzenia tworzenia systemu „GIS Kultura” szczególnie dwa typy błędów stanowiły wyzwanie przy projektowaniu systemu pobierania danych. Pierwszy typ to błędy wynikające ze złego przypisania kategorii do obiektu, nieodzwierciedlające jego rzeczywistej funkcji. Może być to, po prostu, brak odpowiedniej wartości (*value*), mimo właściwej klasyfikacji w kluczu (*key*). Skutkiem tego stanu rzeczy jest często nadużywanie kategorii inne (*other*, *unknown*). Drugi typ to błędy pojawiające się w trakcie wprowadzania danych przez użytkowników do systemu, czyli błędy *drukarskie*. Biorąc pod uwagę te dwa typy błędów, należało zbudować taki mechanizm wydobywania danych z bazy OSM, który pozwala precyzyjnie przekształcać dane z systemu opartego o

⁴Liczba osób, które kiedykolwiek dodały jakiś element do map szacowna jest na 800000 tys.(stan na 10.2016) źródło http://wiki.openstreetmap.org/wiki/Stats#Database_statistics_-_Graphical_charts_.28since_2005.29

tagi (OSM) na system oparty o jednorodne przyporządkowanie obiektu do klasy („GIS Kultura”). Należało uwzględnić także nieprecyzyjny charakter danych. Budowa narzędzia zaczęła się od stworzenia **pliku słownika**, w którym 713 najczęściej używanych **tagów** służących do opisywania obiektów zostało przyporządkowanych do jednego z **29 typów indeksów żywej kultury**. Dodatkowo został użyty **kod określający obiekty niezidentyfikowane**. Plik słownika został przygotowany na tyle szeroko, że znajdują się w nim także nazwy geograficzne miejscowości, osad i przysiółków. Informacje te mogą być pomocne przy generowaniu mapy zasadniczej. Schemat pracy z danymi o punktach z bazy OSM przedstawia Rysunek 4.4.

Rysunek 4.4: Schemat przetwarzania danych z projektu openstreetmap.



Źródło: Mariusz Piotrowski, opracowanie własne

Dane z serwisu OSM pobierane są za pośrednictwem strony Geofabrik⁵, która codziennie udostępnia plik zbiorczy projektu, „przycięty” dla różnych obszarów geograficznych. Warto zauważyć, że dane znajdujące się na stronie z mapą openstreetmap.org, to tylko część danych zbieranych w projekcie. Dane służące renderowaniu⁶ mapy mają mniejszy zakres niż te pobrane np. ze strony Geofabrik. Plik dystrybuowany jest w formacie pbf,

⁵<http://download.geofabrik.de/europe/poland.html>

⁶ Przekształcanie danych geoinformatycznych w mapę obrazkową.

który posiada strukturę podobną do schematu xml, ale jest o 30% mniejszy.⁷ Format ten wymaga zastosowania narzędzi parsujących⁸, które wyodrębnią tylko część informacji zgodnie z intencją zastosowania ich w projekcie GIS Kultura. W celu parsowania pliku został wykorzystany program dystrybuowany na licencji GPL - OsmPoisPbf⁹ Plik wynikowy zmodyfikowano przez użycie specjalnie dopasowanego pliku `filter.txt` - wyodrębniającego specyficzne klucze i wartości.¹⁰ Ostateczny wynik, czyli dane o punktach z dodatkową informacją o dostępności, został wygenerowany przez użycie komendy i dodatkowych parametrów, które wyodrębniały dane o dostępności dla wózków inwalidzkich (*wheelchair*), posiadanej strony internetowej (*url*, *website*), godzin otwarcia (*opening_hours*). W efekcie otrzymywano plik w formacie .csv, który posiada następującą strukturę: dane o nazwie, długości i szerokości geograficznej, kod identyfikacji typu obiektu, dane o dostępności dla osób z niepełnosprawnością, dostępności informacyjnej i dostępności wg godzin otwarcia. Tak spreparowane dane są gotowe do przechowania w bazie PostgreSQL.¹¹

Krajowy Rejestr Sądowy jest informatyczną bazą danych, w której znajdują się informacje o trzech typach podmiotów: 1) przedsiębiorstwach, 2) stowarzyszeniach, innych organizacjach społecznych i zawodowych, fundacjach, 3) oraz publicznych zakładach opieki zdrowotnej. Baza danych gromadzi także informacje o dłużnikach niewypłacalnych. Natomiast Centralna Ewidencja i Informacja o Działalności Gospodarczej zbiera dane o przedsiębiorcach będących osobami fizycznymi. Te dwie bazy stanowią największe źródło danych o oficjalnym życiu gospodarczym i społecznym w kraju. Każdy z podmiotów określa zakres działalności wg schematu Polskiej Klasyfikacji Działalności (PKD). W czasie prowadzenia prac gromadzących dane, podmioty klasyfikowane były wg schematu z roku 2007 (PKD-2007), natomiast w bazie znajdowały się podmioty, które posiadały klasyfikację z

⁷ http://wiki.openstreetmap.org/wiki/PBF_Format

⁸Parser lub Analizator składniowy umożliwia przetworzenie tekstu czytelnego dla człowieka w strukturę danych przydatną dla oprogramowania komputera. źródło:https://pl.wikipedia.org/wiki/Analizator_składniowy

⁹<https://github.com/MorbZ/OsmPoisPbf>

¹⁰Aneks 3

¹¹Wolno dostępny system zarządzania relacyjnymi bazami danych źródło: <https://pl.wikipedia.org/wiki/PostgreSQL>

wcześniejszego okresu - z roku 2004. Pełną informację o schemacie PKD prezentuje dokument wyjaśniający, znajdujący się na stronie GUS.¹² Za aktualność danych w KRS i CEIDG odpowiada podmiot składający wniosek, przedsiębiorca, stowarzyszenie. Na nim też ciąży odpowiedzialność za aktualizację danych, np. aktualizację PKD, niemniej jednak za niedopełnienie tego obowiązku nie są przewidziane żadne sankcje. Może to wyjaśniać, dlaczego dane te są nieprecyzyjne, szczególnie w przypadku bazy CEIDG.

W przypadku opisanych powyżej źródeł danych pojawiły się trzy trudności. Pierwsza jest związana z ekstrakcją tych danych ze stron KRS i CEIDG. Nie ma możliwości pobrania ich w formie ustrukturyzowanego pliku, każdy rekord należy pobrać osobno. Dane udostępniane są w formie pliku pdf, z informacjami tylko o jednym rekordzie. Drugi problem jest związany z potrzebą przypisania współrzędnych do danych adresowych. W ramach prac projektowych zadanie to zostało wykonane przez Fundację ePaństwo, która pobrała i przypisała współrzędne do danych. Wreszcie, kolejnym zagadnieniem było określenie typu działalności wg Polskiej Klasyfikacji Działalności i przełożenie jej na indeksy obiektów żywej kultury. W tym celu ponownie został przygotowany plik słownikowy klasyfikujący działalność PKD pod kątem indeksów obiektów żywej kultury. W tym przypadku także została wykonana specjalna procedura *translacyjna* - przypisywania kodu OŻK dla poszczególnych podmiotów. Znaczna część rekordów w KRS (także w CEIDG) posiadała wiele różnych kodów działalności. Zastosowaliśmy procedurę przypisywania rang wg działalności ze względu na częstotliwość używania pewnych kodów przynależących do obiektów żywej kultury.¹³ Dodatkowo ujawnił się problem z przypisaniem indeksów dla obiektów, które mają określoną pozycję PKD 96.09 - *Pozostała działalność usługowa, gdzie indziej niesklasyfikowana*. Można domniemywać, że będą tam obiekty należące do indeksu magii/medycyny para- i niekonwencjonalnej. Wnioski z analizy danych KRS i CEIDG pozwoliły zauważyć, że kategorie PKD są szczegółowe w przypadku różnorodnych form przemysłu i rolnictwa, natomiast w przypadku działalności usługowej są bardziej ogólne. Działalność kulturalna, a szczególnie animacyjna realizowana poza tradycyjnymi instytucjami kultu-

¹²http://stat.gov.pl/Klasyfikacje/doc/pkd_07/pdf/3_PKD-2007-wyjasnienia.pdf

¹³Aneks 4 – Algorytm P. Michalski.

ry, w obecnej formule informacji o gospodarce jest pomijana.

W trakcie prac nad projektem zmieniał się status dostępności danych publicznych. Szczególnie dotyczy to zasobów Centralnego Ośrodka Dokumentacji Geodezyjnej i Kartograficznej (CODGiK). Nieodpłatnie w ramach usługi PRG dystrybuowane są pliki z geometriami granic administracyjnych (plik z rozszerzeniami .shp)¹⁴ oraz dane z Bazy Danych Obiektów Ogólnogeograficznych (BDOO) (udostępniane jako pliki .gml)¹⁵. Niemniej jednak z punktu widzenia projektu GIS Kultura jakość udostępnianych danych publicznych nie okazała się zadowalająca. Udostępnione w 2015 roku dane z obrysami gmin, które mają pozwalać na badanie relacji w ramach administracji samorządowej, posiadały nieaktualne dane z kodem identyfikacji TERC w ramach systemu TERYT. Dane te należało ręcznie poprawić do stanu faktycznego na rok 2015.¹⁶

W ramach prac nad systemem analitycznym w projekcie pojawiła się także potrzeba weryfikacji hipotez historycznych. Często formułowaną hipotezą dotyczącą nierównomiernego rozwoju kraju jest odnoszenie się do historii jego obecnych obszarów (granice rozbiorów, historia tzw. ziem odzyskanych). W celu weryfikacji tych hipotez, przygotowaliśmy odpowiednie bazy. Jednostkom samorządu terytorialnego została przypisana klasyfikacja historyczna, zgodnie z typem: teren gminy znajdował się w zaborze rosyjskim, pruskim, austriackim bądź na obszarze przyłączonym do terytorium RP po 1945 roku. Zestawione dane prezentuje Rysunek 4.5. Podstawą do przygotowania danych były poprawione dane o gminy z CODGiK.

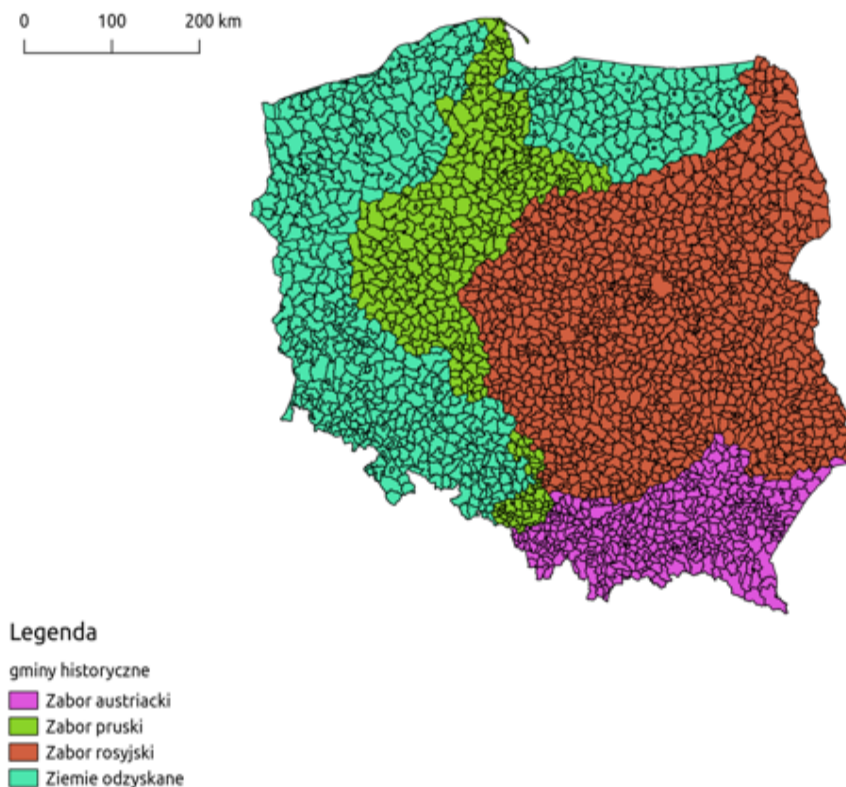
Aby w ramach projektu „GIS Kultura” możliwe było badanie dostępności w wymiarze społecznym, bazy danych przestrzennych powinny być zasilane danymi o charakterze socjodemograficznym. Minimalnym wymaganiem wobec takich analiz są dane o rozmieszczeniu ludności. Niestety, precyzyjne i satysfakcjonujące dane nie są ogólnodostępne. W celu zasilenia systemu zostały wykorzystane ostatnie dane ze spisu ludności (NSP 2011), które upubliczniono na poziomie obwodów spisowych i przełożono do pre-

¹⁴<http://codgik.gov.pl/index.php/darmowe-dane/prg.html>

¹⁵<http://codgik.gov.pl/index.php/darmowe-dane/bdo250gis.html>

¹⁶W styczniu 2017 roku na stronie pojawiły się jednak nowe, zaktualizowane pliki z granicami administracyjnymi

Rysunek 4.5: Gminy wg typu historycznego (stan na 1.01.2016).



Źródło: Mariusz Piotrowski, opracowanie własne

zentacji na siatce o boku 1 km.¹⁷ Wątpliwości, co do jakości danych wynikają z tego, że odzwierciedlają on stan sprzed 6 lat. Dodatkowo procedura ostatniego Narodowego Spisu Powszechnego to raczej duże badanie społeczne, niż spis powszechny w typowej dla niego formie.¹⁸

Sytuacja z danymi publicznymi, którym można przypisać aspekt geograficzny, z każdym rokiem ulega zmianie. Upubliczniane są nowe zbiory danych. Dzieje się to na poziomie centralnym i lokalnym. Niektóre gminy

¹⁷https://geo.stat.gov.pl/start/-/asset_publisher/jNfJiIujcyRp/content/id/36734

¹⁸Przykładowe uwagi, ze względu na sposób dostępu do samospisu drogą elektroniczną znajdują się artykuły <http://www.niepelnosprawni.pl/ledge/x/86254>

prezentują zestawienia tych danych w formie geoportali tematycznych. Bardzo rozbudowane serwisy tego typu mają Warszawa, Gdańsk, Kraków, Wrocław. Często, dane z poziomu gmin prezentowane są w formie surowej informacji, do której można mieć dostęp za pomocą interfejsów programistycznych (API). Jednak w trakcie prac nad projektem, to właśnie zdobywanie danych wymuszało najbardziej dynamiczne reagowanie na zmiany związane z podejściem do publikacji informacji. Gdy nasza Fundacja rozpoczynała projekt, szeroki dostęp do informacji publicznej był w powijakach. Mimo tego, że świadomość wagi tych informacji się zwiększyła, rzeczywiste podejście do problemu dostępności danych publicznych jest ciągle bardzo różnie interpretowane przez instytucje publiczne. Do pozytywnych zjawisk należą takie inicjatywy jak danepubliczne.gov.pl, portal Ministerstwa Cyfryzacji, który gromadzi dokumenty z różnych organów. Jest to właściwy kierunek działań, niemniej jednak można mieć wiele zastrzeżeń do funkcjonalności portalu z punktu widzenia jego potencjalnych użytkowników. Na stronie <https://danepubliczne.gov.pl/group/spoleczenstwo> widnieje 7 formatów danych, w których można pobrać informacje, z czego najpopularniejszym jest format pdf (brak w nim definicji schematu publikacji danych), drugim jest xls (format zamknięty i wykorzystywany przez oprogramowanie MS Office), trzecim jest format zip, czyli skompresowane archiwum, w którym może kryć się każdy rodzaj pliku. Format csv, który można uznać za minimalny względem publikacji danych jest użyty tylko raz. Są to dane syntetyczne o liczbie zameldowanych osób we Wrocławiu¹⁹

Nie tylko administracja publiczna zbiera i udostępnia dane o życiu społecznym. W projekcie „GIS Kultura” jako dodatkową warstwę referencyjną wykorzystane zostały dane z projektu otwartzabytki.pl prowadzonego przez Centrum Cyfrowe Projekt: Polska. Dane w projekcie mają dwojaki charakter, główną częścią są dane o zabytkach zbierane przez Narodowy Instytut Dziedzictwa, reszta - to dane zbierane przez społeczność, która wytworzyła się przy projekcie. Z punktu widzenia projektu ten – społeczny - wymiar wspólnotowego definiowania znaczeń, określania pewnych miejsc jako historyczne i znaczące, sprawia, że dostępna jest tu nowa jakość danych o rzeczywistości społecznej. Dane te nie stanowią źródła wiedzy o infrastruk-

¹⁹https://danepubliczne.gov.pl/group/spoleczenstwo?res_format=CSV

turze, ale pozwalają odpowiedzieć na pytanie, w jaki sposób przestrzeń jest definiowana ze względu na znaczenie dziedzictwa i pamięć o nim czyli jak interpretowane są lokalne i ponadlokalne tradycje?²⁰ Właśnie z tego punktu widzenia, dane te można potraktować jako interesujący wskaźnik wytwarzania wspólnych pamięci w społecznościach lokalnych. Dane z tego projektu są dystrybuowane w formie plików geojson. Niestety nie są udostępniane pełne zasoby opisowe, lecz jedynie podstawowe elementy identyfikujące obiekt ze względu na cechy przestrzenne, nazwę, czas powstania.

Proces radzenia sobie z tak różnymi typami danych z plików szczegółowo prezentuje część raportu stworzona przez zespół partnera projektu, czyli pracowników Wydziału Geodezji i Kartografii Politechniki Warszawskiej. Etap ten w procesie budowy i zarządzania systemem określa się jako transformacja.²¹ Charakterystyka zebranego materiału znajduje się w kolejnym rozdziale.

²⁰W interpretacjach tego wymiaru warto pamiętać o koncepcji tradycji Erica Hobsbawna i Terence'a Rangera, por. tychże, *Tradycja wynaleziona*, Kraków: Wyd. UJ, 2008.

²¹Miłosz Gnat, Raport końcowy z realizacji bazy danych przestrzennych i geoportalu Obiektów Żywej Kultury

Rozdział 5

Charakterystyka źródeł danych

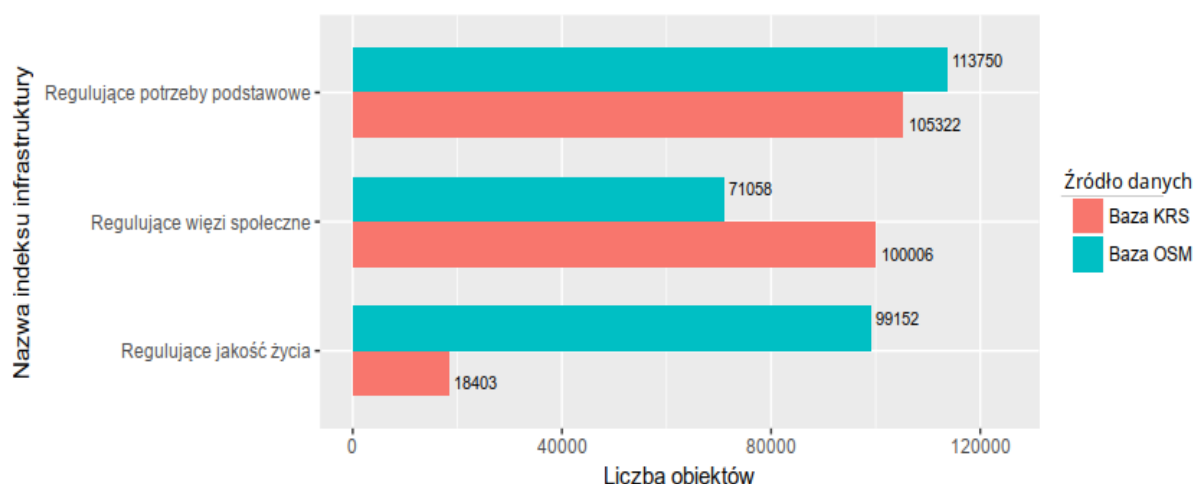
Statystyczny opis rozkładów indeksów żywej kultury i innych zmiennych prezentujących materiały źródłowe jest tu przedstawiony na przykładzie baz KRS i OSM. Te dwie bazy posiadają najwyższy wskaźnik dopasowania trzech elementów: nazwy (jako komponentu identyfikującego), lokalizatora (jako komponentu przestrzennego) oraz typu działalności (jako komponentu funkcjonalnego). Aby uznać, że dane spełniają minimalny warunek jakości, przyjęliśmy, że oprócz określenia przestrzennego rekordu musi być zaznaczona nazwa lub przynależność do konkretnego indeksu infrastruktury żywej kultury.

Szczególna sytuacja wystąpiła w wypadku danych z Centralnej Ewidencji i Informacji o Działalności Gospodarczej. Problem z bazą CE-IDG ma charakter prawny i jest związany z danymi osobowymi. Jako podmiot, który pozyskał te dane nie mamy prawa ich redystrybucji, gdyż baza ta gromadzi informacje o osobach prywatnych. Natomiast kiedy usunie się dane o nazwie, przy braku przypisanych kategorii PKD, pozostają punkty, które nie posiadają atrybutu identyfikacyjnego funkcji. Nie można więc ich przedstawić w formie zestawienia terytorialnego.

Są tutaj opisane dane punktowe, pozwalające zestawiać liczebności

obiektów; pominięte zaś zostały dane zebrane w projekcie, które służą np. do wyznaczania odległości wg dróg czy pomiarów wielkości obiektów (lasów, jezior, niektórych budynków i obiektów użyteczności publicznej). Charakterystykę zebranego materiału rozpocznę od najwyższego poziomu uogólnienia danych, czyli kategorii grupujących funkcje społeczne. Najwięcej obiektów żywej kultury jest związanych z regulacją potrzeb podstawowych (219 072 obiektów), a najmniej z jakością życia (117 555 obiektów). Ogólna liczba obiektów żywej kultury, które mają precyzyjną charakterystykę funkcji wyniosła 507 691. Ich szczegółowy spis z uwzględnieniem bazy, z której pochodzą prezentuje Rysunek 5.1.

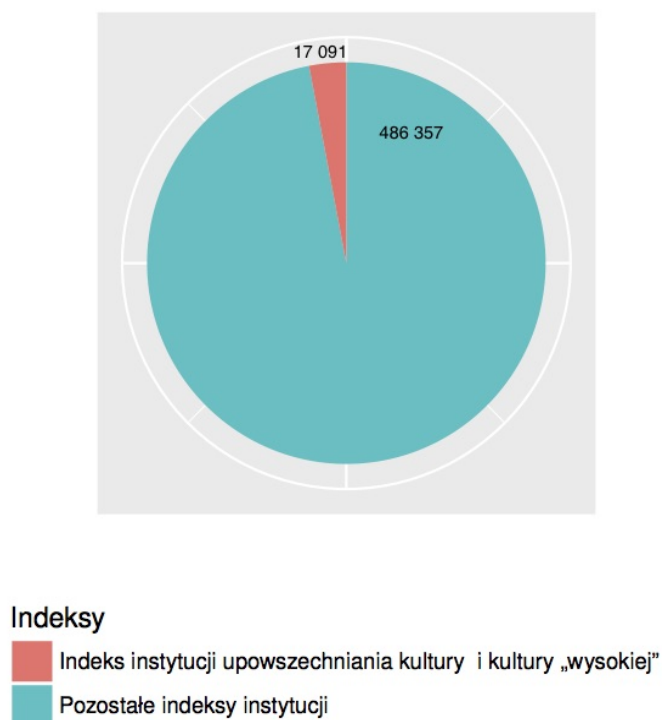
Rysunek 5.1: Obiekty żywej kultury wg funkcji regulującej (n=507 691).



Źródło: Mariusz Piotrowski, opracowanie własne

Dla porównania, w bazie BDL GUS obiektów infrastruktury kultury jest 14 tys. Są to takie obiekty, jak biblioteki, muzea, domy kultury, sale wystawiennicze, teatry itd. Z dwóch baz danych systemu „GIS Kultu-ra” zebrano informację o ponad 17 tys. obiektach o podobnej charakterystyce. W tym kontekście warto przypomnieć, że tzw. wąska definicja kultury stosowana w statystyce GUS odpowiada dwóm indeksom – indeksowi kultury „wysokiej” i indeksowi infrastruktury upowszechniania kultury. Dane te stanowią tylko 3% całości danych zgromadzonych w systemie (Rysunek 5.2).

Rysunek 5.2: Obiekty wąsko rozumianej kultury w bazie GIS Kultura (n=507 691).



Źródło: Mariusz Piotrowski, opracowanie własne

Jeśli weźmie się pod uwagę wymiar przestrzenny, można znaleźć obszary, w których zebrane informacje odbiegają od danych gromadzonych przez GUS. Szczegółowe zestawienie prezentuje Rysunek 5.3. Więcej obiektów w bazie infrastruktury kultury BDL niż udało się nam zebrać występuje w województwach wielkopolskim, kujawsko-pomorskim, opolskim, podlaskim. Możliwa interpretacja takiego stanu rzeczy wskazywałaby na mniejsze zainteresowanie takimi obiektami społeczności openstreetmap mapującej w tych województwach. Natomiast baza „GIS Kultura”, w przypadku województwa dolnośląskiego pokazuje dwukrotnie większą liczbę obiektów niż te, które są w sprawozdawczości GUS. Dla prototypu systemu zbierania informacji, taki wynik można uznać za sukces. Oczywiście jednoznaczne określenie z czego wynikają różnice w liczbie instytucji pomiędzy systemami,

wymaga osobnych badań.¹ Indeksy były projektowane w taki sposób, żeby zbierać szerszą informację niż ta, dostępna dzięki GUS. Różnice na niekorzyść systemu GIS Kultura mogą wynikać z dwóch przyczyn. Po pierwsze, dane o obiektach zostały faktycznie zebrane, lecz nie udało się przypisać właściwego indeksu. W bazie danych ponad 327 tys. obiektów nie posiada przypisanej kategorii. Po drugie, użyte przez nas bazy danych, szczególnie openstreetmap, wykazują koncentrację w stosunku do pewnych obszarów kraju. Jednak patrząc na dynamikę projektu można mieć nadzieję, że ta sytuacja ulegnie poprawie i także te obszary zostaną zmapowane.

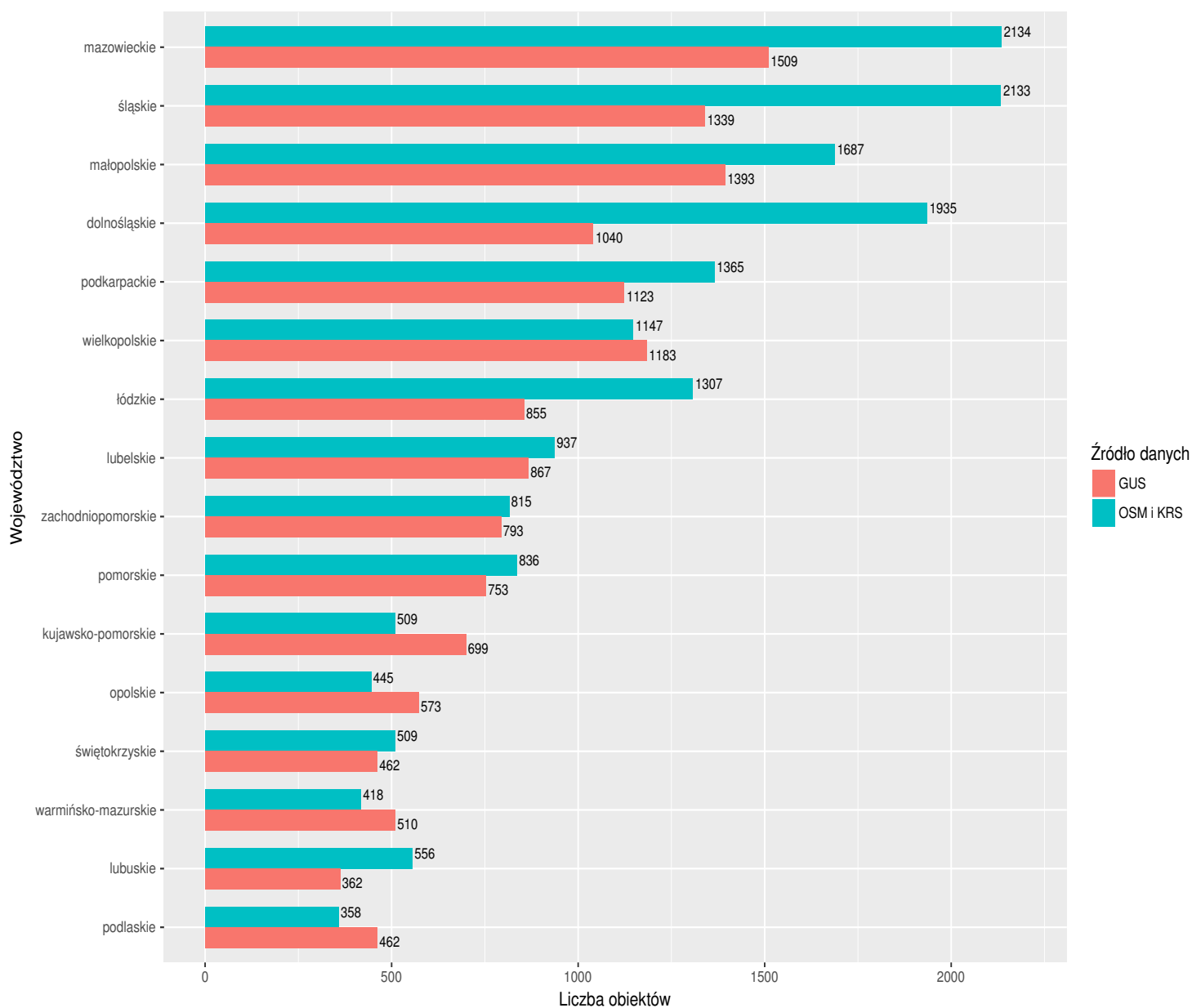
Nie wszystkie indeksy w wyniku automatycznego przetwarzania danych zostały uzupełnione. Stało się tak w przypadku indeksu infrastruktury nieformalnej/subkulturowej oraz infrastruktury administrowania kulturą. Mimo tego dane zgromadzone w systemie są dużo liczniejsze, niż te zbierane przez GUS i udostępniane w BDL. Szczegółowy rozkład liczebności rekordów KRS wg typów indeksów OZK prezentuje Rysunek 5.4. Na wykresie tym przedstawione jest porównanie baz obiektów uzyskanych z KRS i tych z bazy OSM, która jest omawiana w dalszej części niniejszego opracowania.

Prezentację materiałów źródłowych rozpocznę od charakterystyki danych KRS. Ogólna liczba rekordów pozyskanych z KRS (stan na lipiec 2015 roku) wyniosła 401 374 (są to te, które posiadały nazwę). Jednak liczba rekordów, które udało się pomyślnie zgeokodować wyniosła 374 771. Skuteczność geokodowania wynika np. z formatu zapisu adresów w bazie rekordów, w której nie ma żadnego ujednoliconego sposobu tych zapisów. Z całkowitej liczby obiektów, którym można było przypisać współrzędne geograficzne - 223 731 rekordów uzyskało jednoznaczne oznaczenie działalności w bazie PKD i procedury wyznaczania rang. Niewiele ponad 151 tys. rekordów pozostało *pustych*, czyli bez opisu lub z opisem dalece niekompletnym.

Skuteczność działania użytego algorytmu w przypadku organizacji

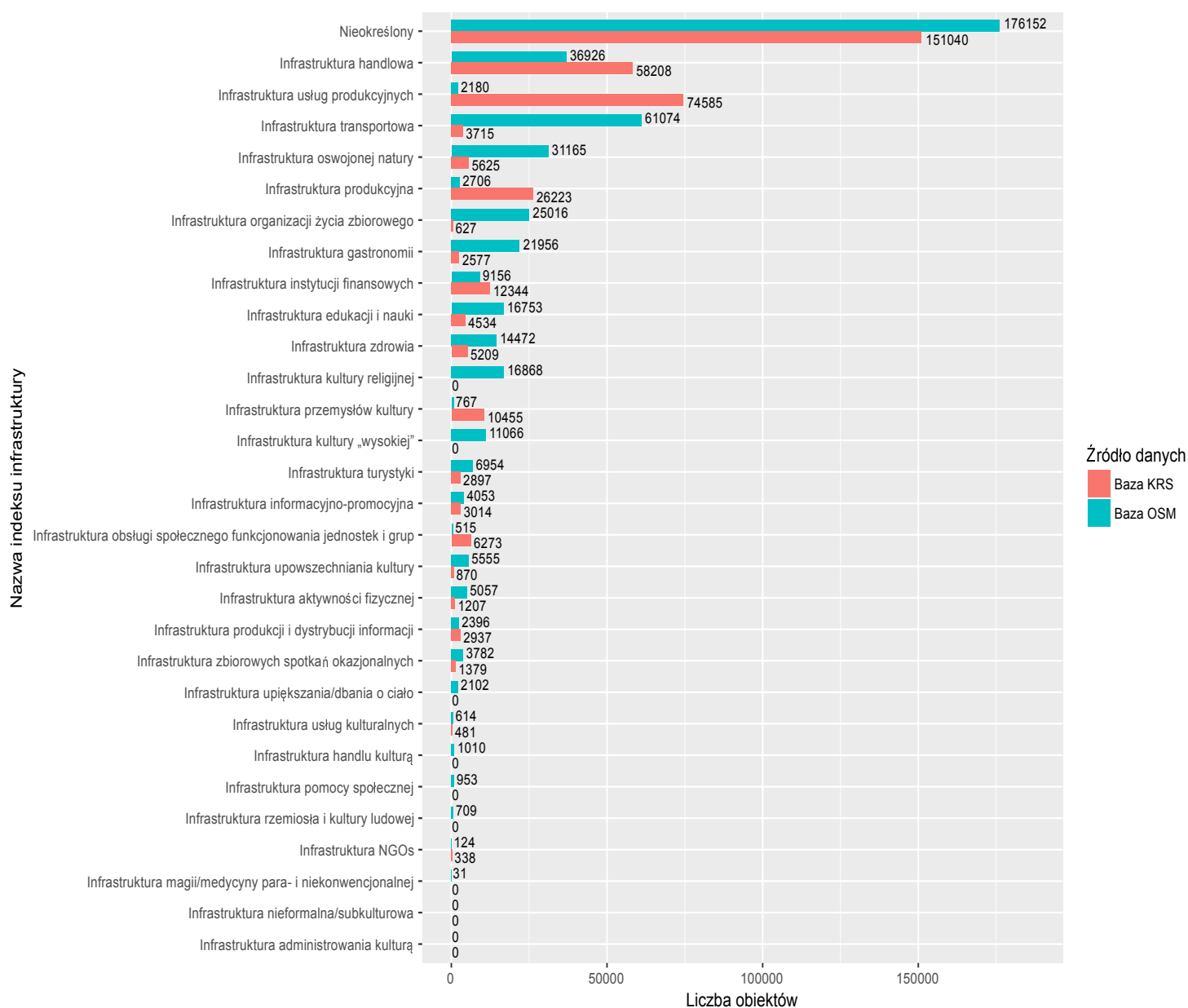
¹Warto przypomnieć, że drastyczne różnice w liczbach obiektów infrastrukturalnych między GUS a innymi statystykami są m.in. dokumentowane na przykładzie muzeów – por. Dorota Folga-Januszewska, Raport o muzeach, 2009, http://www.kongreskultury.pl/title,Raport_o_muzeach,pid,137.html czy teatrów – por. najnowsze wydawnictwo z 2016 roku: <http://www.institut-teatralny.pl/nabytki/teatr-w-polsce-2016/>; przy czym tkwią, jak wiadomo w sposobie definiowania przez GUS tych (i innych) obiektów.

Rysunek 5.3: Indeksy obiektów kultury wg GUS i obiekty żywej kultury wg indeksów upowszechniania kultury i kultury „wysokiej” w województwach.



Źródło: Mariusz Piotrowski, opracowanie własne

Rysunek 5.4: Obiekty żywej kultury wg indeksów (n=834 883).

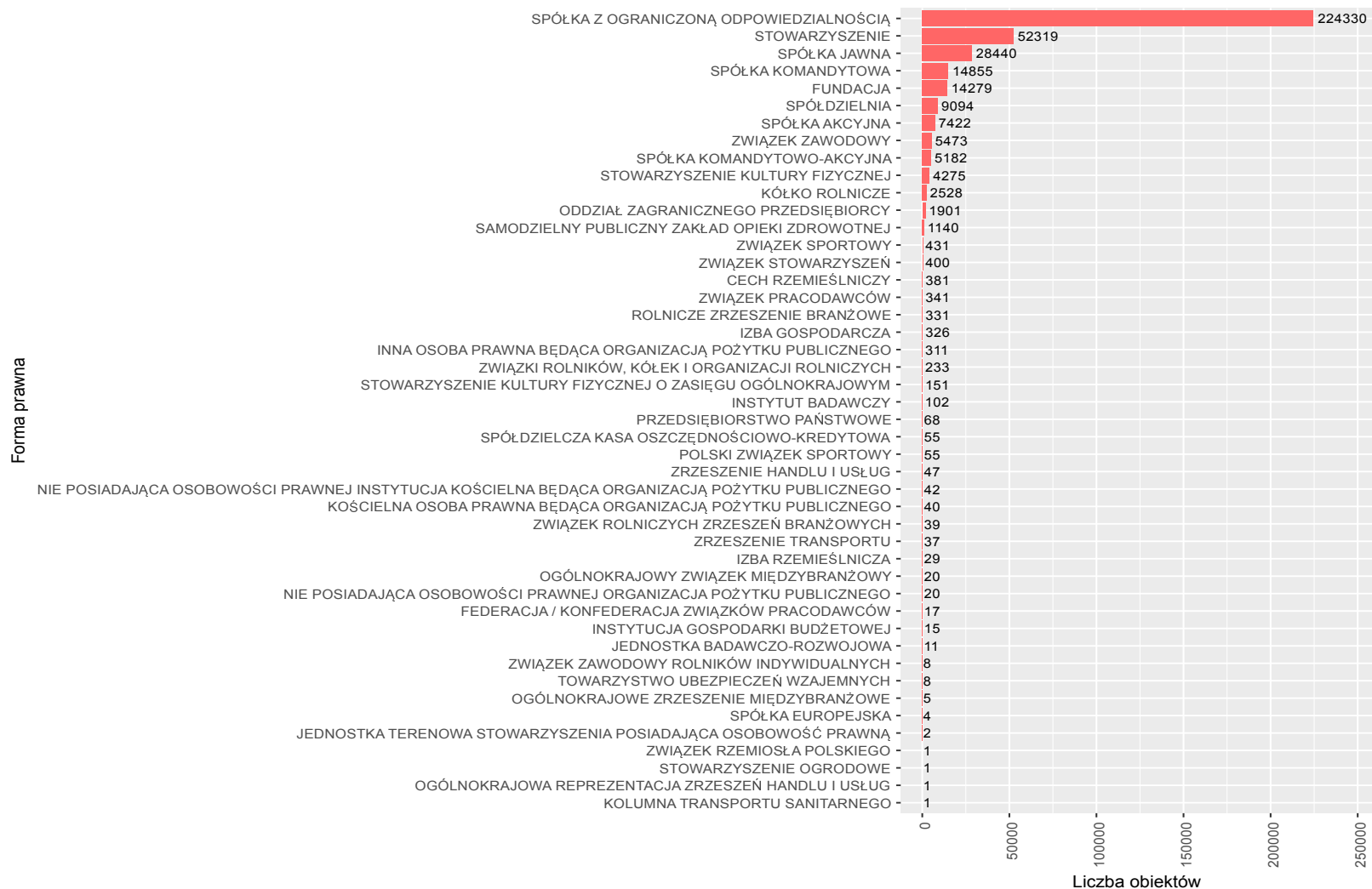


Źródło: Mariusz Piotrowski, opracowanie własne

społecznych i zawodowych była niewielka. Jedynie w przypadku 338 rekordów należących do tej kategorii udało się uzyskać przypisaną funkcję w bazie PKD. Jednak w trakcie analiz możliwe jest przypisanie w bazie typu działalności - indeksu infrastruktury NGO dla ponad 71 tys. rekordów. Redukuje to liczbę nieznanymi obiektów ze 151 tys. do niewielu ponad 80 tys. Masowa obróbka danych, kosztownych do pozyskania, nie pozwoliła na dalsze precyzyjne przyporządkowanie rekordów. Takie działanie powinno więc być kontynuowane w toku dalszych prac, w których należałoby stworzyć infrastrukturę informatyczną do aktualizacji danych z KRS i CEIDG wg typów działalności.

Dodatkową, cenną informacją w wypadku bazy KRS jest możliwość przeprowadzania analiz przestrzennych wg typu organizacyjnego. Można wykorzystać tę zmienną jako zaczątek konstruowania systemu przypisywania szczegółowego indeksu funkcjonalnego, gdy zawiodły inne metody klasyfikacji. Rozkład częstości występowania typów organizacyjnych dla obiektów, którym nie został przypisany indeks infrastruktury, prezentuje Wykres 5. Blisko 45% rekordów w bazie są to spółki z ograniczoną odpowiedzialnością i spółki jawne, które prowadzą bardzo różną działalność, jednak wśród pozostałych 55% rekordów są zarówno stowarzyszenia, fundacje, jak i związki zawodowe i instytuty badawcze. Można traktować je jako elementy wchodzące w skład indeksu infrastruktury NGO. Wówczas liczba elementów nieokreślonych zostanie zmniejszona o blisko połowę. Przechowanie tych danych sprawia, że można się do nich odwołać i poprawiać je ręcznie, w momencie wykonywania analiz na poziomach konkretnych gmin.

Rysunek 5.5: Obiekty z KRS, których funkcji nie określono ze względu na typ prawny (n=151 040).



Źródło: Mariusz Piotrowski, opracowanie własne

Pierwszą możliwą do wykonania analizą przestrzenną jest odpytanie bazy danych wg kategorii - typu infrastruktury żywej kultury (Typ OŻK), ze względu na typ prawny obiektu definiowany w bazie KRS. Wynik takiego zapytania prezentuje Tabela 5.1. Blisko 60 tys. rekordów, to spółki z ograniczoną odpowiedzialnością, którym nie udało się przypisać typu infrastruktury żywej kultury, ale podobna liczba obiektów z tego typu to infrastruktura usług produkcyjnych. Oczywiście wyzwaniem na przyszłość jest zmniejszanie braków w bazie. Warto jednak pamiętać, że nasz projekt jest prototypowy i w skali blisko miliona unikatowych rekordów, tego typu braki są nieuniknione. Można więc mówić tu raczej o szacunkach, niż o pewności wyników.

Tabela 5.1: Typ Obiektu Żywej Kultury ze względu na typ obiektu KRS - 10 najbardziej licznych.

Typ OZK	Typ obiektu KRS	Procent	Liczba
Infrastruktura usług produkcyjnych	SPÓŁKA Z OGRANICZONĄ ODPOWIEDZIALNOŚCIĄ	15,823	59252
None	SPÓŁKA Z OGRANICZONĄ ODPOWIEDZIALNOŚCIĄ	15,767	59044
None	STOWARZYSZENIE	13,248	49610
Infrastruktura handlowa	SPÓŁKA Z OGRANICZONĄ ODPOWIEDZIALNOŚCIĄ	11	41190
Infrastruktura produkcyjna	SPÓŁKA Z OGRANICZONĄ ODPOWIEDZIALNOŚCIĄ	5,405	20240
None	FUNDACJA	2,794	10463
None	SPÓŁKA JAWNA	2,349	8796
Infrastruktura instytucji finansowych	SPÓŁKA Z OGRANICZONĄ ODPOWIEDZIALNOŚCIĄ	2,253	8435
Infrastruktura przemysł kultury	SPÓŁKA Z OGRANICZONĄ ODPOWIEDZIALNOŚCIĄ	1,936	7250
Infrastruktura handlowa	SPÓŁKA JAWNA	1,869	6999
None	ZWIĄZEK ZAWODOWY	1,452	5437

Źródło: Mariusz Piotrowski, opracowanie własne

Tabela 5.2: Typ Obiektu Żywej Kultury ze względu na typ obiektu KRS wg. województw - 10 z największą liczebnością.

Typ obiektu KRS	Typ OZK	województwo	Procent	Liczba
SPÓŁKA Z OGRANICZONĄ ODPOWIEDZIALNOŚCIĄ	Infrastruktura transportowa	mazowieckie	0,264	988
		śląskie	0,074	278
		wielkopolskie	0,071	266
		dolnośląskie	0,057	215
		pomorskie	0,053	200
		małopolskie	0,053	200
		łódzkie	0,035	132
		lubuskie	0,031	116
		podkarpackie	0,03	112
		kujawsko-pomorskie	0,022	81
SPÓŁKA KOMANDYTOWA	Infrastruktura transportowa	mazowieckie	0,018	66

Źródło: Mariusz Piotrowski, opracowanie własne

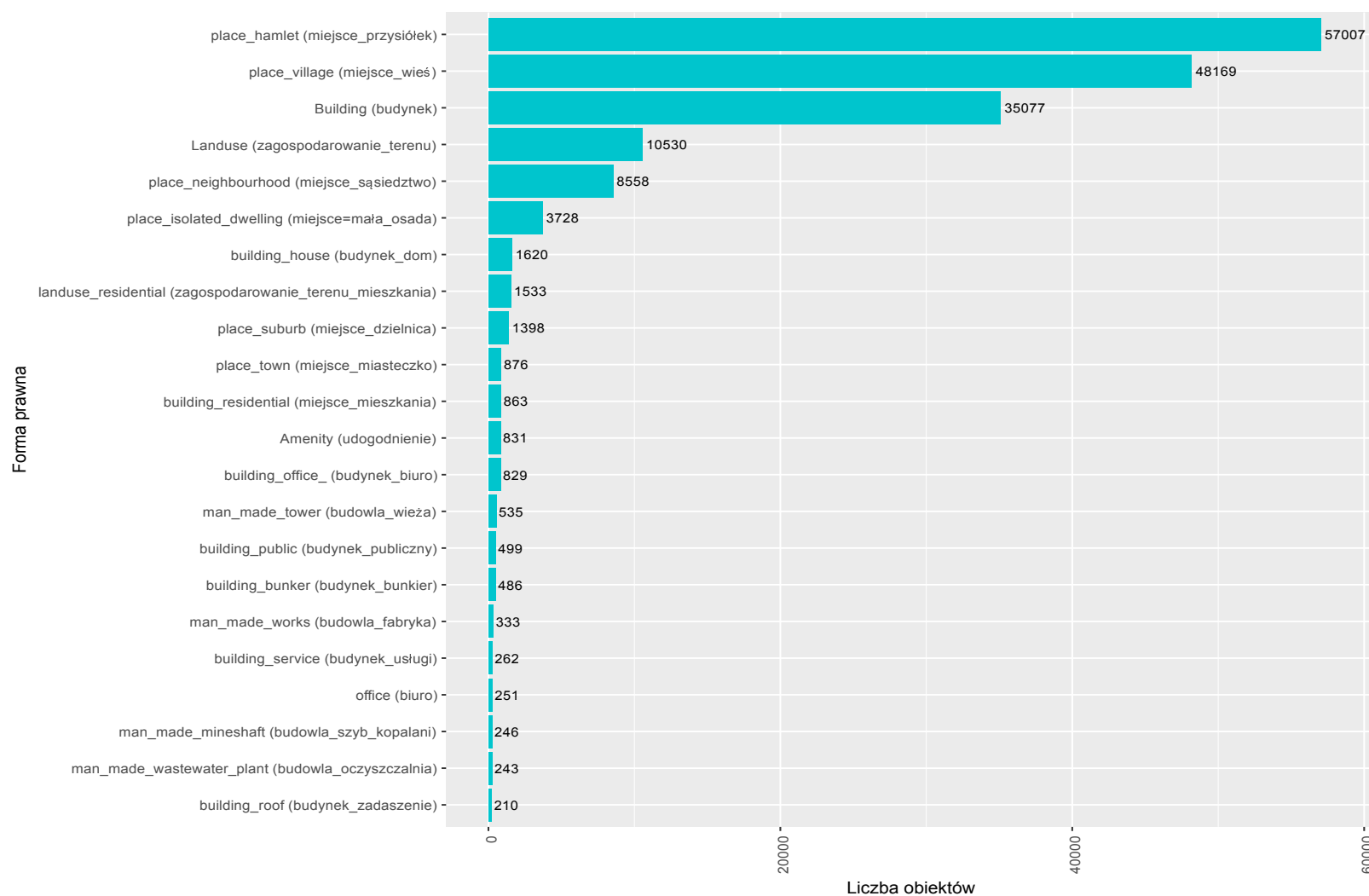
Konstrukcja bazy danych, nawet przed procesem harmonizacji, pozwala na wykonywanie interesujących zestawień o charakterze przestrzennym. Możliwe jest określenie typów obiektów według typów infrastruktury żywej kultury z rozmieszczeniem w województwie, powiecie, gminie. Co więcej, dane o precyzji co do punktu w przestrzeni, pozwalają swobodnie kształtować granice przestrzenne. Zestawienie wg województw przedstawia Tabela 5.2. Liczebności typów infrastrukturalnych ze względu na typ prawny stają się niewielkie w rozkładzie wojewódzkim i nie przekraczają 1000 obiektów. Świadczyć to może o zróżnicowanej strukturze pozyskanych danych.

Drugie serce systemu - baza danych openstreetmap (OSM), składa

się z 460 170 wyodrębnionych obiektów¹. Jest to wynik uzyskany po zastosowaniu pliku słownika na bazie najczęściej pojawiających się kategorii tagów. Został on skonstruowany w oparciu o bazę systemu <https://taginfo.openstreetmap.org>, który pozwala na analizy relacji między tagami oraz na analizy frekwencyjne. Słownik składa się z 713 kategorii (tagów). W efekcie ostatniej operacji wyodrębniania uzyskane zostały 563 tagi, które znalazły się w obrębie Polski co najmniej jeden raz. Rysunek 5.3 na stronie 32. Bliżko 22% przyporządkowanych indeksów infrastruktury dotyczy transportu i handlu. Można to wyjaśnić charakterem projektu społecznościowego, mapy są potrzebne i wykorzystywane w komunikacji, w planowaniu podróży – długich, jak i krótkich (np. drogi do pracy). W projekcie mapowane są przystanki komunikacji publicznej, ale też stacje paliw. Infrastruktura handlowa to miejsca robienia zakupów, począwszy od sklepów osiedlowych po centra handlowe.

Wyjaśnienia wymaga ponownie kwestia dominacji kategorii *Nieokreślone*. Jest to kategoria robocza, w której znajdują się dane nie przyporządkowane do żadnego z indeksów obiektów żywej kultury. Jak już wcześniej zaznaczyłem, są to zarówno błędy w klasyfikacji, literówki w nazwach tagów, ale też zgromadzone nazwy własnych obiektów geograficznych (tag – *place*). Te ostatnie są wykorzystywane w celu generowania map zasadniczych. Rozkład liczebności elementów nieklasyfikowanych wg kategorii żywej kultury prezentuje Rysunek 5.6.

Rysunek 5.6: Obiekty z OSM, których nie zostały przypisane do indeksów obiektów żywej kultury ze względu na charakterystykę tagów (n=176 152).



Źródło: Mariusz Piotrowski, opracowanie własne

Podczas analizy samych tagów z projektu openstreetmap (Tabela 5.3), najczęściej pojawiały te należące do klucza (*key*) place o wartościach (*value*) *hamlet* i *village*. Zgodnie z konsensusem projektu openstreetmap – *hamlet* to sposób oznaczania osad zamieszkiwanych przez mniej niż 200 osób, *village* to zaś miejscowości z ludnością między 1000 a 10000 mieszkańców.

Tabela 5.3: Częstości występowania obiektów z bazy OSM ze względu na typ indeksu obiektu żywej kultury – 10 najczęściej występujących.

Nazwa tagu openstreetmap	Nazwa indeksu OZK	Procent	Liczba
place_hamlet	Nieokreślona	12,388	57007
place_village		10,468	48169
Building		7,623	35077
stop_position_	Infrastruktura transportowa	4,806	22117
amenity_place_of_worship	Infrastruktura kultury religijnej	3,111	14315
Natural	Infrastruktura oswojonej natury	2,733	12578
Landuse	Nieokreślona	2,288	10530
amenity_restaurant	Infrastruktura gastronomii	2,119	9751
amenity_school	Infrastruktura edukacji i nauki	2,104	9683
place_neighbourhood	Nieokreślona	1,86	8558
shop_convenience	Infrastruktura handlowa	1,729	7956

Źródło: Mariusz Piotrowski, opracowanie własne

Trzeci tag w kolejności jest kategorią ogólną. Są to obiekty, które posiadają klucz (*key*) – *building*, natomiast wartość (*value*), jest pusta - z błędem literowym, lub nie należy do elementu zbieranego przez słownik przypisujący do obiektów (np. wartości – *other*). Podobnie ma się sytuacja z kluczem (*key*) - *landuse*.

Na podstawie tych dwóch baz danych widać jak różne obszary one pokrywają. Baza KRS jest bogatym źródłem wiedzy o przemyśle i organizacjach społecznych, baza openstreetmap - o handlu i infrastrukturze transportowej. Użycie tych baz i zaprezentowanych narzędzi może stanowić prototyp systemu aktualizacji danych o życiu społecznym w kraju. Dane są ogólnodostępne i pozwalają na automatyzację procesu określania funkcji dla innych systemów.

Rozdział 6

Funkcjonalność systemu.

Kierunki analizy danych.

Sercem systemu „GIS Kultura” jest baza danych, która umożliwia wykonywanie operacji analitycznych i wizualizację danych na mapach. W części tej zostaną zaprezentowane możliwości pracy na danych przestrzennych, które pozwalają na połączenie różnych podejść metodologicznych w analizach kultury. Analizy te można podzielić na *implicite* i *explicite* przestrzenne. Poniżej przedstawiam różne operacje dokonane w ramach systemu, ale nie są to jeszcze pełne (kompletne, dokładne) badania stanu kultury. W większości tych analiz pod uwagę była brana ograniczona liczba zmiennych.

6.1 Analizy *implicite* przestrzenne.

6.1.1 Wydatki na kulturę ze względu na typ historyczny. Ze-stawienia.

We wcześniejszej części raportu została przedstawiona mapa historyczna Polski (Rysunek 4.5 na stronie 25). Dane umożliwiające jej stworzenie pozwalają np. poszukać odpowiedzi na takie pytania jak: czy wydatki na kulturę z budżetów gmin (pozycja budżetu 921) są podobne w różnych regionach,

czy może się różnić? W tym celu do analizy zostały pobrane dane z Banku Danych Lokalnych o wydatkach gmin w pozycji 921. Następnie zostały one połączone z bazą typów historycznych gmin. Poniżej opisanych jest kilka rodzajów analiz, które przeprowadziliśmy. W pierwszej kolejności określone zostały sumaryczne wydatki na kulturę w zależności od typu historycznego. Wyniki przedstawia Tabela 6.1.

Tabela 6.1: Suma wydatków na kulturę ze względu na typ historyczny (dane na rok 2015).

Typ historyczny	Suma wydatków
Zabór rosyjski	1 937 649 922 zł
Ziemie odzyskane	1 752 195 877 zł
Zabór pruski	999 880 275 zł
Zabór austriacki	721 460 038 zł

Źródło: Mariusz Piotrowski, opracowanie własne

Największe sumaryczne wydatki na kulturę są w gminach, które znajdowały się w obszarze dawnego zaboru rosyjskiego, ale na drugim miejscu pojawił się tu obszar, na którym nie było – jak widać zakorzenionych tradycji. Te trzeba było dopiero wytworzyć.¹ Można też sprawdzić, czy ta sytuacja ulega zmianie jeśli w analizach pominie się Warszawę. Takie zestawienie prezentuje tabela 6.2.

Tabela 6.2: Wydatki na kulturę ze względu na typ historyczny (dane na rok 2015), bez M. st. Warszawa.

Typ historyczny	Suma wydatków
Ziemie odzyskane	1 752 195 877,00 zł
Zabór rosyjski	1 496 260 933,00 zł
Zabór pruski	999 880 275,00 zł
Zabór austriacki	721 460 038,00 zł

Źródło: Mariusz Piotrowski, opracowanie własne

W dalszej kolejności można przeprowadzić analizy, wyodrębniając

¹Por. Elżbieta Berendt (red.), *Mom jo skarb. Dolnośląskie tradycje w procesie przemian*, Wrocław: Wyd. Muzeum Etnograficzne we Wrocławiu, 2009.

typy gmin z zestawienia. Za pomocą tradycyjnych metod łatwo uzyskać wyniki z jednego typu gminy. System pozwala z kolei na elastyczne zadawanie pytań, które pomijają pewne wartości. W efekcie można wskazać wszystkie gminy np. bez typu gmin wiejskich czy miejskich. (Tabela 6.3).

Tabela 6.3: Wydatki na kulturę ze względu na typ historyczny (dane na rok 2015), wg typów gmin.

Typ historyczny	Bez gmin miejskich	Bez gmin wiejskich	Bez gmin miejsko-wiejskich
Zabór rosyjski	749 370 342,00 zł	1 480 807 464,00 zł	1 645 122 038,00 zł
Ziemie odzyskane	679 109 607,00 zł	1 460 320 704,00 zł	1 364 961 443,00 zł
Zabór pruski	428 246 012,00 zł	800 983 739,00 zł	770 530 799,00 zł
Zabór austriacki	341 192 372,00 zł	521 195 394,00 zł	580 532 310,00 zł

Źródło: Mariusz Piotrowski, opracowanie własne

Dzięki zastosowaniu mechanizmu baz danych relacyjnych, używając składni języka SQL do analiz, można przyspieszyć proces weryfikacji hipotez. Wreszcie analizując dwie zmienne, typ historyczny gminy i wydatki na kulturę, można rozszerzyć część analityczną o nową zmienną – liczbę ludności. Dzięki temu możliwe jest określenie wydatków na kulturę na mieszkańca ze względu na typ historyczny. Zestawienie to prezentuje Tabela 6.4.

Tabela 6.4: Wydatki na kulturę na mieszkańca ze względu na typ historyczny (dane na rok 2015).

Typ historyczny	Suma wydatków
Ziemie odzyskane	142,25 zł
Zabór pruski	117,78 zł
Zabór austriacki	99,81 zł
Zabór rosyjski	99,35 zł

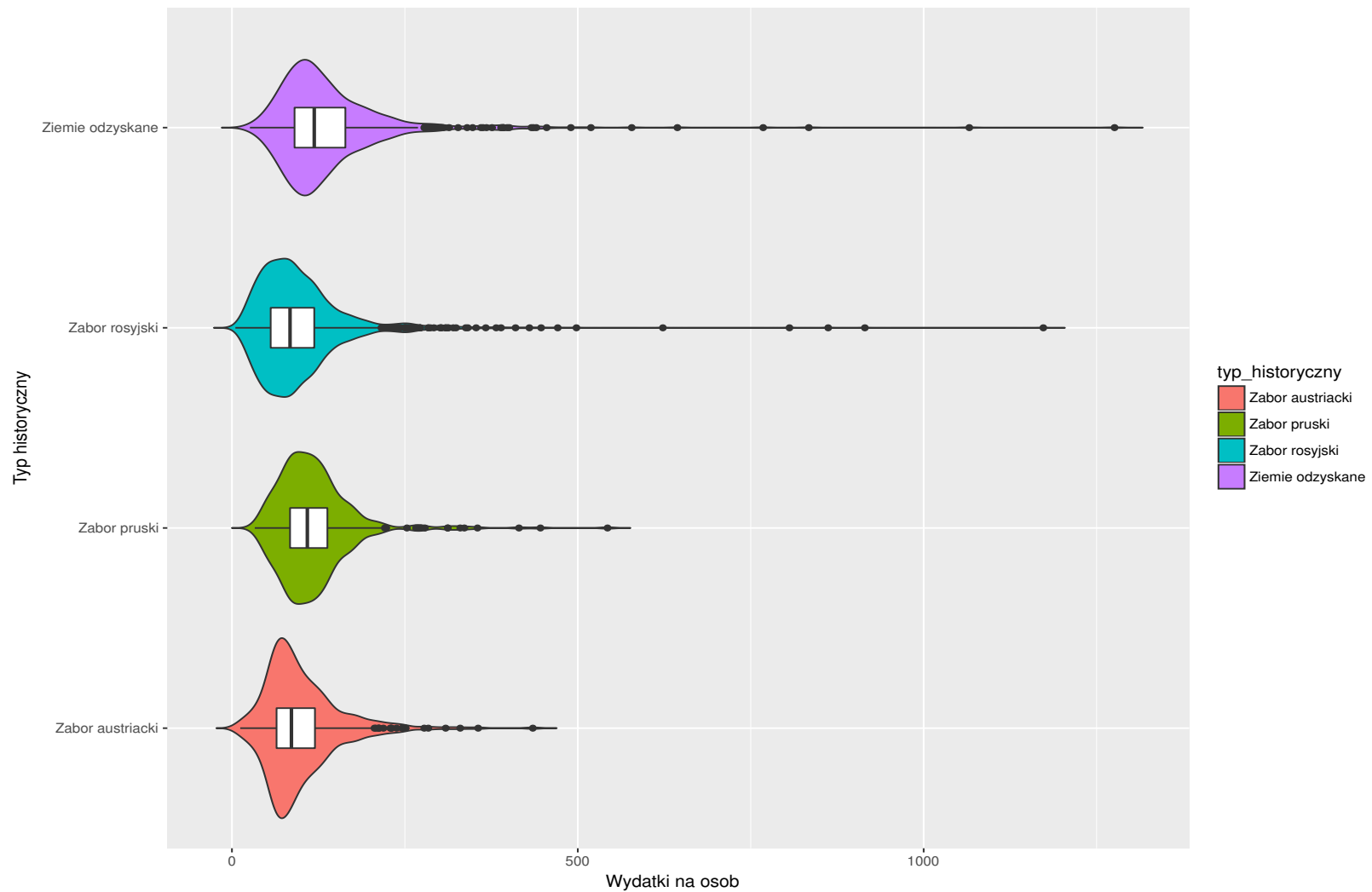
Źródło: Mariusz Piotrowski, opracowanie własne

Ten ostatni wskaźnik można uznać za najbardziej miarodajny w formułowaniu wniosków o rozproszeniu wydatków na kulturę ze względu na parametr przestrzenny. Najmniejsze średnie wydatki na kulturę na mieszkańca są w gminach dawnego zaboru rosyjskiego. Średnia arytmetyczna wyliczana jako wskaźnik dla gmin może być obarczona błędami wynikającymi z dużego rozproszenia danych. Dlatego też za miarę centralną zmiennej warto

obrać medianę. Jednak w przypadku danych o wydatkach na kulturę, różnice pomiędzy średnią arytmetyczną są niewielkie, co przedstawione jest na Rysunku 6.1. Najmniejsze wydatki na kulturę będą w zaborze austriackim (99 zł), o złotówkę wyższe zaś w gminach dawnego zaboru rosyjskiego. Do prezentacji danych wykorzystane zostały dwa wykresy nałożone na siebie: *wykres skrzypcowy*, którego kształt przypomina odbity w lustrze histogram², oraz *wykres pudełkowy*, z zaznaczoną medianą dla każdego z typów historycznych. Olbrzymią zaletą takiego zaprezentowania danych jest możliwość uwzględnienia wartości odstających.

²Dane przekształcone są w oparciu o wskaźnik kernel density estimation – za R. Poniat, O wykorzystaniu wykresów pudełkowych w: „Przyszłość demograficzna Polski” (34) 2014, ss. 103-120

Rysunek 6.1: Wydatki na kulturę ze względu na typ historyczny w gminach.



Źródło: Mariusz Piotrowski, opracowanie własne

6.1.2 Wydatki na kulturę. Wizualizacja danych.

System „GIS Kultura” to także elastyczny mechanizm pracy z danymi. Efekty analiz mogą być dostępne w formie zestawień tabelarycznych, jednak wielokrotnie dane prezentowane w formie map przestrzennych pozwalają na lepszą organizację danych: pozostając przy problematyce wydatków na kulturę z budżetów gmin, można zaprezentować problem, który względnie łatwo przedstawia się przy użyciu map.

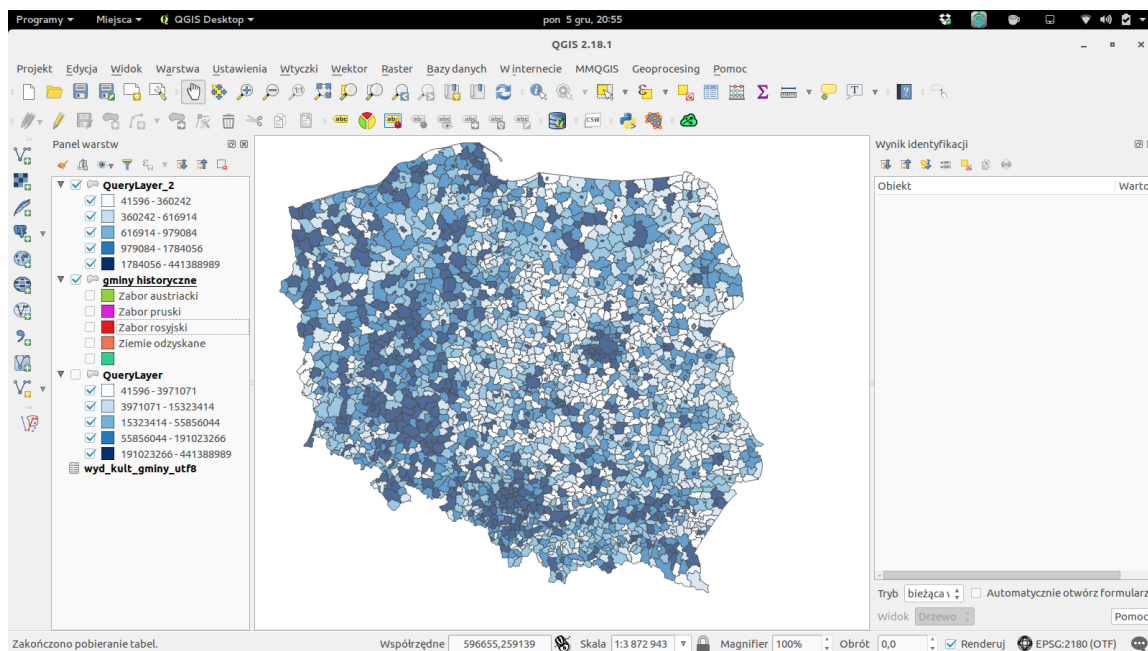
Wizualizacja wydatków budżetowych na poziomie gminy w postaci mapy musi podlegać pewnym generalizacjom. W innym przypadku byłoby to 2478 gmin, które należałoby wyróżnić bądź kolorem, bądź dodając jakiś kształt. Liczba informacji na obrazie byłaby zbyt duża do przetworzenia dla osoby oglądającej wizualizację. Dlatego też, aby zredukować złożoność modelu warto zastosować jakąś formę tworzenia przedziałów. Jednym ze sposobów jest metoda równolicznych przedziałów (metoda kwantyli). Rysunek 6.2 przedstawia użycie metody kwantyli do zaprezentowania wydatków gmin na kulturę, przy użyciu oprogramowania QGIS.

Dzięki użyciu koloru w jednolitej tonacji (jasne odcienie pokazują niskie wartości, ciemne - wysokie wartości) dostrzec można obszary koncentracji wysokich i niskich nakładów. Dalej - można nałożyć kolejny typ mapy i sprawdzić czy pewne dane się wizualnie pokrywają. Rysunek 6.3 pokazuje nałożenie zakresu granic podziałów historycznych na warstwy z mapą rozkładów wydatków.

Warto tu jednak zaznaczyć, że celem tego raportu nie jest rozstrzyganie czy podziały historyczne rzutują na wydatki na kulturę. Aby móc odpowiedzieć na to pytanie należałoby przeprowadzić o wiele szersze badanie, w którym należałoby skonstruować właściwą aparaturę pomiarową, uwzględniającą także wiele innych danych.

Przykładowo tu przywołana tematyka wydatków na kulturę, może być prezentowana dzięki jeszcze jednemu sposobowi pracy z danymi przestrzennymi. Praca analityczna z danymi jest wielokrotnie pracą przy użyciu podstawowych operacji arytmetycznych. Oparcie bazy „GIS Kultura” na

Rysunek 6.2: Użycie metody kwantyli do zaprezentowania wydatków gmin na kulturę, przy użyciu oprogramowania QGIS.



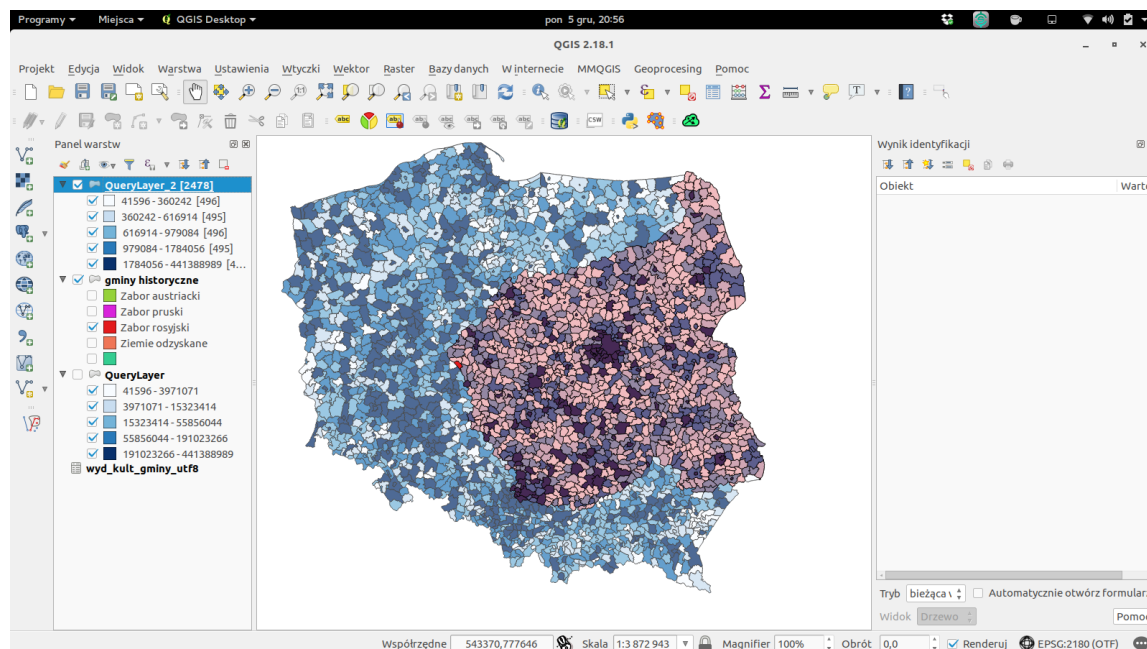
Źródło: Mariusz Piotrowski, opracowanie własne

systemie bazy Postgresql umożliwia wykorzystanie bezpośrednio w jednym narzędziu dostępu do operacji matematycznych. Na wskaźnikach można wykonywać operacje dodawania, odejmowania, mnożenia, dzielenia, ale także inne z obszaru statystyk opisowych.

Pytania, które zostały tu zadane jako następne dotyczą dość intrygującej kwestii: czy istnieją gminy, które zarabiają na kulturze? A jeśli tak, to które to są gminy? W tym celu zgromadziliśmy dane o dochodach w pozycji 921 i od nich zostały odjęte wydatki na kulturę i dziedzictwo narodowe. Zapytanie ograniczone są tutaj w taki sposób, żeby pokazały się tylko wartości większe niż 0 zł. W efekcie otrzymaliśmy listę 51 gmin, które - wg GUS - wykazują zyski w tej pozycji. Listę gmin prezentuje Tabela 6.4.³ na stronie 43.

³Wartość (2) przy nazwach gmin oznacza gminy wiejskie wg kodów TERYT. Taką formę zapisu stosuje GUS w bazie BDL.

Rysunek 6.3: Nałożenie zakresu granic podziałów historycznych na warstwy z mapą rozkładów wydatków.



Źródło: Mariusz Piotrowski, opracowanie własne

Zamiast zaprezentować listę składającą się z 51 elementów, warto ponownie zastosować którąś z metod klasyfikacji. W przypadku, kiedy rozstęp między wartościami jest znaczny (tu wynosi np. 1.892.417 zł) i rozkład nie odbiega od normalnego, albo celem miałyby być wyodrębnienie jednostek dla dalszych badań, można skorzystać z klasyfikacji Jenksa - metody naturalnych przerw⁴ Następną czynnością jest zaprezentowanie tych danych na mapie. W tym celu ponownie użyte zostało oprogramowanie QGIS. Efekt analizy prezentuje Rysunek 6.4 na stronie 50. Przedziały zostały ustalone w taki sposób, żeby podkreślić te, które są wartościami ekstremalnymi; odczytując dane z mapy, można dostrzec brak gmin w województwie warmińsko-mazurskim oraz dużą koncentrację gmin w województwie łódzkim.

⁴ Jest to jedna z metod analizy skupień, w której redukuje się wariancję w obrębie klas, i zwiększa się wariancję pomiędzy klasami.

Tabela 6.5: Gminy, które wykazały dodatnie saldo w pozycji 921, 5 gmin z najwyższymi kwotami.

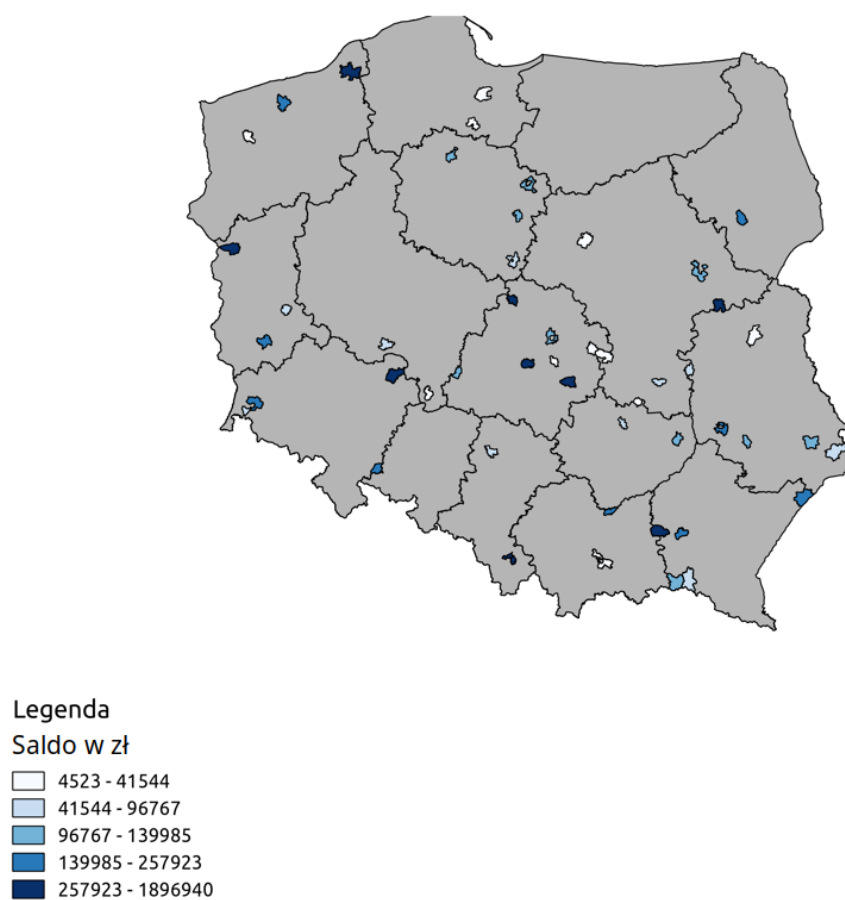
Gmina	województwo	saldo	dochody	wydatki
Krośnice (2)	dolnośląskie	1 896 940,09 zł	12334599,82	10437659,73
Łodygowice (2)	śląskie	1 260 009,36 zł	2553198,82	1293189,46
Wilkowice (2)	śląskie	768 804,73 zł	2538069,61	1769264,88
Malechowo (2)	zachodniopomorskie	686 577,77 zł	1437842,85	751265,08
Daszyna (2)	łódzkie	641 499,64 zł	835808,67	194309,03

Źródło: Mariusz Piotrowski, opracowanie własne

Dane, które posiadają jakąś formę identyfikacji terytorialnej można opracować na wiele sposobów. Dotychczas zostały zaprezentowane te, które za punkt wyjścia przyjmują hipotezę przestrzenną i te, które przyjmują hipotezę o wielkości wartości. W wypadku badań implicite przestrzennych, najbardziej podstawowe analizy są związane z klasyfikacją danych. W zależności od rodzaju analiz wykorzystuje się różne sposoby. Rysunki 6.5 i 6.6 prezentuje dwie kontrastujące metody.

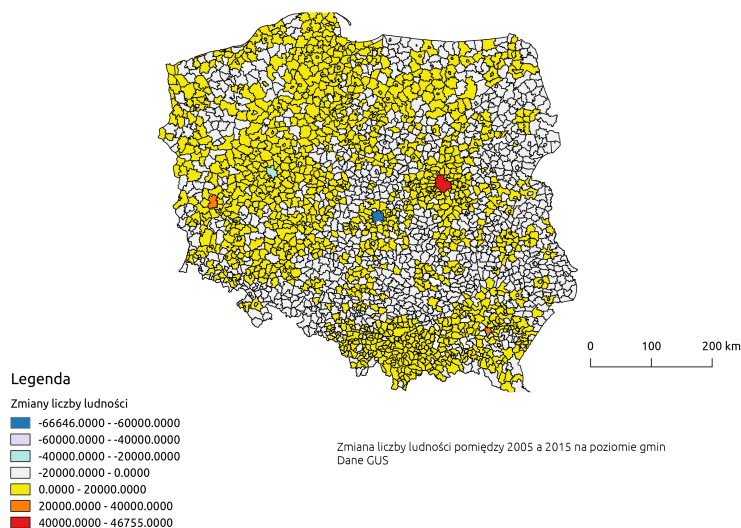
W obu przypadkach prezentowane są dane o migracjach mieszkańców gmin pomiędzy rokiem 2005 a 2015. Rysunek 6.5 prezentuje dane w oparciu o metodę klasyfikacji wg równych przedziałów. Wartości ujemne i dodatnie wyróżnione są w oparciu o ciepłą i zimną gamę kolorystyczną. Gamy zimne oznaczają wartości poniżej 0, gamy ciepłe oznaczają wartości powyżej 0. Na Rysunku 6.6 klasy są stworzone metodą kwantyli – wydzielono 8 równolicznych przedziałów. Sposób prezentacji oparty jest o jednolitą gamę kolorystyczną, gdzie początkowe wartości mają najmniejsze nasycenie barwy. Wyraźnie dostrzec można tutaj tendencje suburbanizacji wokół pewnych miast np. Poznania, Wrocławia, Łodzi.

Rysunek 6.4: Gminy, które wykazały dodatnie saldo w pozycji 921, metoda klasyfikacji - kwantyle.



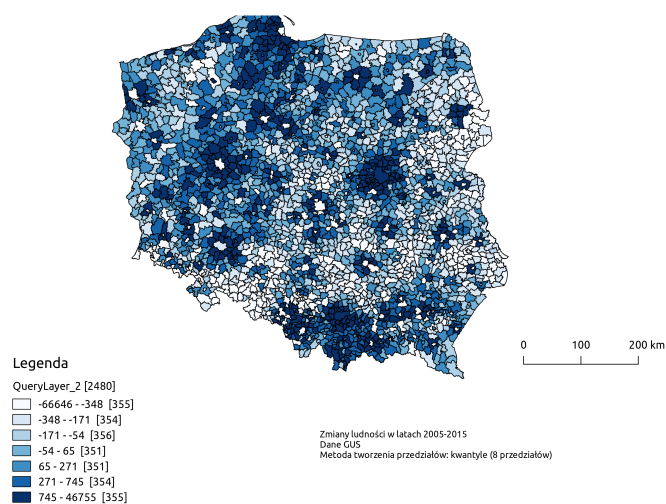
Źródło: Mariusz Piotrowski, opracowanie własne

Rysunek 6.5: Migracje ludności między gminami w Polsce pomiędzy rokiem 2005 a 2015. Przedziały tworzone metodą równych podziałów.



Źródło: Mariusz Piotrowski, opracowanie własne

Rysunek 6.6: Migracje ludności między gminami w Polsce pomiędzy rokiem 2005 a 2015. Przedziały tworzone metodą kwantylową.



Źródło: Mariusz Piotrowski, opracowanie własne

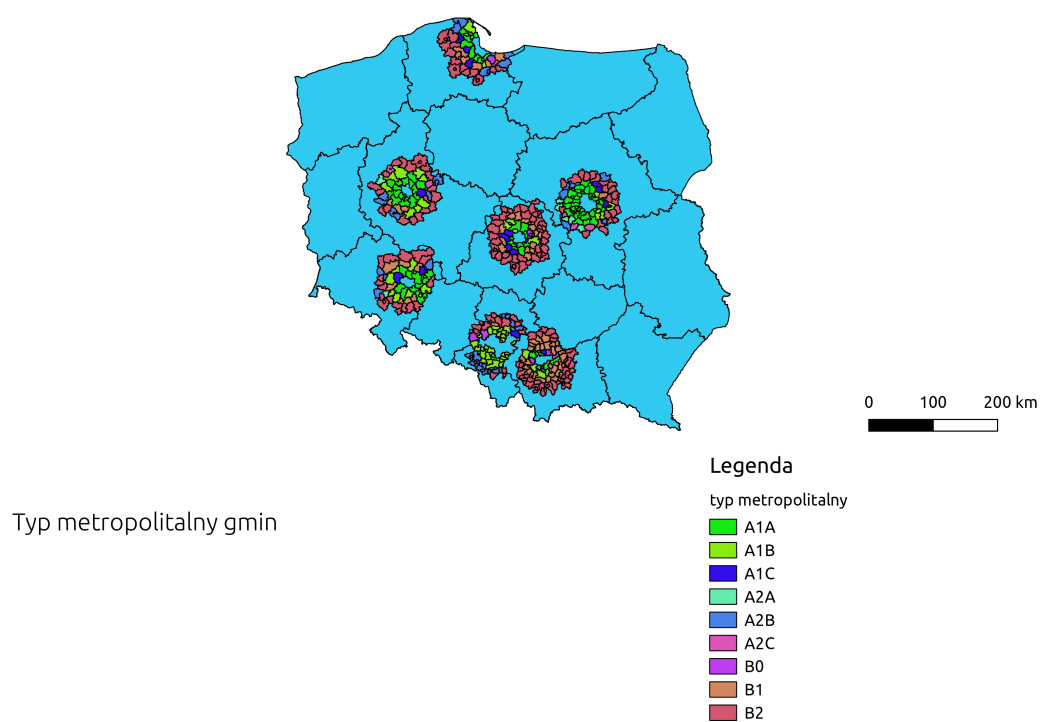
Analiza danych przestrzennych pozwala na zaobserwowanie wzorców, które w samej analizie tabelarycznej byłby trudne do uchwycenia. Przypadek dotyczący migracji ludności naprowadza na poszukiwanie tendencji rozwojowych, wokół konkretnych obszarów miejskich. Pewną propozycją analityczną może być raport Macieja Smetkowskiego, Bohdana Jałowickiego, Grzegorza Gorzelaka pt. „Obszary metropolitalne w Polsce: problemy rozwojowe i delimitacja”⁵.

W celu analizy wydatków na kulturę w obszarach metropolitalnych należało w pierwszym kroku stworzyć i zaktualizować bazę danych o gminach z 2008 roku. Dzięki temu możliwe było zwizualizowanie tych danych (Rysunek 6.6). Pytanie o wielkość wydatków na kulturę i dziedzictwo narodowe jest związane z pytaniem do bazy tylko o te gminy. Wydatki sumaryczne gmin z typu A1 najbliższego ośrodkowi metropolitalnemu i B2 znajdującego się w najdalszej strefie wpływu przedstawia Tabela 6.7.

Zestawienie danych może stanowić przyczynek do dalszych analiz procesów suburbanizacji. Szczególnie interesująca w tym kontekście może być okrug łódzki, w którym relacja pomiędzy gminami z typu A1 do B2 wskazuje na znaczny udział w metropolii słabo zintegrowanych gmin. Dodatkowo te najdalej wysunięte na wschód niemal sąsiadują z metropolią warszawską. Bardzo podobna sytuacja jest w Konurbacji Śląskiej, gminy typu B2 graniczą z metropolią krakowską. Spoglądając na politykę kulturalną przez pryzmat relacji centrum-peryferie na tych obszarach pomija się więc wyraźną autonomię tych gmin. Ujmowane *en bloc* ujawniają siłę słabych gmin w tych metropoliach, lub też słabość centrów tych metropolii.

⁵http://www.euroreg.uw.edu.pl/dane/web_euroreg_publications_files/602/obszary_metropolitalne_w_polsce_problemy_rozwojowe_i_delimitacja.pdf

Rysunek 6.7: Gminy wg typu metropolitalnego.



Źródło: Mariusz Piotrowski, opracowanie własne

Tabela 6.6: Wydatki sumaryczne gmin z typu A1 najbliższego ośrodkowi metropolitalnemu i B2 znajdującego się w najdalszej strefie wpływu.

Nazwa ośrodka metropolitalnego	Typ metropolitalny	Suma wydatków na kulturę
Warszawa	A1A	98 946 759,00 zł
	B2	12 323 772,00 zł
Poznań	A1A	46 078 855,00 zł
	B2	21 725 798,00 zł
Wrocław	A1A	24 383 639,00 zł
	B2	42 993 725,00 zł
Trójmiasto	A1A	20 493 920,00 zł
	B2	25 023 636,00 zł
Kraków	A1A	16 534 151,00 zł
	B2	33 905 353,00 zł
Łódź	A1A	9 228 513,00 zł
	B2	50 822 263,00 zł
Konurbacja Śląska	A1A	2 304 228,00 zł
	B2	22 768 769,00 zł

Źródło: Mariusz Piotrowski, opracowanie własne

6.2 Analizy explicite przestrzenne.

Możliwości analityczne systemu zaprezentowane w poprzedniej części raportu bazują na identyfikacji przestrzennej zgodnej z kodami TERC – systemu Teryt. Analizy są realizowane do poziomu gminy. Jednak pytanie przestrzenne, które jest najbardziej podstawowe, dotyczy odległości wyrażonej w jednostce miary, w metrach, bądź kilometrach. Dla pytania jak wygląda otoczenie kulturalne, bardziej pożądaną i przystającą do sposobu myślenia odpowiedzią z punktu widzenia uczestnika wydarzeń kulturalnych jest odpowiedź, która pokaże obiekty w odległości 1 km od zdefiniowanego punktu, niż odpowiedź, która będzie zawierała obiekty w granicach administracyjnych. W tej części niniejszego opracowania zostaną zaprezentowane możliwości, jakie daje system „GIS Kultura” w poszukiwaniach odpowiedzi na pytanie: jak wygląda otoczenie kulturalne danego obiektu ze względu na odległość wyrażoną w jednostkach miary?

6.2.1 Obiekty wokół punktu. Jeszcze raz o wydatkach gmin.

Mapa gmin w otoczeniu metropolitalnym jest efektem stworzenia wielowymiarowych wskaźników, uwzględniających zmienne gospodarcze i społeczne. Jednak zanim przystąpi się do budowy wskaźników bazując na danych GUS, można rozpocząć analizy od wyznaczenia gmin, które znajdują się w pewnej zdefiniowanej odległości od przyjętego punktu. Pojawiają się tutaj dwa zagadnienia, w jaki sposób wyznaczyć punkt początkowy, i w jaki sposób wyznaczyć odległość. Do wyznaczenia punktu, od którego rozpoczynamy analizy, można zastosować metodę szukania centroidu, czyli przestrzennego środka pewnej powierzchni, lub zrobić to arbitralnie – definiując dany obiekt jako punkt początkowy. Nawet tak ogólne sformułowanie, że środkiem będzie centrum miasta sprawia, iż możemy szukać centroidu dzielnicy, w której znajdują się instytucje municypalne lub - bardziej ogólnie - wiązać środek np. z ratuszem miejskim czy dworcem kolejowym itd. Podobnie z wyznaczaniem odległości - można mówić o odległości, którą da się pokonać pieszo, rowerem, samochodem, komunikacją publiczną lub, bardziej ogólnie, odległość ta może być wyznaczana po linii prostej od punktu. W tej czę-

ści raportu zademonstrowane jest wykorzystanie systemu „GIS Kultura” do wyznaczenia gmin, które znajdują się w odległości 25 km od Ratusza Miejskiego we Wrocławiu oraz określenie wydatków na kulturę na osobę w tych gminach.

Dzięki silnikowi bazy relacyjnej i składni SQL można zadać pytanie jedną komendą.

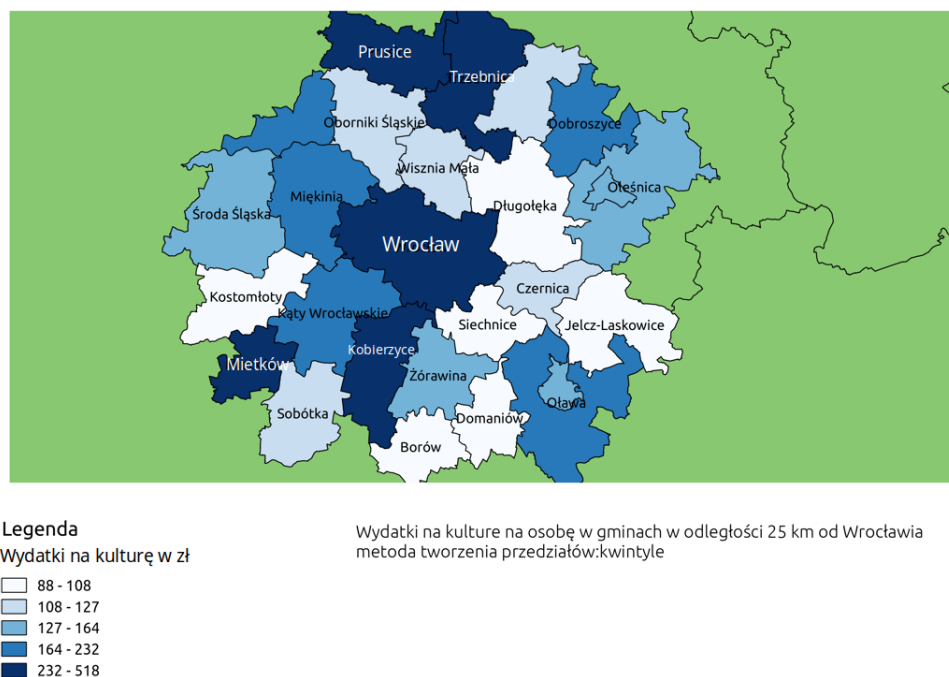
```
%\begin{lstlisting}[frame=single]
select sum(wyd_kult_gminy.wyd_2015::integer)::money, metropol_metropolia,
metropol_typ, way, nazwa_2016
from bazy.wyd_kult_gminy
left join bazy.ludnosc_wg_lat
on lpad(bazy.wyd_kult_gminy.kod::text,7,'0')::text = bazy.ludnosc_wg_lat.teryt::text
left join bazy.zmiany_teryt_gminy
on bazy.ludnosc_wg_lat.teryt::text = zmiany_teryt_gminy.teryt
left join gminy_osm_teryt
on bazy.ludnosc_wg_lat.teryt::text = gminy_osm_teryt.teryt
cross join
(Select st_point(17.031924, 51.109566)::geography AS ref_geog) as r
where st_dwithin(way,
  ST_Transform(
    ST_SetSRID(ST_Point(17.031924, 51.109566 ), 4326), 2180), 25000)
group by metropol_metropolia, metropol_typ, way,nazwa_2016
order by metropol_metropolia desc, metropol_typ asc
%\end{lstlisting}
```

Taka komenda przy użyciu systemu pozwala na prezentację wyników w formie graficznej. Do prezentacji został wykorzystany program QGIS. Efekt takiej analizy prezentuje Rysunek 6.8.

Dzięki tego rodzaju danym, zwizualizowanym na mapie, możliwe jest stwierdzenie, że najwyższe wydatki (ostatni przedział między 232 a 518 zł) są (poza Wrocławiem) w gminach Kobierzyce, Mietków, Prusice, Trzebnica. Z czego tylko Kobierzyce graniczą z Wrocławiem. W przypadku najniższych wydatków w gminach graniczących z Wrocławiem, dostrzec można dwie – Długołękę i Siechnice (wydatki pomiędzy 88 a 108 zł na osobę).

To standardowe pytanie o wydatki, które pojawia się, kiedy rozważana jest kwestia polityk kulturalnych, można zadać w sposób, który bardziej dotyczy uczestników wydarzeń kulturalnych, a mniej decydentów politycznych. Odległość od gmin, które posiadają wysokie wydatki na kulturę może być bardziej precyzyjnym wskaźnikiem dostępności kulturalnej, niż

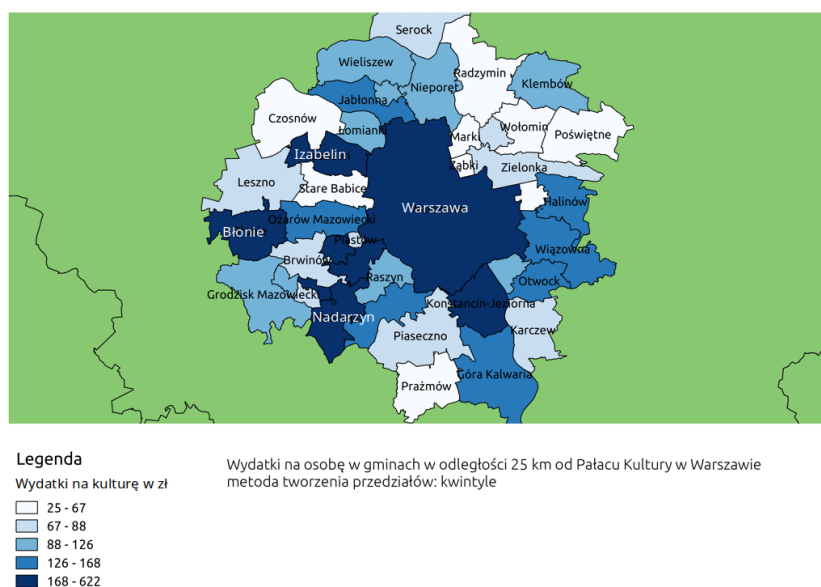
Rysunek 6.8: Wydatki na kulturę na mieszkańca gmin w okolicach Wrocławia.



Źródło: Mariusz Piotrowski, opracowanie własne

same wydatki na kulturę w gminie, w której mieszkają badane osoby. Jeśli mielibyśmy używać tradycyjnych technik badawczych, aby odpowiedzieć na tak postawiony problem badawczy, należałoby zastosować co najmniej dwa narzędzia: mapę i arkusz kalkulacyjny. Dopiero narzędzia klasy GIS pozwalają na integrację danych i ich wizualizację. Dzięki temu można swobodnie modyfikować zapytania, zmieniając tylko zmienne, punkt wyjściowy, odległość czy wskaźnik ilościowy. Przykłady takich modyfikacji prezentują kolejne mapy (Rysunek 6.9). W pierwszym wypadku dane dotyczą wskaźnika sumarycznego wydatków na kulturę dla gmin w odległości 25 km od Ratusza Miejskiego we Wrocławiu. W drugim, dane o wydatkach na kulturę, w przeliczeniu na mieszkańca są wygenerowane dla gmin znajdujących się w odległości 25 km od Pałacu Kultury i Nauki w Warszawie.

Źródło: Mariusz Piotrowski, opracowanie własne



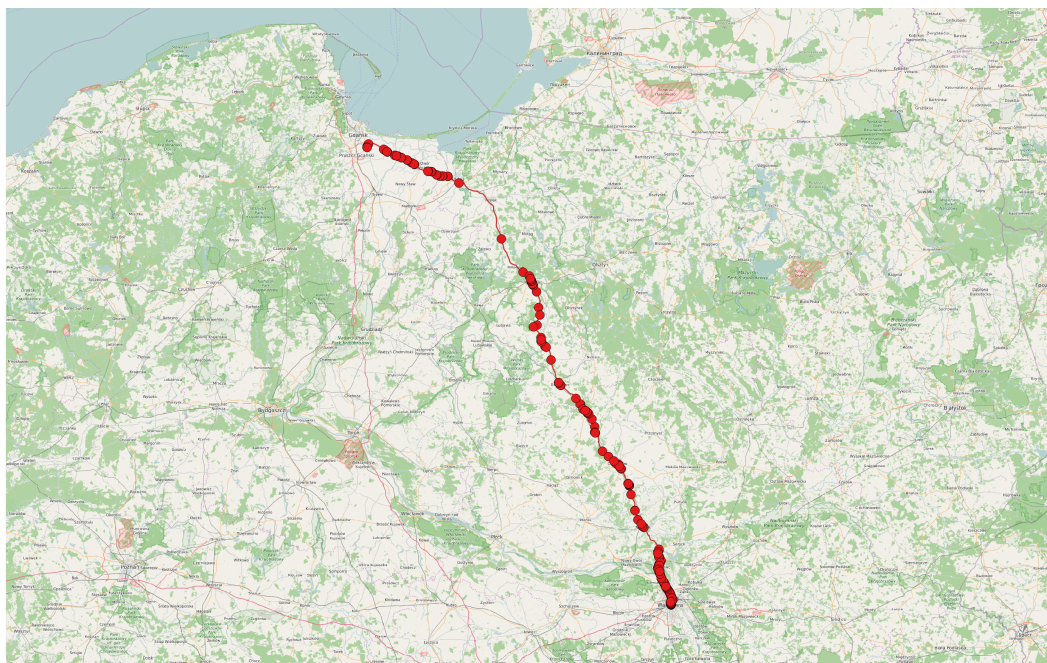
6.2.2 Obiekty przy drodze.

Prototypowa baza systemu GIS Kultura to jednak przede wszystkim możliwość wykonywania analiz na poziomie pojedynczych obiektów. System, dzięki danym o siatce dróg pozwala odpowiadać na pytania o dostępność z uwzględnieniem istniejącej infrastruktury. Z punktu widzenia uczestnika wydarzeń kulturalnych pytanie, które może zostać zadane bazie danych, mogłoby brzmieć następująco – gdzie znajduje się najbliższe kino i jakie obiekty znajdują się na trasie dojazdu do niego? W tej części raportu zostanie zaprezentowana odpowiedź na podobne pytanie: jakie obiekty znajdują się w odległości 50 m od najszybszej trasy między Warszawą a Gdańskiem, przy założeniu że poruszamy się komunikacją samochodową? Aby odpowiedzieć należało wykonać kilka czynności. Przy użyciu oprogramowania QGIS i wtyczki „optymalna droga”, została wyznaczona trasa pomiędzy dwoma punktami, zaś jako kryterium kosztu dotarcia obrany został czas. Czas dotarcia określany jest przez założenie, że podróż odbywa się maksymalną dopuszczalną prędkością na danym odcinku (taka informacja znajduje się w bazie danych, w pliku z geometriami dróg). Kolejną czynnością jest wyznaczenie bufora odległości od wyznaczonej trasy. Bufor został określony na 50 m. Następnie, baza danych została odpytana, jakie obiekty żywej kultury znajdują się w wyznaczonym buforze. Efekt pracy prezentuje Rysunek 6.10

W odległości 50 m od najszybszej trasy samochodowej Warszawa – Gdańsk znajduje się 544 obiektów żywej kultury. Informacje, którą można dodatkowo uzyskać to pełen wgląd w dane o każdym z obiektów. Na najniższym poziomie może być to nazwa i rodzaj indeksu obiektu żywej kultury. Przykładowe 10 obiektów znajdujących się przy trasie prezentuje Tabela 6.7.

Dane o obiektach ze względu na typ indeksu żywej kultury prezentuje Wykres 8. Blisko 20% obiektów pozostało niezidentyfikowanych. Tutaj ponownie uwidacznia się problem z automatyczną klasyfikacją danych. Jednak używając tradycyjnych, nadzorowanych metod klasyfikacji, możliwe jest bardziej precyzyjne przyporządkowanie obiektów do indeksów żywej kultury. Dane tu prezentowane pochodzą z bazy openstreetmap, sprzed harmonizacji, co sprawia, że więcej wysiłku należy włożyć w ich uporządkowanie. Natomiast nawet takie dane można poddawać dalszej obróbce. Wśród ponad 100

Rysunek 6.10: Obiekty w odległości 50 m od najszybszej trasy Warszawa-Gdańsk (N=544).



Źródło: Mariusz Piotrowski, opracowanie własne

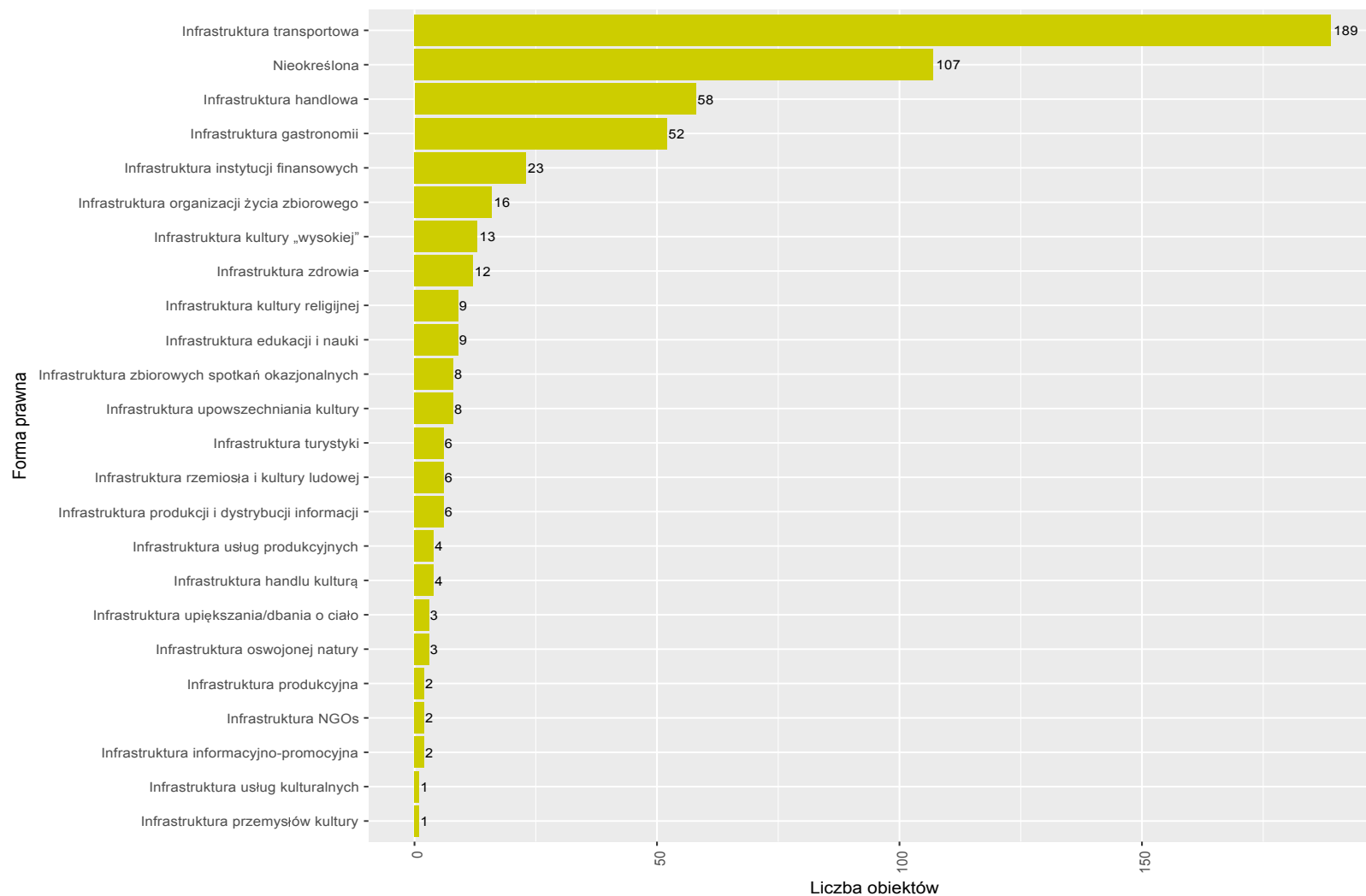
obiektów kategorii *Nieokreślone* można szukać kategorii typów w. słownika openstreetmap. Dane prezentuje Rysunek 6.12.

Tabela 6.7: Przykładowe 9 obiektów znajdujących się w odległości 50 m od najszybszej trasy Warszawa-Gdańsk.

Obiekt OSM	Nazwa	Typ OZK
stop_position_	Ratuszowa-ZOO 04	Infrastruktura transportowa
stop_position_	Siwińskiego 02	Infrastruktura transportowa
shop_supermarket	Biedronka	Infrastruktura handlowa
office_company	zabart.com	Infrastruktura usług produkcyjnych
amenity_fuel	Stacja-LPG	Infrastruktura transportowa
shop_convenience	Delikatesy Panorama	Infrastruktura handlowa
amenity_bank	Alior Bank	Infrastruktura instytucji finansowych
office	Biuro rachunkowe Jovix	Nieokreślona
craft	Serwis Estey	Infrastruktura rzemiosła i kultury ludowej

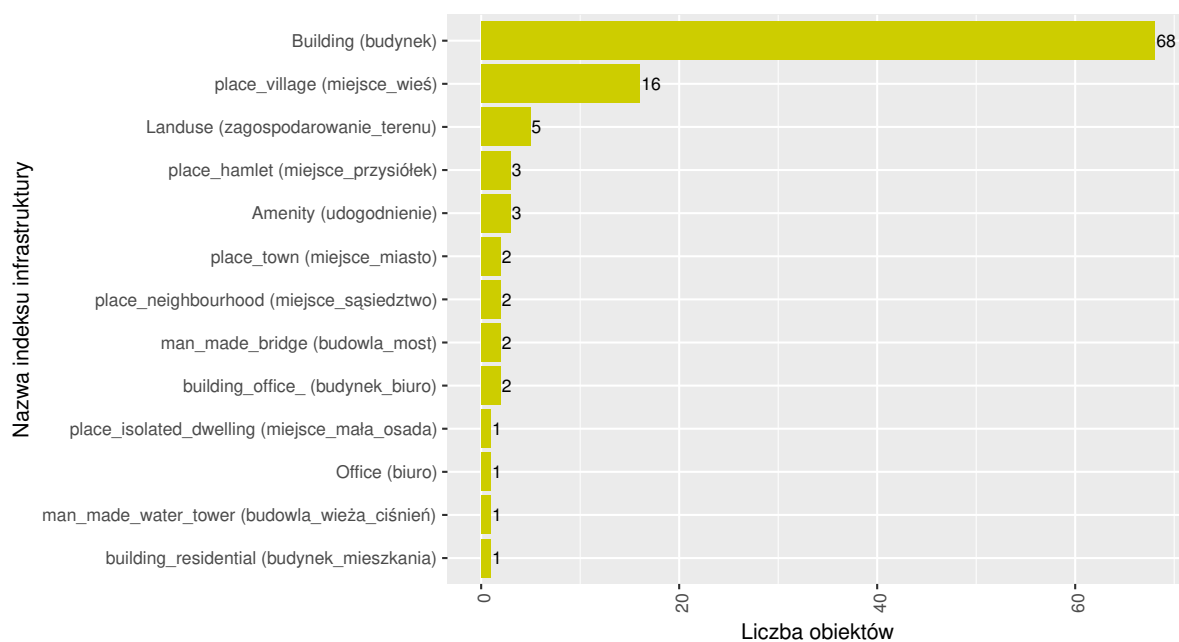
Źródło: Mariusz Piotrowski, opracowanie własne

Rysunek 6.11: Obiekty żywej kultury na trasie Warszawa-Gdańsk (n=544).



Źródło: Mariusz Piotrowski, opracowanie własne

Rysunek 6.12: Obiekty na trasie Warszawa-Gdańsk, bez przypisanego indeksu żywej kultury (n=107).



Źródło: Mariusz Piotrowski, opracowanie własne

6.2.3 Obiekty przy trasie komunikacji publicznej.

Poprzedni opisany przypadek pozwala analizować sytuacje, w których dostępność do infrastruktury jest związana z poruszaniem się pieszo lub samochodem. Jednak kiedy dostęp do kultury potraktuje się jako formę usługi publicznej, gwarantowanej przez jednostki samorządu terytorialnego można zadać pytanie czy istnieje infrastruktura transportowa umożliwiająca dostęp do tych usług? Wskaźnikiem takiej dostępności będzie po pierwsze - bliskość przystanku autobusowego; a po drugie, informacja czy i jakie linie autobusowe zatrzymują się na tym przystanku. W tym celu można wykorzystać dane serwisu openstreetmap. Usługa transportu publicznego w gminach jest realizowana przez władze gminne w różny sposób. Skutkuje to brakiem ogólnokrajowej koordynacji, a w konsekwencji bazy danych o tego typu usługach. Dane są zbierane przede wszystkim przez serwisy komercyjne (w Polsce najpopularniejszy serwis tego typu to jakdojade.pl). Serwis społecznościowy openstreetmap stanowi więc alternatywę dla usług zamkniętych i pozwala na analizy dostępności komunikacyjnej.

W tej części zostanie zaprezentowana odpowiedź na następujące pytanie – przy jakiej linii autobusowej w Warszawie znajduje się najwięcej obiektów żywej kultury? Następnie zostaną one zliczone i zwizualizowane. W tym celu ponownie można wykorzystać bazę relacyjną postgresql i składnię SQL. Założeniem jest, że interesują nas tylko obiekty leżące w odległości 50 m od dróg, którymi porusza się autobus. Komenda prezentuje się następująco:

```
SELECT
  r.name,
  COUNT(DISTINCT public.osm_poi_custom_acces_30_10_2016.id) As tot
FROM
  public.osm_poi_custom_acces_30_10_2016
INNER JOIN public.linie_relacje_komunikacja_publiczna_osm_20_10_2016 As r
ON ST_DWithin(osm_poi_custom_acces_30_10_2016.geom, r.way, 50)
WHERE r.name LIKE '%Bus%'
GROUP BY r.name
ORDER BY tot DESC
LIMIT 8;
```

Efektem działania komendy jest lista 8 linii autobusowych, przy

Tabela 6.8: Linie autobusowe, przy trasie których znajduje się największa liczba obiektów żywej kultury

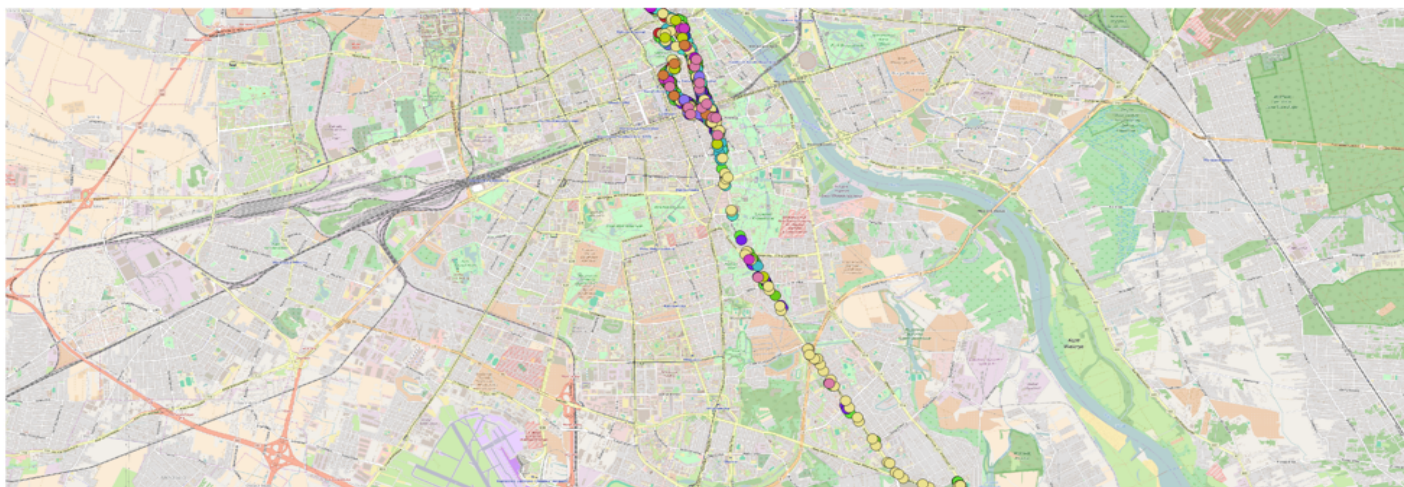
Nr linii	Nazwa linii	Liczba obiektów
Bus 116:	Chomiczówka =>Wilanów	930
	Wilanów =>Chomiczówka	874
Bus 521:	Szczęśliwice =>Falenica	818
	Falenica =>Szczyliwice	811
Bus 518:	Nowodwory =>Dworzec Centralny	756
Bus 517:	Ursus-Niedźwiadek =>Witebska	752
Bus N44:	Zajezdnia Żoliborz =>Dworzec Centralny	744
Bus N03:	Nowodwory =>Ursynów Płn.	727

Źródło: Mariusz Piotrowski, opracowanie własne

których system odnalazł największą liczbę obiektów żywej kultury. Dane prezentuje Tabela 6.8. Wszystkie wyniki to linie autobusowe komunikacji ZTM Warszawa. Baza danych o transporcie openstreetmap jest na tyle precyzyjna, że wskazuje także kierunek trasy autobusowej.

Najwięcej obiektów żywej kultury (930) znajduje na trasie linii autobusu 116 z Chomiczówki do Wilanowa. Druga w kolejności to linia 116, ale jeżdżąca w przeciwnym kierunku (liczba obiektów jest mniejsza o 56). Różnica wynika z tego, że autobusy jeżdżą różnymi drogami, np. część dróg jest jednokierunkowa. Następną czynnością pozostaje określenie jakiego typu są to obiekty. Jest to czynność analogiczna do tej, przy pomocy której określano były obiekty na trasie Warszawa-Gdańsk. Efekt takiego działania prezentuje Rysunek 6.13. Na trasie linii 116 najwięcej jest obiektów infrastruktury gastronomicznej (233), transportowej (221) i handlowej (156); mając jednak do dyspozycji pełen wgląd w dane, można określić jakiego rodzaju są to obiekty. Przykładowe informacje o obiektach prezentuje 6.9.

Rysunek 6.13: Obiekty żywej kultury na trasie linii autobusowej 116, wg typów indeksów. Legenda sortowana alfabetycznie.



Legenda

Obiekty Żywej Kultury [930]

- | | |
|--|---|
| ● Infrastruktura aktywności fizycznej [2] | ● Infrastruktura produkcji i dystrybucji informacji [4] |
| ● Infrastruktura edukacji i nauki [7] | ● Infrastruktura produkcyjna [1] |
| ● Infrastruktura gastronomii [233] | ● Infrastruktura przemysłów kultury [1] |
| ● Infrastruktura handlowa [156] | ● Infrastruktura rzemiosła i kultury ludowej [5] |
| ● Infrastruktura handlu kulturą [21] | ● Infrastruktura transportowa [221] |
| ● Infrastruktura informacyjno-promocyjna [3] | ● Infrastruktura turystyki [12] |
| ● Infrastruktura instytucji finansowych [51] | ● Infrastruktura upiększania/dbania o ciało [15] |
| ● Infrastruktura kultury „wysokiej” [29] | ● Infrastruktura upowszechniania kultury [11] |
| ● Infrastruktura kultury religijnej [7] | ● Infrastruktura usług kulturalnych [4] |
| ● Infrastruktura NGOs [1] | ● Infrastruktura usług produkcyjnych [1] |
| ● Infrastruktura obsługi społecznego funkcjonowania jednostek i grup [1] | ● Infrastruktura zbiorowych spotkań okazjonalnych [8] |
| ● Infrastruktura organizacji życia zbiorowego [27] | ● Infrastruktura zdrowia [30] |
| ● Infrastruktura oswojonej natury [2] | ● Nieokreślona [77] |

Tabela 6.9: Przykładowe obiekty żywej kultury przy trasie linii autobusowej 116, z uwzględnieniem dostępu dla osób poruszających się na wózkach inwalidzkich.

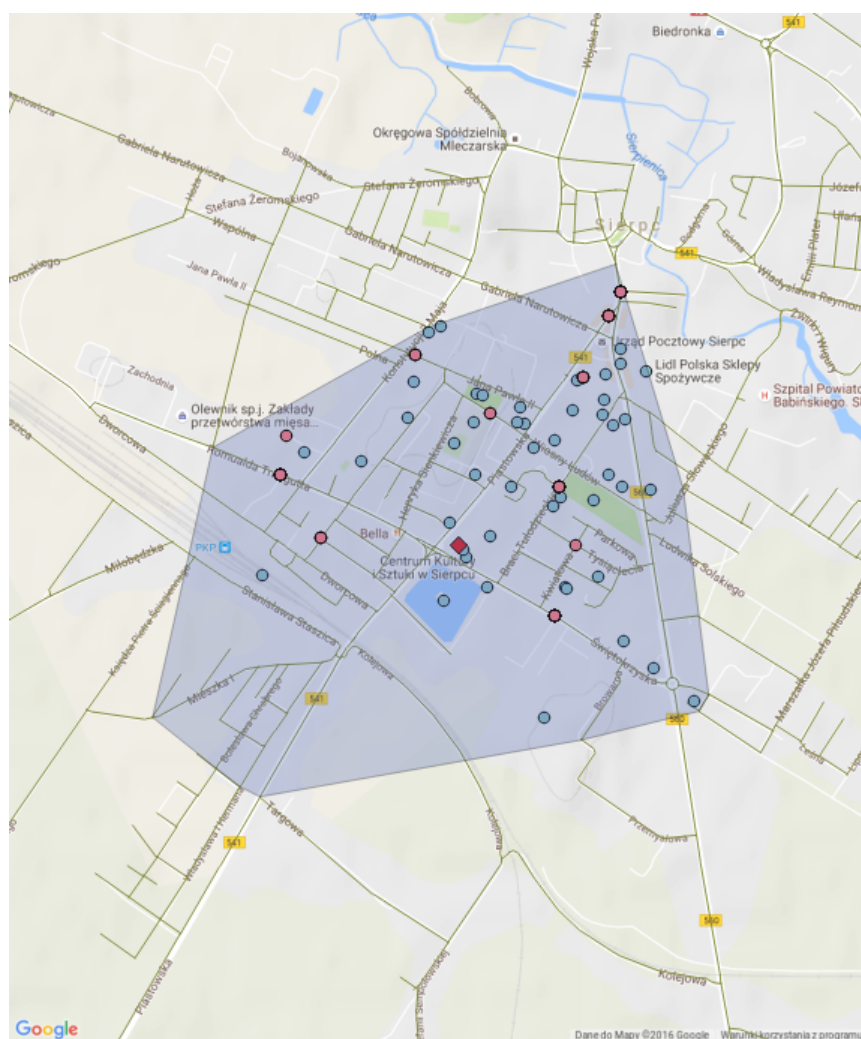
Typ obiektu OSM	Nazwa	Kategoria infrastruktury	czy dostępny dla osób z niepełnośprawnością
Building	Grocery	nieokreślona	Brak danych
Amenity	Ministerstwo Kultury i Dziedzictwa Narodowego; Departament Mecenatu Państwa	nieokreślona	Brak danych
shop_clothes	Ermenegildo Zegna	handlowa	Brak danych
amenity_bank	BZ WBK	instytucji finansowych	Brak danych
amenity_bank	Pekao	instytucji finansowych	Brak danych
shop_newsagent	Looksim	usług kulturalnych	Brak danych
stop_position_platform	Krucza 02	transportowa	Tak

6.2.4 Zasięg dostępności obiektu. Metoda izochronu.

Wśród wielu zapytań przestrzennych, które są możliwe do zrealizowania w projekcie „GIS Kultura”, może pojawić się i takie, które wyrażone przez uczestnika wydarzenia kultury będzie brzmiało – do jakich obiektów kultury mogę dotrzeć w ciągu 10 minut od miejsca zamieszkania? Odległość w nieformalnej komunikacji wyrażana jest przez czas, jaki należy poświęcić, żeby dotrzeć do pewnego miejsca. W takim określeniu związku czasu i przestrzeni konieczne jest założenie o prędkości poruszania się. W przypadku, kiedy osoba porusza się pieszo – założeniem jest, że średnio pokonuje 5 km na godzinę, dla samochodów zmiennych jest więcej, żeby wspomnieć o dwóch – jest to dopuszczalna prędkość na drodze i natężenie ruchu na niej.

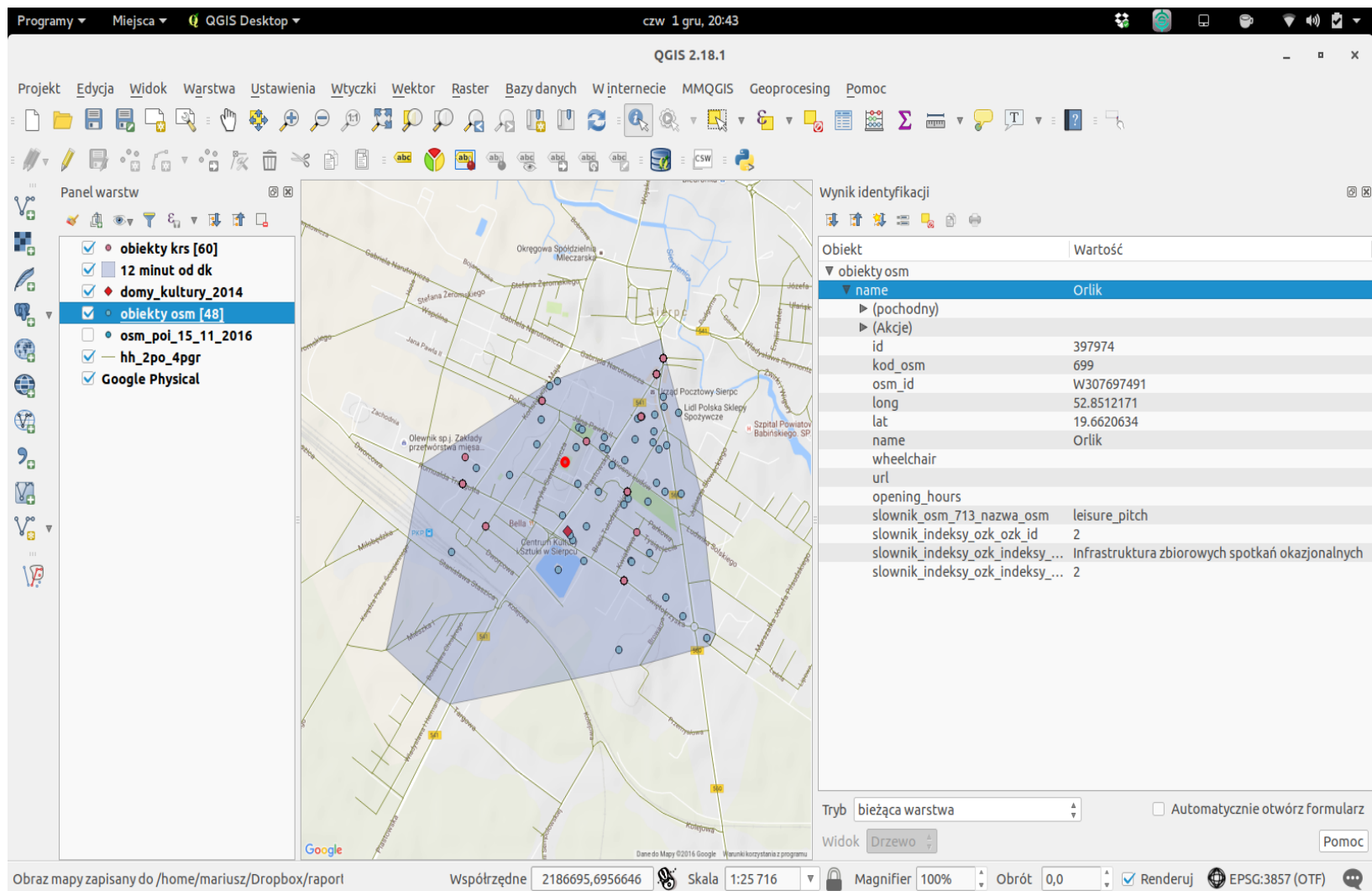
Zdobywanie informacji o natężeniu ruchu wymaga systemu, który pozwala na ciągle zdobywanie informacji o zachowaniach komunikacyjnych użytkowników dróg. Dlatego też w systemie „GIS Kultura” przyjęliśmy bardziej ogólne założenia, wynikające ze średniej prędkości poruszania się dla pieszego, i maksymalnej prędkości dopuszczalnej dla samochodu. Poniżej przedstawiony jest opis możliwości analiz, które wychodzą od zapytania dotyczącego dostępności wg czasu dotarcia. Prezentację graficzną przedstawia Rysunki 6.14 i 6.15.

Rysunek 6.14: Obiekty żywej kultury w dostępności 12 minut od Domu Kultury (Sierpiec).



Źródło: Mariusz Piotrowski, opracowanie własne

Rysunek 6.15: Informacje o obiektach żywej kultury w dostępności 12 minut od Domu Kultury.



Źródło: Mariusz Piotrowski, opracowanie własne

Rozdział 7

Zakończenie

Budowa systemu geoinformacji o obiektach żywej kultury „GIS Kultura” jest przedsięwzięciem prototypowym. Jego możliwości analityczne zostały dotychczas wykorzystane m.in. w projektach prowadzonych przez Fundację Obserwatorium Żywej Kultury – Sieć Badawczą takich, jak: „Wpływ diagnoz i badań na realizację polityk kulturalnych w gminach i województwach” czy też badania publiczności Starego Teatru w Krakowie. Z zainteresowaniem spotkały się prezentacje projektu „GIS Kultura” w środowisku naukowym, np. na XIV Ogólnopolskim Zjeździe Socjologicznym czy dla animatorów kultury, którzy brali udział w projekcie Dom Kultury + oraz podczas paneli organizowanych przy okazji realizacji pierwszego wspomnianego projektu. Konstrukcja systemu, zarówno w warstwie wykorzystanych źródeł danych, jak i podejścia do analiz, pozwala na jego wykorzystanie przez teoretyków, w empirycznych badaniach kultury oraz – co szczególnie ważne – przez praktyków. Myślenie o zasobach kulturalnych, nieoczywistych, ale dostępnych przestrzennie, chociażby w kategoriach odległości pomiędzy różnymi podmiotami, pozwala na racjonalne wykorzystanie synergii w działaniach animatorów kultury. Sieciowanie rozmaitych instytucji kultury, będące wynikiem nie tylko pokrewnej funkcji czy specyfiki działań, ale także bliskości i dostępności przestrzennej, może być zasobem, chociaż rzadko dotąd brany pod uwagę i opisywanym, to jednak potencjalnie bardzo wartościowym, co starałem się pokazać w niniejszym raporcie.

Bibliografia

- Jochen Albrecht, *Key concepts & techniques in GIS*, SAGE Publications, 2007
- Rafał Drozdowski, Barbara Fatyga, Mirosław Filiciak, Marek Krajewski, Tomasz Szlendak, *Praktyki kulturalne Polaków*, Wydawnictwo UMK, Toruń, 2014
- Richard Florida, *Narodziny klasy kreatywnej*, NCK, Warszawa, 2010
- Waldemar Izdebski, *Dobre praktyki udziału gmin i powiatów w tworzeniu infrastruktury danych przestrzennych w Polsce*, Geo-System, Warszawa, 2015
- Bohdan Jałowiecki, Wojciech Łukowski (red.), *Szata informacyjna miasta*, Scholar, Warszawa, 2008
- Dobiesław Jdrzejczyk, *Geografia humanistyczna miasta*, Wydawn. Akademickie "Dialog", 2004
- Tomasz Kukołowicz (red.), *Statystyka kultury w Polsce i Europie. Aktualne zagadnienia*, NCK, Warszawa, 2015
- Antonina Kłoskowska, *Spoleczne ramy kultury*, PWN, Warszawa, 1972
- Antonina Kłoskowska, *Kultura masowa: krytyka i obrona*, PWN, Warszawa, 2006
- Antonina Kłoskowska, *Socjologia kultury*, PWN, Warszawa, 2007

- Elaine Lewinnek, *Mapping Chicago, Imagining Metropolises: Reconsidering the Zonal Model of Urban Growth* [w:] *"Journal of Urban History"*, 36(2):197–225, 2010
- Beata Medyńska-Gulij, *Kartografia i geowizualizacja*, PWN, Warszawa, 2011
- *Nowa sprawozdawczość instytucji kultury*, Kraków: MIK, 2015
- Zbigniew Taylor, *Przestrzenna dostępność miejsc zatrudnienia, kształcenia i usług a codzienna ruchliwość ludności wiejskiej*, [w:] *„Prace Geograficzne IGiPZ PAN”*, nr 171, Wrocław 1999.
- Andrzej Tyszką, *Interesy i ideały kultury struktura społeczeństwa i udział w kulturze*, PWN, Warszawa, 1987
- Aleksander Wallis, *Informacja i gwar*, PIW, Warszawa, 1979
- Aleksander Wallis, *Hierarchia miast* [w:] *"Studia socjologiczne"*, 1(200), 2011
- Grzegorz Wclawowicz, *Geografia społeczna miast*, PWN, Warszawa, 2007
- James D Wright, *The Founding Fathers of Sociology : Francis Galton, Adolphe Quetelet , and Charles Booth Or What Do People You Probably Never Heard of Have to Do with the Foundations of Sociology?* [w:] *"Journal of Applied Social Science"*, 3(2):63–72, 2016
- Janusz Ziółkowski, *Urbanizacja, miasto, osiedle. Studia socjologiczne*, PWN, Warszawa, 1965.
- Florian Znaniecki, Socjologiczne podstawy ekologii ludzkiej,[w] *„Ruch Prawniczy, Ekonomiczny i Socjologiczny”*, z. 1/1938.
- Wiesława Żyszkowska, Waldemar Alfred Spallek, Dorota Borowicz, *Kartografia tematyczna*, PWN, Warszawa, 2012