

# **Analiza skali wykluczenia komunikacyjnego na obszarze Polski wraz z rekomendacjami zmian legislacyjnych w kontekście publicznego transportu zbiorowego T-INCLUDED**

## **Zadanie 3 Opracowanie łatwego do aktualizacji i wykorzystania przez podmioty trzecie podsystemu informacji o funkcjonowaniu PTZ**

---

**Opracowanie łatwego do aktualizacji, interpretacyjnego  
i multimodalnego podsystemu informacji o funkcjonowaniu  
PTZ w Polsce**

---

30 kwietnia 2025 r.



Rzeczpospolita  
Polska

**NCBR**  
Narodowe Centrum Badań i Rozwoju



POLITECHNIKA POZNAŃSKA



Politechnika  
Śląska



POLITECHNIKA  
GDAŃSKA

Politechnika  
Warszawska

### INFORMACJE O WYKONAWCY

#### **Politechnika Poznańska**

pl. M. Skłodowskiej-Curie 5, 60-965 Poznań

NIP: 7770003699

REGON: 000001608

e-mail: [biuro.rektora@put.poznan.pl](mailto:biuro.rektora@put.poznan.pl)

strona: <http://www.put.poznan.pl>

### ZESPÓŁ WYKONAWCÓW PRAC

#### **Wydział Inżynierii Lądowej i Transportu Politechniki Poznańskiej**

ul. Piotrowo 3, 61-138 Poznań

strona: <https://wilit.put.poznan.pl>

Wykonawcy:

- Maciej Bieńczak – IT, ZST
- Szymon Fierek – IT, ZST
- Marcin Kiciński – IT, ZST
- Marek Łukaszczyk – IT
- Andrzej Markiewicz – IT
- Jeremi Rychlewski – IIL, ZBMiDK
- Waldemar Walerjańczyk – IT, ZST
- Paweł Zmuda-Trzebiatowski – IT, ZST (*kierownik projektu na PP*)

# Spis treści

<b>1. Wprowadzenie</b>	<b>4</b>
<b>2. Analiza rozwiązań technologicznych i uruchomienie serwera bazodanowego</b>	<b>5</b>
2.1. Wprowadzenie	5
2.2. Kryteria analizy dostępnych rozwiązań technologicznych i bazodanowych	5
2.3. Przegląd istniejących rozwiązań oprogramowania	6
2.4. Przegląd istniejących rozwiązań sprzętowych	7
2.5. Wybór konfiguracji sprzętowej	9
2.6. Wybór konfiguracji oprogramowania	11
2.7. Podsumowanie w zakresie analizy rozwiązań technologicznych i uruchomienia serwera bazodanowego	12
<b>3. Analiza porównawcza pozyskanych danych o funkcjonowaniu PTZ</b>	<b>13</b>
3.1. Założenia wstępne dotyczące analizy porównawczej pozyskanych danych o funkcjonowaniu PTZ	13
3.2. Pozyskanie i wprowadzenie do bazy informacji o rozkładach jazdy	13
3.3. Ocena zgodności danych ze standardem GTFS	19
3.4. Ocena standardów wprowadzania informacji do systemu o funkcjonowaniu PTZ	21
<b>4. Analityczne przetestowanie rozwiązań kreatorów dodawania danych o funkcjonowaniu PTZ</b>	<b>23</b>
4.1. Ocena istniejących kreatorów danych o funkcjonowaniu PTZ	23
4.2. Opracowanie wstępnej wersji autorskiego narzędzia pozwalającego na cyfryzację analogowych danych o funkcjonowaniu PTZ	26
<b>5. Ocena osiągnięcia kamieni milowych</b>	<b>29</b>
<b>6. Podsumowanie</b>	<b>31</b>

## 1. Wprowadzenie

Zadanie 3 w projekcie zakładało cztery podstawowe działania, tj.:

- D1 - Analiza rozwiązań technologicznych i uruchomienie serwera bazodanowego;
- D2 - Analiza porównawcza pozyskanych danych o funkcjonowaniu PTZ, zarówno w formie plików GTFS, jak i analogowych, pod kątem potrzeb w zakresie standaryzacji zapisu informacji o funkcjonowaniu PTZ oraz potrzeb rozszerzenia standardu GTFS;
- D3 - Analityczne przetestowanie rozwiązań kreatorów umożliwiających przygotowanie lub cyfryzację danych o funkcjonowaniu PTZ zgodnie ze standardem GTFS oraz narzędzi umożliwiających import rezultatów do bazy danych;
- D4 - Opracowanie raportu częściowego z realizacji Zad. 3.

Spodziewane efekty tych działań zostały sformułowane w postaci trzech kamieni milowych, które sformułowano w sposób następujący:

- KM3.1 - Wyniki laboratoryjnego testowania bazy danych informacjami o funkcjonowaniu PTZ, przy czym parametrem tego kamienia milowego jest pozyskanie i wprowadzenie do bazy informacji o rozkładach jazdy od min. 20 organizatorów PTZ lub przewoźników.
- KM3.2 - Wstępna diagnoza dotycząca potrzeb dostosowania standardu GTFS oraz potrzeb standaryzacyjnych wprowadzania danych o PTZ na potrzeby projektu. Parametrem oceny tego kamienia są: listy zawierające minimum wskazanie problemów i sposobu rozwiązania dla danych pozyskanych od minimum 20 organizatorów lub przewoźników PTZ, w zakresie: 1) zgodności danych ze standardem GTFS - 1 lista, 2) standardów wprowadzania informacji do systemu o funkcjonowaniu PTZ - 1 lista.
- KM3.3 - Opracowanie wstępnej wersji narzędzia pozwalającego na cyfryzację danych analogowych. Za parametr oceny tego kamienia przyjęto minimum 1 interfejs użytkownika umożliwiający prowadzenie prac przez wszystkich konsorcjantów.

## **2. Analiza rozwiązań technologicznych i uruchomienie serwera bazodanowego**

### **2.1. Wprowadzenie**

Właściwy dobór rozwiązań software'owych, adekwatnych do modelowanej problematyki konfiguracji sprzętowej oraz metodyki organizacji prac programistycznych będzie rzutować na efektywność prowadzonych prac badawczych, a w dalszej konsekwencji na możliwość realizacji podstawowego celu projektu. Problematyka analizy rozwiązań technologicznych i uruchomienia serwera bazodanowo-aplikacyjnego została wydzielona jako pierwsze działanie w ramach realizacji Zad. 3 projektu.

W tym celu przeprowadzono analizę dostępnych rozwiązań technologicznych, przyjęto kryteria zastosowane przy ich wyborze, a ostatecznie na podstawie tych kryteriów dokonano wyboru systemów. Rozpatrywanymi elementami oprogramowania systemu są:

- system operacyjny,
- silnik bazodanowy,
- warstwa backend aplikacji,
- warstwa frontend aplikacji.

Ze względu na równoległe prowadzenie prac badawczych i rozwojowo technologicznych podjęty został również problem doboru metodyki pracy odpowiedniej dla specyfiki zespołu projektowego.

### **2.2. Kryteria analizy dostępnych rozwiązań technologicznych i bazodanowych**

W celu obiektywnej oceny istniejących na rynku rozwiązań sformułowano zestaw kryteriów, dzięki którym możliwa jest systemowa ocena poszczególnych rozwiązań. Przyjęto następujący zbiór dziewięciu kryteriów:

- 1) Licencja - rodzaj licencji rozpatrywanego oprogramowania;

- 2) Dostęp do źródeł - otwartość/zamkniętość źródeł, możliwość utworzenia własnej gałęzi - uzależnione od licencji;
- 3) Wbudowane funkcje - funkcje domyślnie zawarte lub w łatwy sposób możliwe do dołączenia w oprogramowaniu;
- 4) Wydajność oraz skalowalność rozwiązania;
- 5) Łatwość integracji - czy wybrany element jest możliwy do integracji z innym rozpatrywanymi elementami oprogramowania;
- 6) Popularność w przemyśle - skala wykorzystania w przemyśle dla podobnych rozwiązań;
- 7) Koszt;
- 8) Rozszerzalność - dostępność rozszerzeń (wtyczek, bibliotek) elementu systemu;
- 9) Wsparcie producenta i dokumentacja techniczna.

### 2.3. Przegląd istniejących rozwiązań oprogramowania

W wyniku realizacji prac koncepcyjnych przyjęto podział istniejących rozwiązań oprogramowania na:

- 1) System operacyjny - podstawowa część rozwiązania. Obecnie dla rozwiązań serwerowych *on premise* rynek podzielony jest między systemy Microsoft Server oraz rozwiązania z rodziny systemów Linux.
- 2) Silnik bazy danych - wśród najczęściej stosowanych implementacji baz danych w standardzie ANSI SQL wyróżnić można SQL Server firmy Microsoft, Oracle Database firmy Oracle, a z rozwiązań open source: MySQL czy darmowy PostgreSQL<sup>1</sup> rozwijany przez Uniwersytet Kalifornijski.
- 3) Warstwa backend aplikacji - ze względu na architekturę aplikacji pod uwagę należy wziąć rozwiązania mające zaprojektowane z naciskiem na wykorzystanie REST API z naciskiem na modularność i podział na osobne jednostki wdrożeniowe. Wiodące

---

<sup>1</sup> PostgreSQL: The World's Most Advanced Open Source Relational Database, on-line: <https://www.postgresql.org/> (dostęp 5.11.2021r.)

obecnie rozwiązania to - Django dla języka Python, Express.js dla server side javascript, ASP.NET Core dla C#, Laravel dla PHP, Spring Boot dla platformy Java<sup>2</sup>

- 4) Warstwa frontend - Oprócz technologii implementującej wymagania logiczne aplikacji należy również wziąć pod uwagę warstwę prezentacji. W przemyśle istnieje bardzo duża przestrzeń rozwiązań tego problemu, jednak dominują rozwiązania oparte na Javascript - wśród nich wyróżnić można React.js rozwijany przez Facebook, Angular rozwijany przez Google oraz Vue.js rozwijany przez społeczność.

### 2.4. Przegląd istniejących rozwiązań sprzętowych

Rynek serwerów jest obecnie oparty o dwie dominujące architektury - ARM (np. Ampere Altra) oraz x86 (AMD Epyc, Intel Xeon), które na potrzeby obliczeniowe, w tym związane ze sztuczną inteligencją, mogą być wsparte akceleratorami, np. NVIDIA Tensor Core. Obie architektury wspierają konfiguracje 1-2 lub więcej (4-8) procesorowe. Oba rozwiązania wykorzystują pamięci DDR4 3200 w trybie wielokanałowym (4, 6, 8) i o pojemności do 4TB na gniazdo procesora. Podobna jest też szybkość magistrali PCI Express wykorzystywanej do komunikacji z akceleratorami, niektórymi pamięciami masowymi, a także innymi modułami, jak np. karty sieciowe (wersja 4, 64 lub 128 linii). Rozwiązania ARM oferują w chwili obecnej większą maksymalną liczbę rdzeni na procesor (80) względem rozwiązań x86 (40-64), jednak maksymalna częstotliwość taktowania, a także dostępna pojemność pamięci podręcznej mogą być niższe, np. odpowiednio 3,3 GHz i 32MB dla 80 rdzeniowego Ampere Q80-33 vs. 4,1GHz i 256MB dla 8 rdzeniowego AMD EPYC 72F3. Ta. 1, poniżej pokazuje porównanie podstawowych architektur. Zawarto w niej porównanie architektur procesorów serwerowych za serwisem Anandtech.com.

---

<sup>2</sup> Stackoverflow technology survey 2021, on-line: <https://insights.stackoverflow.com/survey/2021#technology> (dostęp 01.11.2021r.)

Tab. 1. Porównanie architektur dostępnych procesorów serwerowych (źródło: [anandtech.com](https://www.anandtech.com)<sup>3</sup>)

Cecha	Intel Xeon	AMD EPYC	Ampere Altra
Platforma	Ice Lake	Milan (7003)	QuickSilver
Procesor (top)	8380	7763	Q80-33
Architektura	Sunny Cove	Zen 3	N1
Chipset	płyta główna	na CPU	na CPU
Liczba rdzeni	40	64	80
TDP [W]	270	280	250
Częstotliwość bazowa	2300	2450	3300
Częstotliwość turbo - 1 rdzeń	3400	3500	3300
Częstotliwość turbo - wszystkie rdzenie	3000	3200	3300
Cache L3 [MB]	60	256	32
PCIe (standard x liczba linii)	4.0 x 64	4.0 x 128	4.0 x 128
pamięć DDR4 - taktowanie [MHz]	3200	3200	3200
pamięć - liczba kanałów	8	8	8
pamięć maksymalna [TB]	4	4	4
Obsługa technologii Optane	tak	nie	nie
Cena [USD]	8099	7890	4050

Drugim istotnym elementem rozwiązań serwerowych jest pamięć masowa. W chwili obecnej dominują trzy typy złącz (SATA, SAS i NVME) oraz dwa rodzaje dysków (HDD, SSD). Najwolniejsze pod względem liczby operacji odczytu i zapisu są dyski HDD (talerzowe) działające na interfejsie SATA (6 Gbps). Interfejs SAS zapewnia transfery rzędu 24 Gbps, tj. 2x12 Gbps, przy czym dostępne są też tak zwane dyski value SAS wykorzystujące tylko jedno złącze 12Gbps przy jednoczesnym niewiele większym koszcie niż dyski SATA. Najszybszymi, ale jednocześnie najdroższymi pamięciami masowymi są dyski SSD (NVME) podłączane bezpośrednio w magistralę PCI Express. Na rynku oferowane są dyski o zróżnicowanej trwałości, którą można zaliczyć do trzech głównych grup, tj.:

- względnie najmniej trwałe, a zarazem najtańsze, przeznaczone do serwerów, w których dominują operacje odczytu (tzw. read intensive), tj. do większości typowych zastosowań www - parametr DWDP (Drive Writes Per Day - liczba zapisów dziennie) na poziomie 1;

<sup>3</sup> <https://www.anandtech.com/show/16594/intel-3rd-gen-xeon-scalable-review/2>



- względnie najbardziej trwałe, a zarazem najdroższe, przeznaczone do serwerów, w których dokonywanych jest wiele operacji zapisu, np. w centrach danych i obliczeniowych - parametr DWDP na poziomie 10;
- mieszane przeznaczenie (mixed use) - parametr DWDP na poziomie 3.

Pojemność pojedynczego dysku wynosi od kilkuset GB do zazwyczaj około 8TB. Dyski mogą być jednak łączone w macierze RAID w różnych konfiguracjach. Rozwiązaniem dostępnym wyłącznie na platformie Intela jest technologia Optane, której celem jest przyspieszenie operacji dyskowych.

Ponadto na rynku oferowanych jest wiele rozwiązań w zakresie podłączenia serwerów do sieci, zdalnego zarządzania (np. moduły iDRAC) czy bezpieczeństwa i niezawodności pracy (redundancja zasilania czy możliwość wymiany dysków podczas pracy serwera). Rozwiązania te powinny być dostosowane do obowiązujących w serwerowni.

### 2.5. Wybór konfiguracji sprzętowej

W związku z badawczym charakterem projektu, w którym dopiero odkryta ma być część parametrów, mających wpływ na wybór konfiguracji sprzętowej, dobór sprzętu oparty został na doświadczeniach z oprogramowaniem, którego zakres jest w pewnym sensie wspólny z projektowanym systemem, tj. bazami danych geograficznych, systemami GIS, planerami podróży czy narzędziami do modelowania ruchu. Przykłady oczekiwanych parametrów technicznych sprzętu można znaleźć w licznych źródłach (np. <sup>4,5,6,7,8,9,10</sup>).

Zespół autorski przeprowadził także wstępne szacowanie, w oparciu o istniejące publicznie dostępne bazy danych. Biorąc pod uwagę, że analizy będą prowadzone w skali kraju, sam zbiór Bazy Danych Obiektów Topograficznych 10k po rozpakowaniu zajmuje

---

<sup>4</sup> [http://wiki.gis.com/wiki/index.php/Capacity\\_Planning\\_Tool](http://wiki.gis.com/wiki/index.php/Capacity_Planning_Tool)

<sup>5</sup> <https://www.postgresql.org/message-id/54DBE7F2.4010202@ase-research.org>

<sup>6</sup> <https://gis.stackexchange.com/questions/122310/hardware-requirements-for-one-instance-geoserver-postgis>

<sup>7</sup> <https://github.com/GIScience/openrouteservice/wiki/System-Requirements>

<sup>8</sup> <https://ask.openrouteservice.org/t/requirements-based-on-osm-data/2016/6>

<sup>9</sup> <https://ask.openrouteservice.org/t/question-about-ram-requirement/3306>

<sup>10</sup> <https://github-wiki-see.page/m/GIScience/openrouteservice/wiki/System-Requirements>

blisko 20GB, a jego charakter jest wyłącznie pomocniczy w rozpatrywanym problemie oceny skali wykluczenia komunikacyjnego liczonej dla 125 tys. - 8 mln punktów adresowych (w zależności od przyjętych założeń wyboru) dostępnych w bazie Państwowego Rejestru Nazw Geograficznych.

Należy również zauważyć, że pozyskanie sprzętu odbywało się w warunkach przetargowych w dość trudnym, pandemicznym roku (i związanym z tym kryzysem na rynku półprzewodników - por. np.<sup>11</sup>). Stąd na ostateczną konfigurację sprzętu wpływ miały także takie czynniki, jak dostępność niezależnych, obiektywnych testów sprzętu (tj. SPEC INT2017) czy fizyczna, a nie jedynie katalogowa, dostępność określonych konfiguracji. Poniżej przedstawione zostały oczekiwania wobec kluczowych elementów sprzętu:

- procesor - oczekiwana względnie duża wydajność z rdzenia, przy relatywnie dużej liczbie rdzeni w związku ze spodziewanymi problemami;
- pamięć RAM - ponad 256GB RAM (możliwie szybkiej);
- dyski twarde - ponad 5,5TB miejsca przy założeniu redundancji RAID1;
- karta graficzna/akcelerator obliczeń - brak wstępnych potrzeb we wskazanym zakresie przy pozostawieniu wolnego miejsca na podstawowy akcelerator AI (na wypadek, gdyby ostatecznie wykorzystane algorytmy wyznaczające wykluczenie komunikacyjne korzystały z takich rozwiązań;
- inne: karta sieciowa, moduł zarządzania zdalnego, zasilanie.

Ostatecznie zakupiono serwer Dell PowerEdge R650, który wyposażony został w:

- jeden 18-rdzeniowy procesor Intel Xeon Gold 6354, z możliwością rozbudowy o drugi,
- 384 GB RAM DDR4 pracujące w trybie 6-kanałowym, z możliwością rozbudowy do 2TB pracujące w trybie 8-kanałowym,
- cztery dyski twarde SSD vSAS o łącznej pojemności 11,52TB, pracujące pod kontrolą fizycznego kontrolera RAID PERC745; z możliwością rozbudowy o kolejne cztery dyski, a także utworzenia macierzy RAID,

---

11

[https://pl.wikipedia.org/wiki/Niedob%C3%B3r\\_p%C3%B3%C5%82przewodnik%C3%B3w\\_\(2020%E2%80%932021\)](https://pl.wikipedia.org/wiki/Niedob%C3%B3r_p%C3%B3%C5%82przewodnik%C3%B3w_(2020%E2%80%932021))

- karta sieciowa zgodna z przyłączami w miejscu kolokacji, tj. 1Gbps; z możliwością rozszerzenia o inne karty sieciowe,
- zintegrowana karta graficzna,
- moduł zarządzania zdalnego iDRAC, a także dwa redundantne zasilacze.

### 2.6. Wybór konfiguracji oprogramowania

Pozyskany serwer Dell PowerEdge R650 deklaruje wsparcie dla systemów Linux, MS Windows Server, VMWare oraz Citrix. W ocenie autorów najniższe koszty zakupu przy jednoczesnym zapewnieniu stosunkowo dużego wsparcia dla oprogramowania (bazy danych, systemy GIS, planery podróży) zapewniają systemy linuxowe, które są rozpowszechniane na licencjach open source. Producent sprzętu zadeklarował wsparcie dla trzech rozwiązań tego typu, tj. RedHat Enterprise Linux Server, Novell SuSE Linux Enterprise Server oraz Ubuntu Server. Zdecydowano się na ten ostatni przez wzgląd na cenę (0 PLN), stosunkowo dużą popularność oraz to, że wspierana przez platformę sprzętową wersja, tj. 20.04 LTS, ma zadeklarowane wsparcie przynajmniej do kwietnia 2025<sup>12</sup>. W związku z tym w trakcie trwania projektu prawdopodobnie możliwe będzie uniknięcie prac związanych z poważną migracją między wersjami systemu operacyjnego.

W kontekście przedmiotowego projektu baza danych jest elementem oprogramowania o największym znaczeniu dla działania wynikowej aplikacji - szybkości obliczeń i możliwości implementacji algorytmów. Podobnie, jak w przypadku systemu operacyjnego podjęto decyzję o wykorzystaniu bazy danych PostgreSQL, która jest popularnym (duże wsparcie społeczności) i opensourcowym rozwiązaniem. Ponadto dysponuje rozszerzeniami ułatwiającymi manipulowanie danymi przestrzennymi - zainstalowano PostGIS.

Podobne podejście zastosowano także przy wyborze narzędzi wspierających prace programistyczne - zainstalowano lokalne instancje Gitlaba oraz Dockera. Pierwsze ze wskazanych narzędzi pozwala przyspieszyć prace programistyczne, natomiast drugie konteneryzować, co jest zgodne ze współczesnymi dobrymi praktykami programistycznymi.

---

<sup>12</sup> <https://wiki.ubuntu.com/Releases>

W przypadku warstwy backend aplikacji (poza silnikiem bazodanowym), podjęto decyzję o wykorzystaniu “.net core” ze względu na aktywne wsparcie firmy Microsoft (poprawki wydajnościowe, częste nowe wydania), otwarty kod źródłowy (bezpieczeństwo) i rozbudowaną społeczność, co niesie ze sobą dużą liczbę bibliotek ułatwiających pracę, a także multiplatformowość i idącą za tym możliwość konteneryzacji w dockerze.

W warstwie frontend aplikacji zdecydowano się na “react js”, gdyż sam w sobie nie pociąga za sobą określonego podejścia do architektury aplikacji umożliwiając jednocześnie realizację jednego z założeń projektowych, tj. aplikacji działającej w chmurze i dostępnej przez przeglądarkę internetową bez przywiązania do konkretnych platform sprzętowych i systemów operacyjnych. Ponadto react w przeciwieństwie do angulara jest biblioteką, a nie frameworkiem, co daje większą dowolność w rozwoju aplikacji. React ma wsparcie Facebooka (Meta) i społeczności co sprawia, że można korzystać ze sprawdzonej biblioteki komponentów, takich jak np. leaflet zapewniający wsparcie dla webmap.

Ponadto przyjęte rozwiązania zarówno w warstwie backend, jak i frontend oraz pomocniczych narzędzi programistycznych są zgodne z tymi, dla których największe doświadczenie posiadają pozyskani do projektu starsi specjaliści w zakresie projektowania i rozwoju systemów informatycznych w transporcie. Umożliwia to większą efektywność ich prac, tj. możliwość poświęcenia mniejszej ilości czasu na zapoznavanie się z samymi narzędziami.

### **2.7. Podsumowanie w zakresie analizy rozwiązań technologicznych i uruchomienia serwera bazodanowego**

Należy zauważyć, że podjęte decyzje dotyczące zarówno architektury sprzętowej, jak i programowej, zapewniają pewną elastyczność oraz otwartość na wprowadzenie nowych rozwiązań w przypadku, gdyby takowe okazały się niezbędne w toku realizacji kolejnych zadań.

### **3. Analiza porównawcza pozyskanych danych o funkcjonowaniu PTZ**

#### **3.1. Założenia wstępne dotyczące analizy porównawczej pozyskanych danych o funkcjonowaniu PTZ**

W ramach Działania 2 pozyskano dane o rozkładach jazdy, które wprowadzono do tworzonego systemu informacji o zagrożeniu wykluczeniem komunikacyjnym. Dokonano także ich analizy porównawczej pod kątem potrzeb w zakresie standaryzacji zapisu informacji. Analiza była prowadzona zarówno dla danych w formie plików GTFS, jak i analogowych, a jej głównym celem było określenie potrzeb rozszerzenia podstawowego standardu GTFS, tak by mógł prawidłowo opisywać rozkłady jazdy, które realnie funkcjonują w Polsce, a także by możliwa była realizacja innych zadań związanych z projektem. W efekcie, w ramach działania powstały dwie listy:

- zgodności pozyskanych danych dotyczących rozkładów jazdy ze standardem GTFS, w tym ich kompletności, szczególnie pod kątem opisu danych, które będą wykorzystane w projekcie;
- standardów wprowadzania informacji do systemu o funkcjonowaniu PTZ, w tym ocena zgodności i niezgodności w zakresie dotychczas wprowadzanych danych.

#### **3.2. Pozyskanie i wprowadzenie do bazy informacji o rozkładach jazdy**

W celu testowania przygotowywanego systemu, do bazy danych zaczęto ładowanie plików GTFS. Większość plików pozyskano ze źródeł internetowych. Wykorzystano do tego celu zapytania w wyszukiwarce google, wg frazy "GTFS Polska" i "GTFS w Polsce". Przejrzano także ogólnodostępne repozytoria / bazy adresowe, w tym:

- międzynarodowe:
  - <https://transitfeeds.com/l/434-poland>  
(kopia: <https://openmobilitydata.org/l/434-poland>)

- <https://database.mobilitydata.org/>
- <https://www.transit.land/operators>
- MMTIS DE: <https://mobilithek.info/>
- MMTIS DE (archiwum):  
<https://service.mdm-portal.de/mdm-portal-application/>
- MMTIS CZ: <https://data.gov.cz/datov%C3%A9-sady>
- MMTIS EE: <https://web.peatus.ee/>
- MMTIS SK: <https://odoprave.info/wps/portal/pub>
- MMTIS SK: <https://data.gov.sk/>  
(<https://data.gov.sk/en/dataset/https-www-zsr-sk-files-pre-cestujucich-cestovny-poriadok-gtfs-gtfs-zip>)
- MMTIS LT: <https://www.visimarsrutai.lt/gtfs/>
- MMTIS LV:  
<https://lvceli.lv/en/road-network/statistical-data/transport-sector-open-data/>
- ogólnopolskie:
  - Krajowy Punkt Dostępu do usług informacji o podróżach multimodalnych -  
<https://dane.gov.pl/pl/dataset/1739,krajowy-punkt-dostepowy-kpd-multimodalne-usugi-informacji-o-podrozach>
  - <https://dane.gov.pl/>
  - <https://mkuran.pl/gtfs/>
  - <https://przyjazdy.pl/gtfs> (dawny gtfs.pl)

Należy zauważyć, że w żadnym z ww. repozytoriów nie było wszystkich z ww. plików - repozytoria były wzajemnie niekompletne.

W efekcie do bazy danych wczytano dane od 52 przewoźników lub organizatorów ptz<sup>13</sup>. Uzupełniono je przygotowanymi ręcznie danymi dla publicznej linii obsługiwanej przez Średzką Kolej Powiatową. Uzyskane wyniki podsumowano w tab. 2, poniżej.

---

<sup>13</sup> W trakcie trwania prac projektowych lista ta została rozszerzona z pierwotnych 43 podmiotów.

Tab. 2. Tabela otwartych źródeł plików GTFS zawierających dane na terytorium Polski.

L.p.	Organizator/przewoźnik	Link do ostatniego gtfs static	GTFS aktualny?	Dostęp przez API?
1	Średzka Kolej Powiatowa	wprowadzony ręcznie		
2	Białostocka komunikacja miejska	<a href="http://komunikacja.bialystok.pl/cms/File/download/gtfs/google_transit.zip">http://komunikacja.bialystok.pl/cms/File/download/gtfs/google_transit.zip</a>	TAK	NIE
3	ZKM Bielawa	<a href="https://transitfeeds.com/p/przyjazdy-pl/1206/latest/download">https://transitfeeds.com/p/przyjazdy-pl/1206/latest/download</a>	NIE (2019-05-12)	NIE
4	MZK Bolesławiec	<a href="https://transitfeeds.com/p/przyjazdy-pl-boleslawiec/1209/latest/download">https://transitfeeds.com/p/przyjazdy-pl-boleslawiec/1209/latest/download</a>	NIE (2020-11-17)	NIE
5	ZDMiKP Bydgoszcz	<a href="https://zdmikp.bydgoszcz.pl/rozklady/paczka/gtfs/gtfs.zip">https://zdmikp.bydgoszcz.pl/rozklady/paczka/gtfs/gtfs.zip</a> (nieaktualne, 2022 z <a href="https://zdmikp.bydgoszcz.pl/pl/transport/rozklad-jazdy/">https://zdmikp.bydgoszcz.pl/pl/transport/rozklad-jazdy/</a> ) <a href="https://mkuran.pl/gtfs/bydgoszcz.zip">https://mkuran.pl/gtfs/bydgoszcz.zip</a> <a href="https://transitfeeds.com/p/zdmikp-w-bydgoszczy/1109">https://transitfeeds.com/p/zdmikp-w-bydgoszczy/1109</a> (nieaktualne, 2021)	NIE TAK	NIE
6	Zarząd komunikacji miejskiej w Elblągu	<a href="https://przyjazdy.pl/latest/elblag.zip">https://przyjazdy.pl/latest/elblag.zip</a>	TAK	NIE
7	MZK Elk	<a href="https://mkuran.pl/gtfs/elk.zip">https://mkuran.pl/gtfs/elk.zip</a>	TAK	NIE
8	ZTM Gdańska i ZKM Gdynia	<a href="https://mkuran.pl/gtfs/tristar.zip">https://mkuran.pl/gtfs/tristar.zip</a> (oba) <a href="https://dane.gov.pl/pl/dataset/1803/resource/2056.rozkad-jazdy-gtfs/table">https://dane.gov.pl/pl/dataset/1803/resource/2056.rozkad-jazdy-gtfs/table</a> (tylko Gdańsk) <a href="http://api.zdiz.gdynia.pl/pt/gtfs.zip">http://api.zdiz.gdynia.pl/pt/gtfs.zip</a> (tylko Gdynia)	TAK TAK	NIE
9	Gmina Zakład Komunalny (GZK Bystry)	<a href="https://mkuran.pl/gtfs/gizycko.zip">https://mkuran.pl/gtfs/gizycko.zip</a>	TAK	NIE
10	Zarząd Transportu Metropolitalnego z siedzibą w Katowicach	<a href="https://otwartedane.metropoliagzm.pl/dataset/rozklady-jazdy-i-lokalizacja-przystankow-gtfs">https://otwartedane.metropoliagzm.pl/dataset/rozklady-jazdy-i-lokalizacja-przystankow-gtfs</a>	TAK	NIE
11	ZTM Kielce	<a href="https://mkuran.pl/gtfs/kielce.zip">https://mkuran.pl/gtfs/kielce.zip</a>	TAK	NIE
12	MKP Kraków i ZTP Kraków	<a href="https://gtfs.ztp.krakow.pl/">https://gtfs.ztp.krakow.pl/</a>	TAK	NIE
13	Miejska Komunikacja Samochodowa Sp. z o.o. Krosno	<a href="https://przyjazdy.pl/latest/krosno.zip">https://przyjazdy.pl/latest/krosno.zip</a>	NIE ( )	NIE
14	Zarząd Transportu Miejskiego Lublin	<a href="https://mkuran.pl/gtfs/lublin.zip">https://mkuran.pl/gtfs/lublin.zip</a>	TAK	NIE
15	Komunikacja Miejska Łomianki	<a href="https://mkuran.pl/feed/kml/kml-latest.zip">https://mkuran.pl/feed/kml/kml-latest.zip</a>	TAK	NIE
16	Miejskie Przedsiębiorstwo Komunikacyjne w Łomży	<a href="https://mkuran.pl/gtfs/lomza.zip">https://mkuran.pl/gtfs/lomza.zip</a>	TAK	NIE
17	Miejskie Przedsiębiorstwo Komunikacyjne w Łodzi	<a href="https://przyjazdy.pl/latest/lodz.zip">https://przyjazdy.pl/latest/lodz.zip</a>	TAK	NIE

L.p.	Organizator/przewoźnik	Link do ostatniego gtfs static	GTFS aktualny?	Dostęp przez API?
18	MPK Nowy Sącz	<a href="https://transitfeeds.com/p/przyjazdy-pl-nowy-sacz">https://transitfeeds.com/p/przyjazdy-pl-nowy-sacz</a>	NIE (2020-11-16)	NIE
19	Zarząd Dróg Zieleni i Transportu Olsztyn	ftp://hq.zdzit.olsztyn.eu:3521/tts_xml/ login: tts hasło: DaneTTS1	TAK	NIE
20	Komunikacja Miejska Płock	<a href="https://transitfeeds.com/p/przyjazdy-pl-plock/1211/latest/download">https://transitfeeds.com/p/przyjazdy-pl-plock/1211/latest/download</a>	NIE (2020-01-26)	NIE
21	Zarząd Transportu Miejskiego Poznań	<a href="https://www.ztm.poznan.pl/pl/dla-deweloperow/gtfsFiles">https://www.ztm.poznan.pl/pl/dla-deweloperow/gtfsFiles</a>	TAK	TAK
22	Miejski Zakład Komunikacji sp. z o.o. w Przemyślu	<a href="https://przyjazdy.pl/latest/przemysl.zip">https://przyjazdy.pl/latest/przemysl.zip</a>	NIE ( )	NIE
23	MZDiK Radom	<a href="https://mkuran.pl/feed/radom/radom-latest.zip">https://mkuran.pl/feed/radom/radom-latest.zip</a>	TAK	NIE
24	ZTZ Rybnik	<a href="https://mkuran.pl/gtfs/rybnik.zip">https://mkuran.pl/gtfs/rybnik.zip</a>	TAK	NIE
25	Zarząd Transportu Miejskiego Rzeszów	<a href="http://193.19.167.109:4080/gtfs.zip">http://193.19.167.109:4080/gtfs.zip</a> (z transit.land)  <a href="https://mkuran.pl/gtfs/rzeszow.zip">https://mkuran.pl/gtfs/rzeszow.zip</a>  (nie wiadomo, który jest aktualniejszy, prawdopodobnie mkuran, bo taka jest data transxchange - strona ztm.erzeczow.pl oraz chmura nie działają) GTFSy przygotowane w różny sposób, ale zbliżone)	TAK? TAK?	NIE
26	Zarząd Dróg i Transportu Miejskiego Szczecin	<a href="https://www.zditm.szczecin.pl/rozklady/GTFS/latest/">https://www.zditm.szczecin.pl/rozklady/GTFS/latest/</a>	TAK	NIE
27	Komunikacja Autobusowa w Świnoujściu	<a href="https://mkuran.pl/gtfs/swinoujscie.zip">https://mkuran.pl/gtfs/swinoujscie.zip</a>	TAK	NIE
28	Urząd Miasta Torunia	<a href="http://158.75.35.18/~wgk/gtfs-TORUN.zip">http://158.75.35.18/~wgk/gtfs-TORUN.zip</a> (MMTIS) <a href="https://mkuran.pl/gtfs/torun.zip">https://mkuran.pl/gtfs/torun.zip</a>  tak na oko to samo, ale mkuran prawidłowo spakował, a w MMTIS jest spakowany folder	TAK	NIE
29	Związek Transportu Miejskiego Warszawa	<a href="https://mkuran.pl/feed/ztm/ztm-latest.zip">https://mkuran.pl/feed/ztm/ztm-latest.zip</a>  <a href="https://dane.gov.pl/pl/dataset/1803/resource/2056.rozkad-jazdy-gtfs/table">https://dane.gov.pl/pl/dataset/1803/resource/2056.rozkad-jazdy-gtfs/table</a>	TAK	NIE
30	Wielicka Komunikacja Wewnętrzna	<a href="https://otwarte.dane.malopolska.pl/devportal/apis">https://otwarte.dane.malopolska.pl/devportal/apis</a>  <a href="http://gtfs.pyrfekt.com/feeds/wieliczka_gtfs.zip">http://gtfs.pyrfekt.com/feeds/wieliczka_gtfs.zip</a>	TAK  NIE (2019-02-22)	TAK  NIE
31	MPK Włocławek	<a href="https://mkuran.pl/gtfs/wloclawek.zip">https://mkuran.pl/gtfs/wloclawek.zip</a>	TAK	NIE
32	Wydział Transportu Urzędu Miasta Wrocławia	<a href="https://www.wroclaw.pl/open-data/dataset/rozkład-jazdytransportupublicznegoplik_data">https://www.wroclaw.pl/open-data/dataset/rozkład-jazdytransportupublicznegoplik_data</a>	TAK	NIE



L.p.	Organizator/przewoźnik	Link do ostatniego gtfs static	GTFS aktualny?	Dostęp przez API?
		<a href="https://dane.gov.pl/pl/dataset/80/resource/1823/table">https://dane.gov.pl/pl/dataset/80/resource/1823/table</a>		
33	Flixbus	<a href="https://transport.data.gouv.fr/resources/11681?id=11681&amp;locale=en">https://transport.data.gouv.fr/resources/11681?id=11681&amp;locale=en</a> <a href="http://gtfs.gis.flix.tech/gtfs_generic_eu.zip">http://gtfs.gis.flix.tech/gtfs_generic_eu.zip</a>	TAK	TAK
34	PKS Poznań	<a href="http://pksonline.pl/download/pkspoznan.zip">http://pksonline.pl/download/pkspoznan.zip</a>	TAK	NIE
35	Koleje Dolnośląskie	<a href="https://mkuran.pl/gtfs/kolejedolnoslaskie.zip">https://mkuran.pl/gtfs/kolejedolnoslaskie.zip</a>	TAK	NIE
36	Koleje Małopolskie	<a href="https://www.kolejemalopolskie.com.pl/pl/rozklady-jazdy/gtfs">https://www.kolejemalopolskie.com.pl/pl/rozklady-jazdy/gtfs</a> <a href="https://dane.gov.pl/pl/dataset/3083/resource/46165.dane-gtfs-transportu-kolejowego/table">https://dane.gov.pl/pl/dataset/3083/resource/46165.dane-gtfs-transportu-kolejowego/table</a>	SKA - TAK ALD - TAK	NIE
37	Koleje Mazowieckie	<a href="https://mkuran.pl/gtfs/kolejemazowieckie.zip">https://mkuran.pl/gtfs/kolejemazowieckie.zip</a>	TAK	NIE
38	Koleje Wielkopolskie	<a href="https://przyjazdy.pl/latest/kw.zip">https://przyjazdy.pl/latest/kw.zip</a>	TAK do 2023-04-30	NIE
39	Łódzka Kolej Aglomeracyjna	<a href="https://kolej-lka.pl/pliki/pn0e6eg45qcl4hd5/gtfs-2022-2023/zip/">https://kolej-lka.pl/pliki/pn0e6eg45qcl4hd5/gtfs-2022-2023/zip/</a> <a href="https://kolej-lka.pl/pliki/pn0e6eg45qcl4hd5/gtfs-2023-2024/zip/">https://kolej-lka.pl/pliki/pn0e6eg45qcl4hd5/gtfs-2023-2024/zip/</a> <a href="https://kolej-lka.pl/pliki/pn0e6eg45qcl4hd5/gtfs-2024-2025/zip/">https://kolej-lka.pl/pliki/pn0e6eg45qcl4hd5/gtfs-2024-2025/zip/</a> <a href="https://przyjazdy.pl/gtfs">https://przyjazdy.pl/gtfs</a>	kolej - TAK przyjazdy - NIE	NIE
40	PKP Intercity	<a href="https://mkuran.pl/gtfs/pkpic.zip">https://mkuran.pl/gtfs/pkpic.zip</a>	TAK	NIE
41	PolRegio	<a href="https://mkuran.pl/gtfs/polregio.zip">https://mkuran.pl/gtfs/polregio.zip</a>	TAK	NIE
42	PKP Szybka Kolej Miejska w Trójmieście Sp. z o.o.	<a href="https://bip.skm.pkp.pl/c60/rozklad-jazdy">https://bip.skm.pkp.pl/c60/rozklad-jazdy</a> <a href="https://przyjazdy.pl/latest/skmt.zip">https://przyjazdy.pl/latest/skmt.zip</a>	TAK	NIE
43	Warszawska Kolej Dojazdowa	<a href="https://mkuran.pl/feed/wkd/wkd-latest.zip">https://mkuran.pl/feed/wkd/wkd-latest.zip</a>	TAK	NIE
44	Miejski Zakład Komunikacji w Gorzowie Wielkopolskim	<a href="https://przyjazdy.pl/latest/gorzow.zip">https://przyjazdy.pl/latest/gorzow.zip</a>	TAK	NIE
45	MZK Wejherowo	<a href="https://mkuran.pl/gtfs/wejherowo.zip">https://mkuran.pl/gtfs/wejherowo.zip</a>	TAK	NIE
46	KPA Kombus (tylko część linii - Kórnik)	<a href="http://kombus.pl/pl/deweloperzy/specyfikacja-i-gtfs/">http://kombus.pl/pl/deweloperzy/specyfikacja-i-gtfs/</a> <a href="https://kombus.pl/gtfs/kombus.zip">https://kombus.pl/gtfs/kombus.zip</a>	TAK	NIE
47	UM Częstochowa / MPK Częstochowa	<a href="https://mpk.czystochowa.pl/public/cloud/index.php/s/utbhxukj2pYM6T?path=%2FDane%20roz%C5%82adowe">https://mpk.czystochowa.pl/public/cloud/index.php/s/utbhxukj2pYM6T?path=%2FDane%20roz%C5%82adowe</a> <a href="https://mpk.czystochowa.pl/public/cloud/index.php/s/utbhxukj2pYM6T/download?path=%2FDane%20roz%C5%82adowe">https://mpk.czystochowa.pl/public/cloud/index.php/s/utbhxukj2pYM6T/download?path=%2FDane%20roz%C5%82adowe</a>	TAK (tylko ręcznie trzeba wchodzić, bo nowe dane są dodawane)	NIE

<sup>15</sup> <https://mkuran.pl/gtfs/>

<sup>16</sup> <https://przyjazdy.pl/gtfs>

3. Niektórzy dostawcy danych nie przyjęli dobrej praktyki utrzymywania jednej ścieżki do najbardziej aktualnego pliku ("latest"), lecz wskazywali w repozytoriach ścieżki do plików aktualnych dla określonego przedziału czasu. Stąd każdorazowo w przypadku pozyskania nieaktualnych danych zespół autorski dokonywał dodatkowej weryfikacji. Dzięki temu udało się pozyskać bardziej aktualne dane dla Łódzkiej Kolei Aglomeracyjnej.
4. Standard GTFS, poza podstawową i zaakceptowaną przez konsorcjum *mobilitydata*, wersją, zawiera także różne propozycje rozszerzeń, zarówno do oficjalnej części, jak i opracowane przez podmioty trzecie. W pozyskanych paczkach pojawiały się takiego typu pola, jednak nie miało to większego wpływu na możliwość wczytania danych - nierozpoznane pola są zazwyczaj ignorowane przez system wczytujący dane.

### 3.3. Ocena zgodności danych ze standardem GTFS

Przeprowadzona analiza pozyskanych cyfrowych danych o rozkładach jazdy, a także wstępny przegląd znanych danych, które występowały w formie analogowej, wskazały na następujące problemy z wykorzystaniem standardowych plików zapisanych w standardzie GTFS, jak również problemy z konkretnymi plikami pobranymi dla polskich organizatorów lub przewoźników:

- Brak odzwierciedlenia wielu organizatorów w jednym pliku GTFS - zazwyczaj 1 plik GTFS jest przygotowywany przez jednego organizatora (a czasem przewoźnika). Przygotowywany w ramach projektu system IT powinien móc obsługiwać wielu organizatorów. W tym celu zaproponowano utworzenie dodatkowej tabeli, która zawierałaby informacje o danych kontaktowych oraz o zasięgu działalności.
- Problemy z obsługą przesiadek - w przypadku większości pobranych GTFS-ów, w składzie archiwum nie było pliku "transfers.txt" lub był on pusty (same nagłówki tabeli). Nawet istnienie takiego pliku nie rozwiązuje do końca problemu przesiadek, gdyż plik ten zawiera jedynie informacje o przesiadkach pomiędzy kursami pojedynczego organizatora. Stąd w ramach obliczeń wskaźnika zagrożeniem wykluczeniem komunikacyjnym wymagane może być tworzenie dodatkowych tabel przesiadek.

Dodatkowym utrudnieniem jest tu brak jednolitego nazewnictwa przystanków, co zostało szerzej opisane w kolejnej tabeli.

- Brak występowania informacji o cenach biletów - takowe znajdowały się jedynie w przypadku nielicznych GTFS-ów pobranych dla Ełku, Wieliczki, Warszawy czy Radomia. Ponadto, w opinii autorów większym problemem jest sam standard opisu taryf w GTFS (zarówno w wersji 1, jak i rozszerzonej wersji 2). Wymagane mogą być w tym zakresie dodatkowe prace standaryzacyjne.
- Dość ubogi opis dostosowania dla osób ze specjalnymi potrzebami. Poza parametrami *wheelchair\_boarding* czy *wheelchair\_accessible*, które przy pomocy jednej wartości opisują przystosowanie dla osób ze specjalnymi potrzebami, odpowiednio przystanków oraz kursów, standard GTFS przewiduje możliwość stosowania pliku *pathways.txt* (np. w obrębie dworca). Plik ten jednak nie był stosowany w pozyskanych GTFS-ach. Stąd także w tym przypadku mogą być wymagane dalsze prace standaryzacyjne.
- Podawanie do repozytoriów GTFS-ów (na przykład Krajowego Punktu Dostępu do usług informacji o podróżach multimodalnych) odnośników z zapisaną w ścieżce konkretną datą, zamiast niezmiennego odnośnika do wersji aktualnej (*latest*). w efekcie dane w repozytoriach mogą być nieaktualne, a pozyskanie danych w wersji aktualnej wymaga podjęcia dodatkowych, niekoniecznie prostych czynności, jak ręczna modyfikacja odnośnika w oparciu o przewidywaną nazwę aktualnego odnośnika.
- W zależności od rodzaju dostawcy GTFS, występują różne podejścia do wykorzystania podziału na linie ("routes") i kursy ("trips"). W szczególności w GTFS-ach kolejowych zdarza się, że wszystkie kursy są przypisane do jednej/kilku linii, które opisują rodzaj usługi (np. pociągi pospieszne, zastępcza komunikacja autobusowa, itd.). Może wpływać to na wykorzystanie tych danych, np. poziom dokładności statystyk dotyczących liczby linii i kursów<sup>17</sup>.

---

<sup>17</sup> Należy przy tym zauważyć, że podobne dość swobodne podejście do określania tego, czym jest linia, może mieć miejsce też w rzeczywistych informacji pozyskiwanych od organizatorów czy przewoźników - zostanie to sprawdzone w zadaniach. 6-9.

### 3.4. Ocena standardów wprowadzania informacji do systemu o funkcjonowaniu PTZ

Jak już wspomniano w poprzednim podrozdziale, przeprowadzona przez zespół autorski analiza wskazała na grupę problemów w zakresie standardu samej informacji, która ma być wprowadzana do konstruowanego systemu IT. Poniżej przedstawiono zestaw uwag będących podstawą do standaryzacji informacji:

- Bardzo zróżnicowane nazewnictwo tych samych przystanków przez różne podmioty, także różnych słupków w ramach grup przystanków. Problemem jest tu brak centralnej bazy informacji o przystankach, w efekcie czego nazwy określone przez zarządzających przystankami są bardzo rozproszone i nieustandaryzowane. Ponadto stwierdzono, że niektórzy zarządcy mogą nie dysponować aktualną informacją o przystankach. W efekcie jeden przystanek może występować w rozkładach jazdy pod wieloma nazwami, które są nadawane w myśl różnych standardów (nazwa ulicy, przy której znajduje się przystanek z ewentualnym kolejną liczbą porządkową; nazwa ulicy poprzecznej wobec tej, przy której znajduje się przystanek; nazwa związana z punktem charakterystycznym, który znajduje się w pobliżu). Brak standaryzacji w tym zakresie ma wpływ na mechanizmy aktualizacji danych rozkładowych. Ostatecznie ustalono, że aktualizacja GTFS-ów nie powinna kasować dotychczasowych przystanków, lecz co najwyżej aktualizować wybrane atrybuty, gdyż przystanek dodany przez GTFS może następnie być wykorzystany także przez innego organizatora (linia dodana ręcznie). Sam mechanizm aktualizacji atrybutów także powinien uwzględniać sytuacje, w których poprawiony "ręcznie" przystanek może zostać nadpisywany przez starszą, niepoprawioną wersję, podczas ponownego importu danych zewnętrznych<sup>18</sup>. Ewentualne usuwanie niewykorzystywanych przystanków będzie wymagać oddzielnych algorytmów. Natomiast obsługa przesiadek powinna być oparta na algorytmach grupujących - nie ma możliwości rozpoznania po nazwie, kodzie przystanku (ten sam przystanek może mieć różne nazwy / kody u różnych organizatorów). Stwierdzono

---

<sup>18</sup> Na poziomie automatycznych mechanizmów importujących trudno jest zawrzeć logikę, która w sposób niezawodny będzie oceniać, które dane są właściwe - stare czy nadpisywane. Nawet w przypadkach, gdy nadpisywane dane będą w danych polach puste (tj. wartości NULL).

także, że należy przygotować procedurę rozstrzygania spornego nazewnictwa i rozpowszechnić ją w obrębie zespołu projektowego.

- Stosowanie oznaczeń "kursuje w dni nauki szkolnej" - brak bezpośredniego odzwierciedlenia w GTFS, ale możliwe odwzorowanie przy pomocy kalendarza (najlepiej przez przygotowanie słowniczka). Jednak problematyczne są kwestie ferii zimowych, które mogą mieć różne terminy w różnych województwach - dotyczy to w szczególności linii przecinających granice województw.
- Podobny problem występował z innymi symbolami, np. w zależności od organizatora stosowane była legenda "Nie kursuje 24.XII" lub "nie kursuje w Wigilię Bożego Narodzenia". Jakkolwiek w standardzie GTFS istnieje możliwość nadania różnych nazw de facto tym samym kalendarzom, należy zastanowić się czy takie posunięcie ma sens. Ponadto w GTFS niekoniecznie funkcjonują wszystkie pola, które pozwoliłyby na kompletne zakodowanie oznaczeń kalendarza (np. "oznaczenie literowe" i "pełna nazwa"). Możliwe, że będzie w tym miejscu konieczne rozszerzenie standardu.
- Podawanie dla jednego przystanku kursów w obu kierunkach, rozróżnienie kierunków np. przez literkę "A" - propozycja powinna zakładać tworzenie dodatkowego przystanku (po drugiej stronie jezdni, jeśli fizycznie nie ma).
- Inne, drobne, takie jak podawanie dodatkowego adresu www w polu email.

## **4. Analityczne przetestowanie rozwiązań kreatorów dodawania danych o funkcjonowaniu PTZ**

### **4.1. Ocena istniejących kreatorów danych o funkcjonowaniu PTZ**

W ramach Działania D3 analityczne przetestowano (koszt wdrożenia, spełnianie potrzeb, poziom interoperacyjności z systemami informatycznymi organizatorów PTZ i przewoźników) rozwiązań kreatorów umożliwiających przygotowanie lub cyfryzację danych o funkcjonowaniu PTZ, zgodnie ze standardem GTFS oraz narzędzi umożliwiających import rezultatów do bazy danych.

W celu analizy kreatorów, oprócz wskazania znanych wcześniej w konsorcjum rozwiązań, przejrano także wyniki zapytania "GTFS creator" w wyszukiwarce Google. Dokonano przeglądu stron internetowych zbierających dane o standardach GTFS i NeTeX. Przejrano m.in.:

- <https://gtfs.org/resources/>
- <https://developers.google.com/transit/gtfs>
- <https://project-awesome.org/CUTR-at-USF/awesome-transit>
- <https://github.com/>

Pozyskane w ten sposób informacje o narzędziach pozwalają podzielić owe narzędzia na cztery podstawowe grupy:

1. Komercyjne programy dla organizatorów lub planistów transportu, które pozwalają na eksport informacji o transporcie zbiorowym do standardu GTFS, np.:
  - a. PTV VISUM
  - b. AGC Busman CB
  - c. DPK Cityliner
  - d. kiedyprzyjedzie.pl
2. Usługi webowe, np.:

- a. WatriFeed  
(<https://digitaltransport4africa.org/watrideed-the-open-source-collaborative-gtfs-data-editor/>)
- b. GTFS Editor (<https://addtransit.com/gtfs-transit-file.php>)
- 3. Opensourcowe narzędzia do samodzielnej instalacji, np.:
  - a. IBI Data Tools (<https://github.com/ibi-group/datatools-ui>)
  - b. WatriFeed (<https://github.com/DataTransport/WatriFeed>)
- 4. Opensourcowe narzędzia wykorzystujące arkusze kalkulacyjne, np.
  - a. GTFS Builder (<https://www.nationalrtap.org/Technology-Tools/GTFS-Builder>)
  - b. xlstoGTFS (<https://github.com/ChamGeeks/xlstoGTFS>)
  - c. gtfs-xlsx (<https://github.com/thzinc/gtfs-xlsx>)

Narzędzia z pierwszej grupy mają dość zaawansowaną funkcjonalność w zakresie przygotowywania danych niezbędnych dla plików GTFS. Nie zawsze jest ona jednak pełna (np. brak eksportu danych o taryfach, pomimo możliwości kodowania tej informacji). Z perspektywy projektu ich kolejną wadą są wysokie koszty licencji, związane z szeregiem funkcjonalności, które nie będą wykorzystywane w projekcie, takich jak makroskopowe modelowanie transportu. Narzędzia te mogą mieć także problemy z modelowaniem niestandardowych rozwiązań publicznego transportu zbiorowego mogącymi występować w Polsce, jak np. rozwiązania typu demand responsive, jak krakowski “Tele Bus”, czy szczeciński “Transport na żądanie”.

Narzędzia z drugiej grupy mogą być dostępne za darmo (WatriFeed) lub za opłatą (GTFS Editor). Wymagają jednak logowania oraz są dostosowane do określonych obszarów, takich jak kraje afrykańskie. Stąd o ile narzędzia te mogą być źródłem inspiracji, nie nadają się do wykorzystania w projekcie T-Included. Ponadto podobnie, jak narzędzia z pierwszej grupy, mogą występować tu problemy z modelowaniem niestandardowych rozwiązań publicznego transportu zbiorowego.

Narzędzia z trzeciej i czwartej grupy, to rozwiązania typu open-source. Jednak reprezentują one zazwyczaj uproszczone lub częściowe podejście. Np. jedno narzędzie usprawnia edycję tabel GTFS, podczas gdy inne umożliwia ich walidację, a jeszcze inne pozwala na automatyzację odwzorowywania trasy przejazdu środka transportu. A zatem



pojedynczo nie umożliwiają one kompleksowego podejścia do wprowadzania informacji o publicznym transporcie zbiorowym. Mogą jednak przyspieszyć tworzenie takiego.

Należy zauważyć, że narzędzia inne niż przynależne do drugiej grupy mogą mieć dodatkowe wymagania dotyczące platform sprzętowych i systemów operacyjnych. Przykładowo przeprowadzone przez zespół projektowy testy narzędzia GTFS Builder wskazały, że nie jest on kompatybilny z arkuszem kalkulacyjnym Excel dla komputerów Mac przez wzgląd na wykorzystywanie specjalistycznych makr. Podobnie część oprogramowania do samodzielnej instalacji może wymagać systemu operacyjnego Microsoft Windows. Ponadto sam wymóg instalacji może być poważną przeszkodą wdrożenia w niektórych organizacjach przez wzgląd na prowadzoną politykę w zakresie cyberbezpieczeństwa. Z kolei narzędzia webowe także mogą mieć pewne ograniczenia w zakresie wymogów co do wykorzystywanej przeglądarki internetowej. Są one jednak mniejsze m.in. dlatego, że większość rozwiązań w tym zakresie jest dostępne za darmo dla użytkownika końcowego, co ułatwia ich ewentualne wdrożenie. Biorąc to pod uwagę oraz potrzebę zapewnienia możliwości jednoczesnej pracy wielu użytkowników na wspólnej bazie, narzędzia webowe powinny być rozwiązaniem preferowanym do wdrożenia w projekcie.

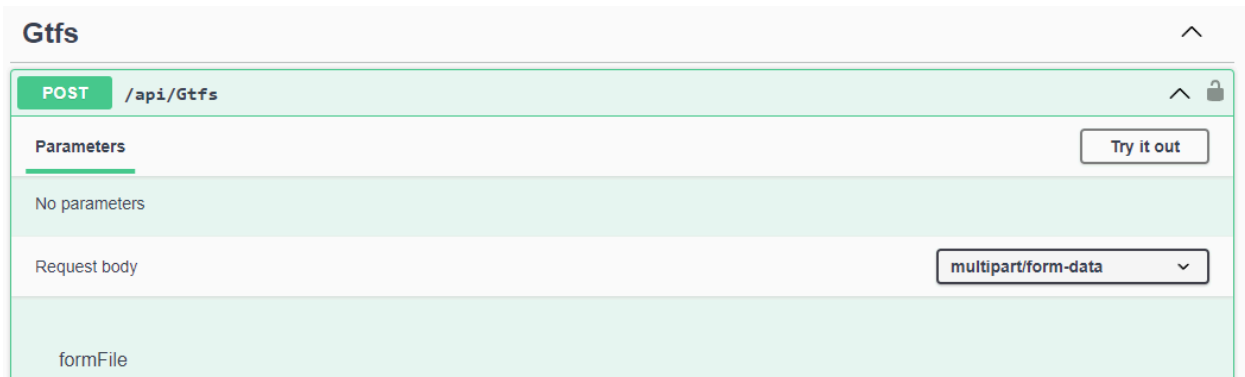
Ponadto dość krytyczną wadą wszelkich przejranych grup narzędzi jest to, że są one dostosowane do pracy z perspektywy pojedynczego organizatora lub przewoźnika. Nie spełnia zatem wymagań projektowych, gdzie założono wprowadzanie danych od wielu organizatorów na raz, przy założeniu że w rozkładach jazdy występują pewne elementy wspólne, np. możliwość korzystania z usług tych samych przewoźników czy stosowanie zbliżonych rozwiązań w zakresie przygotowywania rozkładów jazdy, np. informacja o kursowaniu jedynie w dni nauki szkolnej.

Podsumowując, stwierdzono brak występowania jednego narzędzia, które spełniałoby na raz wszystkie wymagania projektowe. Dostępne narzędzia mogą pełnić w projekcie co najwyżej rolę pomocniczą. Stąd postanowiono rozwinąć własne narzędzie przy możliwie maksymalnym wykorzystaniu dostępnych rozwiązań fragmentarycznych opensource (walidatory, etc.).

#### **4.2. Opracowanie wstępnej wersji autorskiego narzędzia pozwalającego na cyfryzację analogowych danych o funkcjonowaniu PTZ**

Prace prowadzone w ramach działań D2 i D3 pozwoliły na przygotowanie wstępnej wersji narzędzia, które pozwala na cyfryzację danych analogowych o rozkładach jazdy. Narzędzie to ma charakter usługi dostępnej przez Internet. Zostało uruchomione na serwerze wybranym i zakupionym na podstawie prac prowadzonych w D1. Narzędzie składa się z dwóch głównych warstw:

- frontendowej - aplikacja www, która pozwala na ręczne wprowadzanie danych oraz łatwy (przez wybór z dysku) import plików gtfs z innych źródeł - zarówno zewnętrznych wobec projektu (dane pozyskane w tym standardzie od przewoźników i organizatorów), jak i wewnętrznych (dane wprowadzone i przekonwertowane do GTFS-ów przy wykorzystaniu posiadanych przez konsorcjantów narzędzi, jak np. PTV VISUM).
- backendowej - baza danych pozwalająca przechowywać dane niezbędne do przygotowania pliku GTFS wraz z dodatkowymi danymi, które nie są uwzględnione w standardzie, a mają duże znaczenie z perspektywy projektu oraz organizacji PTZ w Polsce (np. dodatkowa tabela wskazująca na organizatora transportu). Oprócz aplikacji frontendowej, backend zapewnia dostęp do danych przez dwa dodatkowe interfejsy:
  - API (standard JSON dla pojedynczych tabel lub plik binarny .zip dla kompletnych danych GTFS), patrz rys. 1, poniżej,
  - edytor tabel SQL (np. DBeaver).



Rys. 1. Zrzut ekranu, pokazujący fragment API, który umożliwia wczytywanie plików GTFS

Ponadto narzędzie to zostało wsparte dwoma komponentami, które zapewniają możliwość cyfryzacji danych niestandardowych lub wymagających prac standaryzacyjnych przewidzianych w Zad. 10, tj.:

- bazę danych przestrzennych (Post-GIS) dla danych punktowych, liniowych i poligonowych,
- dysk sieciowy dla innego rodzaju dokumentów.

Podsumowując, narzędzie to pozwala na wczytywanie GTFS-ów z poziomu GUI lub backendu. Zawiera także komponenty w postaci dysku sieciowego oraz bazy danych GIS (Postresql+PostGIS), która umożliwia np. uwzględnianie w projekcie rzadko występujących rozwiązań, takich jak transport "na żądanie" (opis w postaci poligonu z zasięgiem działania przy dodatkowych atrybutach, jak np. koszt na km). Dostęp do tych komponentów będzie odbywał się z wykorzystaniem standardowych klientów dysków sieciowych czy GIS, np. opensourcowego QGISa.

Narzędzie informatyczne zostało udostępnione wszystkim konsorcjantom dzięki zapewnieniu dostępu VPN. Forma ta z jednej strony zapewnia duży poziom bezpieczeństwa serwera oraz danych przed atakiem cybernetycznym, co ma duże znaczenie szczególnie na początkowych etapach prac. Z drugiej strony pod względem technicznym proste jest szersze udostępnienie tego narzędzia. Należy zauważyć, że zaletą udostępniania aplikacji przez przeglądarkę internetową jest zapewnienie wyższej kompatybilności z różnymi

systemami (w porównaniu do np. rozwiązań bazujących na arkuszach kalkulacyjnych z makrami) czy sam brak wymogu instalacji.

## 5. Ocena osiągnięcia kamieni milowych

Według założeń przedstawionych we wprowadzeniu do niniejszego raportu, autorzy założyli osiągnięcie trzech kamieni milowych. Poniżej przedstawiono argumentację dającą podstawę potwierdzenia ich osiągnięcia.

- KM3.1 - Wyniki laboratoryjnego testowania bazy danych informacjami o funkcjonowaniu PTZ poprzez pozyskanie i wprowadzenie do bazy informacji o rozkładach jazdy od minimum 20 organizatorów PTZ lub przewoźników.

W rozdz. 3.2 przedstawiono listę 52 organizatorów PTZ i przewoźników działających na polskim rynku, dla których pozyskano dane o funkcjonowaniu PTZ - w standardzie GTFS. Dane zostały zweryfikowane, wprowadzone do zbudowanego systemu (bazy danych) i przeprowadzono testy laboratoryjne. **Komplet tych działań potwierdza osiągnięcie KM3.1.**

- KM3.2 - Wstępna diagnoza dotycząca potrzeb dostosowania standardu GTFS oraz potrzeb standaryzacyjnych wprowadzania danych o PTZ w związku z realizacją projektu. Szczegółowe parametry tego kamienia obejmują minimum wskazanie problemów i sposobu rozwiązania dla danych pozyskanych od minimum 20 organizatorów lub przewoźników PTZ, przy czym:
  - 1 lista w zakresie oceny zgodności danych ze standardem GTFS.

W rozdz. 3.3 przedstawiono listę zawierającą 6 fundamentalnych uwag dotyczących zgodności danych z przyjętym standardem GTFS.

- 1 lista w zakresie standardów wprowadzania informacji do systemu o funkcjonowaniu PTZ.

W rozdziale 3.4 przedstawiono listę zawierającą 5 elementów diagnozy z zakresu potrzeb dostosowania standardu GTFS oraz potrzeb standaryzacyjnych wprowadzania danych o PTZ na potrzeby projektu.

Opracowanie obu list zawierających uwagi w zakresie zgodności danych ze standardem GTFS i potrzeb dostosowania standardu GTFS oraz potrzeb standaryzacyjnych wprowadzania danych o PTZ stanowi wypełnienie założeń.

**Kamień milowy KM3.2 można zatem uznać za osiągnięty.**

- KM3.3 - Opracowanie wstępnej wersji narzędzia pozwalającego na cyfryzację danych analogowych. Za parametr oceny tego kamienia przyjęto utworzenie minimum 1 interfejsu użytkownika umożliwiającego prowadzenie prac przez wszystkich konsorcjantów.

W rozdz. 4.2 przedstawiono podstawy budowy autorskiego kreatora cyfryzacji analogowych informacji o funkcjonowaniu PTZ, tj. rozkładów jazdy. Stanowi on realizację rozwiązania eliminującego większość ze wskazanych niedomagań istniejących rozwiązań lub systemów. W efekcie wszyscy członkowie Konsorcjum uzyskali interfejs, który pozwala na cyfryzację danych analogowych, przewidzianą do realizacji w Zad. 6-9. W świetle powyższych rozważań należy uznać, że **KM3.3. został osiągnięty.**

## 6. Podsumowanie

W ramach prac projektowych został zbudowany prototyp autorskiego kreatora cyfrowych rozkładów jazdy będący podsystemem gromadzenia informacji o funkcjonowaniu PTZ, który był następnie rozwijany w ramach realizacji zad. 10. Spełnia on szereg cech, wśród których najważniejsze to:

- forma aplikacji webowej - zlokalizowany na serwerze w ramach infrastruktury informatycznej Politechniki Poznańskiej;
- w celu ograniczenia kosztów cybersecurity w budżecie projektu, podczas jego trwania, korzystanie z kreatora wymagało dostępu przez VPN (Virtual Personal Network) PP, co stanowiło pierwszy poziom weryfikacji użytkownika i podstawowe zabezpieczenie dostępu - drugim był wymóg zalogowania się w samej aplikacji;
- szczegółowe procedury dostępowe zostały przedstawione w raporcie z Zad. 20;
- szczegółowa instrukcja korzystania z kreatora wraz z materiałem szkoleniowym przedstawione są w raporcie Zad. 21;
- kreator jest gotowym rozwiązaniem, które można w pełni wdrożyć dla wszystkich organizatorów PTZ i przewoźników, którzy nie posiadają własnych rozwiązań do tworzenia cyfrowych rozkładów jazdy w standardzie GTFS, zapewnia przy tym import i eksport plików GTFS;
- w przypadku konieczności korzystania ze standardu NeTEx kreator należy rozbudować o konwerter standardu GTFS na NeTEx, co wymaga realizacji dodatkowych prac; w trakcie trwania prac projektowych zespół projektowy nie otrzymał z ani jednego źródła danych w standardzie NeTEx.

Opracowanie kreatora cyfrowych rozkładów jazdy z jednej strony pozwoliło na stworzenie platformy do sprawnego przetwarzania analogowych danych o funkcjonowaniu PTZ do postaci cyfrowej, według standardu GTFS. Jest to efekt stanowiący istotny wkład w realizację Zad. 6-9, prowadzonych przez wszystkich konsorcjantów z zachowaniem zasady jednoczesnej możliwości pracy z systemem. Z drugiej strony budowa takiego rozwiązania pozwoliła na wyeliminowanie szeregu niedoskonałości i barier jakie stwierdzono w trakcie przeglądu testowego zbioru gotowych rozwiązań, zarówno tych prostych i darmowych, jak i

dedykowanych rozwiązań, stanowiących integralną częścią większych systemów informatycznych.

Wiedza zdobyta w trakcie realizacji tego zadania, w szczególności w obszarze pozyskiwania gotowych zbiorów GTFS (niezbędnych do testowego wprowadzenia do budowanego systemu), jak i standaryzacji informacji, pozwoliła na sformułowanie następujących konkluzji natury merytorycznej i formalnej:

- W przestrzeni publicznej, tj. w otwartych repozytoriach, istnieją zbiory GTFS które charakteryzują się niespójnością i wymagają stosowania szczegółowej walidacji i korekty przed ich dalszym wykorzystaniem. Istnieje zatem konieczność wprowadzenia procedur prowadzenia szczegółowej walidacji i korekt zbioru GTFS przed ich udostępnieniem.
- Podstawą utworzenia zbioru GTFS jest, a właściwie powinna być, ustandaryzowana informacja o przystankach na których odbywa się realizacja rozkładów jazdy. Przygotowanie w pełni wiarygodnych informacji o funkcjonowaniu PTZ wymaga najpierw ustalenia wspólnej - jednolitej bazy nazw i lokalizacji przystanków, która powinna być bazą:
  - **ogólnodostępną**, tj. upowszechnioną na potrzeby wszystkich jednostek zajmujących się planowaniem transportu zbiorowego o statusie PTZ/PR;
  - **administrowaną i aktualizowaną** przez wskazane jednostki, np. zarządców infrastruktury drogowej, co zapewni zgodność jej zapisów ze stanem faktycznym;
  - **wymagalną** przy tworzeniu cyfrowych rozkładów jazdy, co zapewni spójność nazewnictw i jednoznaczną identyfikację przystanków stanowiących podstawę budowy linii PTZ/PR.
- Rozkłady jazdy w standardzie cyfrowym (w projekcie przyjęto GTFS) powinny być:
  - **tworzone obowiązkowo** przez wszystkich organizatorów lub operatorów PTZ/PR, jako podstawowy element planowania działalności transportowej - realizacji powierzonych zadań;
  - **publicznie udostępniane**, w postaci krajowego punktu dostępowego, co pozwoli na rozwój szeregu funkcjonalności w obszarze informacji pasażerskiej (dostępna informacja o aktualnym rozkładzie, budowę multimodalnych planerów podróży, sprzedaż biletów itp.);



- **walidowane** przed ich zastosowaniem i udostępnieniem, co zapewni pełną zgodność ze standardem i spójność zawartej informacji; w tym zakresie zaleca się wprowadzenie jednostki certyfikującej zgodność stosowanych procedur cyfryzacji i ostatecznej postaci zbioru GTFS z obowiązującym standardem i stanem faktycznym;
- **koordynowane** w zakresie utrzymania ich aktualności oraz wspólnych elementów, takich jak np. sieć przystanków; koordynacja (w tym kuracja danych) powinna być wykonywana najlepiej przez podmioty, które z jednej strony posiadają niezbędne zasoby, a z drugiej strony są możliwie blisko źródeł powstawania danych (w celu ułatwienia i obniżenia kosztów ich weryfikacji); w przypadku dwóch ostatnich punktów autorzy sugerują szczebel samorządu wojewódzkiego, jako stanowiący najlepszy kompromis między bliskością do źródeł, a wymaganymi zasobami i kompetencjami.
- Istnieje określony zestaw informacji o funkcjonowaniu PTZ, który wymaga szczególnej uwagi przy tworzeniu ustandaryzowanych cyfrowych rozkładów jazdy, w tym m.in:
  - definicja kalendarzy,
  - uwzględnianie przesiadek,
  - definiowanie cenników biletów,
  - informacja o dostosowaniu do osób ze specjalnymi potrzebami,
  - jednoznaczne stosowania podziału na linie i kursy,
  - format zapisu nazwy i aktualności bioru GTFS.
- W raporcie zwrócono uwagę na niezgodności standaryzacyjne w zakresie informacji wprowadzanych do zbiorów GTFS, co narzuca:
  - konieczność ujednolicenia nazewnictwa przystanków (wymaga to zastosowania wspomnianej wcześniej referencyjnej bazy przystanków),
  - stosowania ujednoliconych, tj. spójnych znaczeniowo dla całego kraju, kalendarzy obowiązywania rozkładów,
  - stosowanie jednoznacznego rozróżnienia kierunku obowiązywania rozkładu jazdy,
  - jednoznacznej identyfikacji organizatora i operatora linii PTZ/PR.