

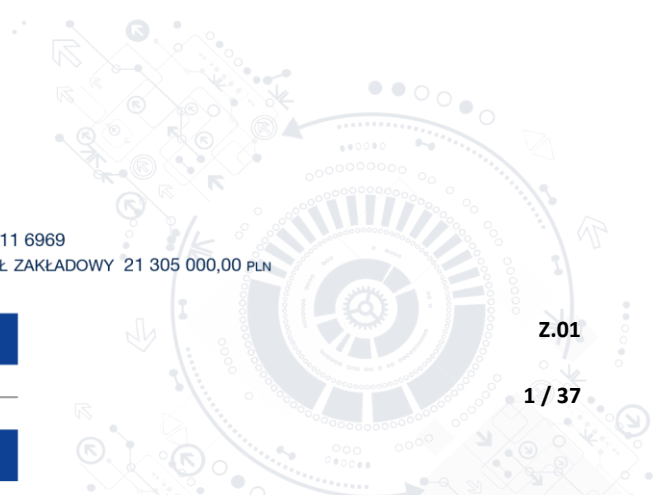
# Założenia do Koncepcji Migracji

INNOBALTICA SP. Z O.O. 80-067 GDAŃSK, UL. RÓWNA 19/21  
NIP 957-10-03-404, REGON 220639884, BANK PKO BP SA 63 1020 1811 0000 0102 0311 6969  
KRS 0000311943, VII WYDZIAŁ GOSPODARCZY KRAJOWEGO REJESTRU SĄDOWEGO KAPITAŁ ZAKŁADOWY 21 305 000,00 PLN



Rzeczpospolita  
Polska

Unia Europejska  
Fundusz Spójności



1	Wprowadzenie .....	5
1.1	Cel dokumentu .....	5
1.2	Definicje pojęć .....	6
2	Zakres przedmiotowy migrowanych danych .....	7
3	Bezpieczeństwo danych .....	8
4	Systemy źródłowe i docelowe .....	9
5	Warunki selekcji migrowanej grupy klientów zakres migrowanych danych .....	10
5.1	Wybór systemów i obiektów do migracji .....	10
5.1.1	Systemy źródłowe dla migracji .....	10
5.1.2	Dane podlegające migracji .....	11
5.1.3	Dane wyłączone z migracji .....	11
5.1.4	Obiekty podlegające migracji z perspektywy organizatora .....	11
5.1.5	Obiekty podlegające migracji z perspektywy pasażera .....	11
5.2	Zakres źródłowy i docelowy migrowanych danych .....	12
5.2.1	Zakres danych migrowanych z systemów źródłowych do systemu docelowego .....	12
5.2.1.1	Użytkownik - dane podlegające migracji <b>Error! Bookmark not defined.</b>	
5.2.1.2	Karta (Nośnik) - dane podlegające migracji <b>Error! Bookmark not defined.</b>	
5.2.1.3	Podróż - dane podlegające migracji <b>.. Error! Bookmark not defined.</b>	
5.2.1.4	Płatność - dane podlegające migracji <b>Error! Bookmark not defined.</b>	
5.2.1.5	Posiadane środki - dane podlegające migracji <b>Error! Bookmark not defined.</b>	
5.2.1.6	Dostępne usługi - dane podlegające migracji (taryfy) .....	12
5.2.1.7	Przystanki - dane podlegające migracji (PKP) .....	12
5.2.1.8	Przystanki - dane podlegające migracji (inne niż PKP) .....	12
5.2.1.9	Trasy - dane podlegające migracji .....	12
5.2.1.10	Rozkłady jazdy - dane podlegające migracji .....	12
6	Środowiska migracji danych (logiczne) .....	13
6.1	Struktura danych w obiektach .....	15

INNOBALTICA SP. Z O.O. 80-067 GDAŃSK, UL. RÓWNA 19/21

NIP 957-10-03-404, REGON 220639884, BANK PKO BP SA 63 1020 1811 0000 0102 0311 6969

KRS 0000311943, VII WYDZIAŁ GOSPODARCZY KRAJOWEGO REJESTRU SĄDOWEGO KAPITAŁ ZAKŁADOWY 21 305 000,00 PLN



Rzeczpospolita  
Polska

Unia Europejska  
Fundusz Spójności



Z.01

2 / 37



6.2	Środowisko Docelowe .....	15
6.3	Środowisko developerskie.....	15
6.4	Środowiska Testowe .....	15
6.5	Środowisko UAT .....	15
6.6	Środowisko Produkcyjne .....	15
7	Koncepcja testów migracji (próbkiowanie) .....	18
8	Dostęp do danych źródłowych (m.in. zidentyfikowane problemy prawne z dostępem do danych) określone w Założeniach do Koncepcji Migracji.....	19
9	Zasady podziału zadań w trakcie Migracji Danych .....	20
10	Organizacja Zespołu Migracji Danych .....	21
10.1	Zespoły Zamawiającego i Wykonawcy .....	21
11	Opis zakresu migrowanych danych .....	22
12	Produkty strumienia Migracji danych (z podziałem i powiązaniem) .....	23
12.1	Produkty migracji danych dostarczane przez Wykonawcę .....	23
12.2	Produkty migracji danych DOSTARCZANE PRZEZ Zamawiającego .....	23
13	Schemat wykonania Migracji Danych .....	24
13.1	Migracja Testowa .....	24
13.1.1	Przygotowanie .....	24
13.1.2	Przeprowadzenie .....	25
13.2	Migracja UAT .....	25
13.3	Migracja Produkcyjna .....	26
14	Algorytm odcięcia danych (dane historyczne) .....	27
14.1	Wybór scenariusza migracji „Z bilansem otwarcia” .....	27
15	Reguły transformacji i walidacji danych .....	28
15.1	Główne zasady transformacji .....	28
15.1.1	Reguły wyodrębnienia zakresu danych oraz reguły walidacji i transformacji danych .....	29
15.1.2	Słowniki mapujące .....	30
15.1.3	Unikanie tworzenia „sztucznych” obiektów .....	30
15.1.4	Reguły uzupełniania brakujących danych .....	30
15.1.5	Unikanie potencjalnie „niebezpiecznego” czyszczenia danych .....	30

15.2	Główne etapy Walidacji Danych .....	31
15.2.1	Zaczytywania ekstraktów do bazy STAGE .....	31
15.2.2	Transformacji.....	31
15.2.3	Loadu .....	31
16	Reguły uzupełniania brakujących danych, niewymaganych w systemach źródłowych, a wymaganych w systemie PZU.....	32
16.1	Zakres działań związanych z POPRAWĄ, JAKOŚCI danych.....	32
16.2	Czyszczenie biznesowe danych – poza zakresem Wykonawcy .....	32
16.3	Postępowanie z brakującymi lub błędnymi danymi .....	33
17	Reguły tworzenia danych referencyjnych podczas migracji.....	34
18	Reguły obsługi błędnych danych, znajdujących się w systemach źródłowych i migrowanych do systemu DOCELOWEGO.....	35
19	Kryteria akceptacji poprawności migracji .....	36
19.1	Sprawdzenie kryteriów rozpoczęcia migracji.....	36
19.2	Wycofanie migracji .....	37

## SPIS TABEL

Tabela nr 1 - Lista do docelowego oszacowania wolumenu migrowanych danych .....	22
---	----

## SPIS DIAGRAMÓW

Środowisko migracji obrazuje Diagram nr 1 - Architektura fizyczna Środowiska Migracji Danych i jego otoczenia - struktury bazodanowe. ....	13
Diagram nr 2 - Architektura fizyczna Środowiska Migracji Danych i jego otoczenia - struktury bazodanowe.....	14

# 1 WPROWADZENIE

## 1.1 CEL DOKUMENTU

Celem dokumentu jest stworzenie i przedstawienie założeń do koncepcji migracji danych, na podstawie której zostanie opracowana Koncepcja Migracji i Plan Migracji oraz będzie przeprowadzona migracja danych z systemów

INNOBALTICA SP. Z O.O. 80-067 GDAŃSK, UL. RÓWNA 19/21  
NIP 957-10-03-404, REGON 220639884, BANK PKO BP SA 63 1020 1811 0000 0102 0311 6969  
KRS 0000311943, VII WYDZIAŁ GOSPODARCZY KRAJOWEGO REJESTRU SĄDOWEGO KAPITAŁ ZAKŁADOWY 21 305 000,00 PLN



Rzeczpospolita  
Polska

Unia Europejska  
Fundusz Spójności



## 1.2 DEFINICJE POJĘĆ

Definicje pojęć wykorzystywanych w dokumentach projektowych:

- **Zamawiający:** InnoBaltica Sp z o.o.
- **Wykonawca:** Podmiot wybrany w przetargu
- **Operatorzy Transportu** – organizacje, zapewniające transport na danym obszarze
- **Karta** – elektroniczny nośnik biletów jednorazowych lub okresowych
- **ETL** (ekstrakt, transformacja, load) – przebieg migracji testowej, lub produkcyjnej
- **Ekstrakt** – dostarczenie plików z danymi, lub całej bazy systemu źródłowego
- **Transformacja** – operacja polegająca na przetworzeniu danych źródłowych na dane możliwe do wczytania do systemu docelowego zgodnie z ustalonymi regułami
- **Load** – wczytanie przetworzonych danych do systemu docelowego

## 2 ZAKRES PRZEDMIOTOWY MIGROWANYCH DANYCH

Ogólny zakres migrowanych danych zostanie przedstawiony w dalszej części tego dokumentu, zaś bardziej szczegółowy – w produkcie **Plan Migracji**.

Zakres migrowanych danych wyznaczają następujące podziały:

- pochodzenie – system źródłowy danych,
- obiekty - zakres tematyczny,
- obiekty - zakres historyczny.

Zagadnienia te omówione są w kolejnych rozdziałach.



### 3 BEZPIECZEŃSTWO DANYCH

**Wszystkie dane pochodzące z systemów źródłowych i ekstraktów danych poddawane migracji do systemów docelowych muszą być zabezpieczone w sposób należyty.**

Konieczne jest zapewnienie poufności, integralności, dostępności i odporności systemów i usług przetwarzania danych.

Proces ekstrakcji, transformacji i ładowania danych **musi być zgodny z dokumentami nadrzędnymi do niniejszego**, w szczególności z obowiązującym prawem oraz obowiązującymi w InnoBaltica, wewnętrznymi politykami i procedurami, w tym w szczególności:

- Rozporządzeniem Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE
- Ustawą z dnia 10 maja 2018 roku w sprawie ochrony danych osobowych. Dz. U. nr 1000 z 2018 roku
- Polityką Bezpieczeństwa Danych Osobowych w InnoBaltica Sp. z o. o.
- Instrukcją Zarządzania Systemami Informatycznymi służącymi do przetwarzania danych osobowych w InnoBaltica Sp. z o. o.

**Wskazane powyżej dokumenty zostaną przekazane do wglądu wykonawcy w momencie podpisania umowy.**

Dane powinny być przesłane wyłącznie za pomocą protokołów zapewniających bezpieczeństwo danych.

Osoby uczestniczące w procesie ekstrakcji, transformacji i ładowania danych muszą być do tego upoważnione przez Wykonawcę, a Zamawiający musi być na bieżąco informowany pisemnie, dopuszczalna jest droga elektroniczna, o nadawaniu, odbieraniu i modyfikowaniu uprawnień poszczególnych osób.

Wykonawca zobowiązany jest informować zamawiającego o wszystkich problemach, zagrożeniach, incydentach i zdarzeniach zgodnie z SIWZ.



## 4 SYSTEMY ŹRÓDŁOWE I DOCELOWE

Potencjalnymi źródłami danych są (Systemy rozpoznane na podstawie dostępnych na dzień powstania niniejszego dokumentu danych):

- Analizator danych PIXEL dla MZK Starogard
- Bilet MSQL
- BUSMAN 100
- BUSMAN 120
- ELGEBRA
- Ereporter
- Ergeria
- eSKM – Bilet
- eSKM – ER SKM
- Infospread Polska Sp.z o. o.
- ISOKE
- KPD – Krajowy Punkt Dostępowy
- KRD - Krajowy rejestr dłużników
- Kurs90
- MITRON ZSDIP
- Mumicom Premium R&G
- Optima
- Pakiet Pixel 3
- Pliki płaskie
- PPBilety
- System roweru publicznego
- Rozkład 7xp
- SYSTEmEG
- Standardowy mechanizm integracji w zakresie identyfikacji i weryfikacji uprawnień w systemach zewnętrznych wraz z kompletną dokumentacją
- TRISTAR
  - SAE w ramach Systemu Zarządzania Transportem Zbiorowym (element Systemu Tristar)
- WinAD
- Winado

Systemami docelowymi są:

- PZUM
- Szyna Danych
- Hurtownia Danych.

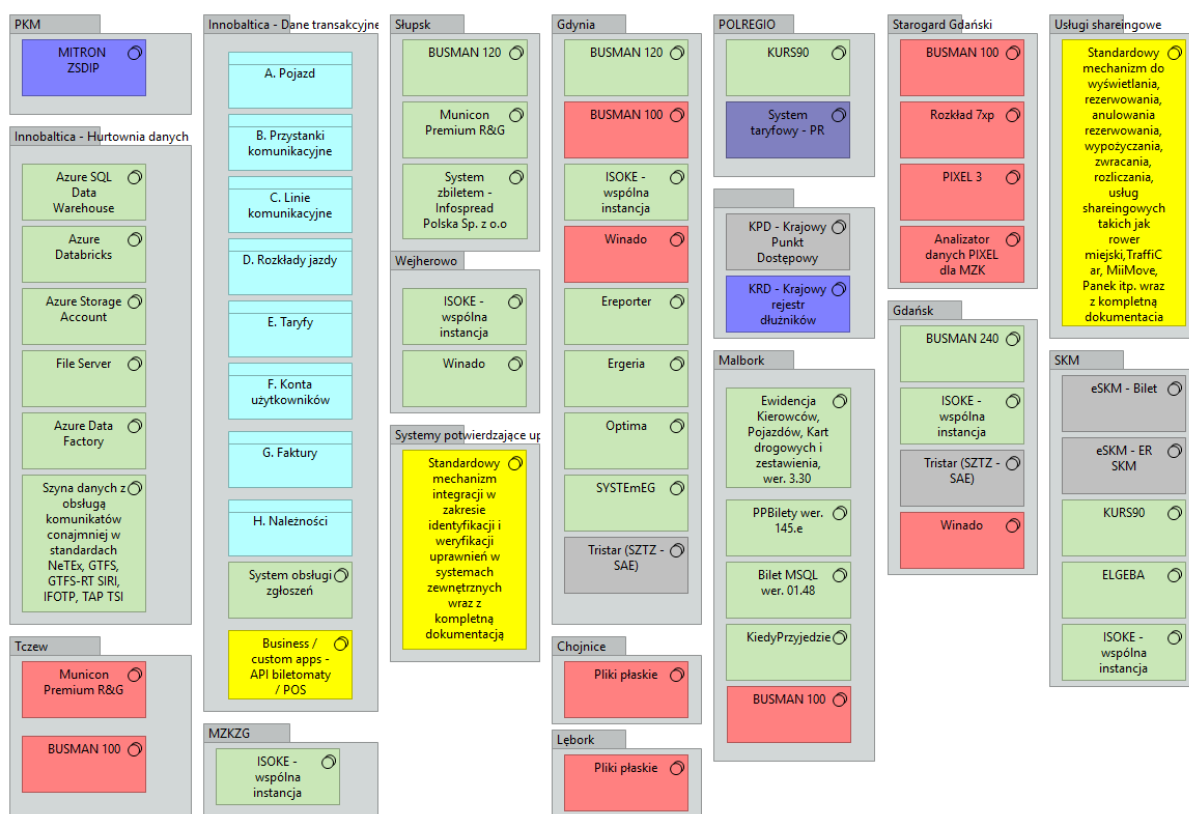
## 5 WARUNKI SELEKCJI MIGROWANEJ GRUPY KLIENTÓW ZAKRES MIGROWANYCH DANYCH

### 5.1 WYBÓR SYSTEMÓW I OBIEKTÓW DO MIGRACJI

Ostateczna lista encji biznesowych i ich atrybutów (razem z ewentualną historią), dla których wskazana jest Migracja Danych, oraz ostateczny szczegółowy opis zakresu migracji, zostanie zdefiniowana w postaci Specyfikacji Migracji Danych w ramach produktu **Plan Migracji**.

#### 5.1.1 Systemy źródłowe dla migracji

Dane mogą być pozyskane z systemów opisanych w punkcie 4 Systemy źródłowe i docelowe (Systemy rozpoznane na podstawie dostępnych na dzień powstania niniejszego dokumentu danych):



Czerwony - Migracja  
 Żółty - Standardowy interfejs  
 Fioletowy - Integracja (Pobieranie danych z możliwością przekazania parametrów)  
 Szary - Integracja  
 Zielony - Potencjalne źródło danych  
 Błękitny - Potencjalne artefakty systemu PZUM

INNOBALTICA SP. Z O.O. 80-067 GDAŃSK, UL. RÓWNA 19/21

NIP 957-10-03-404, REGON 220639884, BANK PKO BP SA 63 1020 1811 0000 0102 0311 6969

KRS 0000311943, VII WYDZIAŁ GOSPODARCTWA KRAJOWEGO REJESTRU SĄDOWEGO KAPITAŁ ZAKŁADOWY 21 305 000,00 PLN



Rzeczpospolita  
Polska

Unia Europejska  
Fundusz Spójności



Z.01

10 / 37



### 5.1.2 Dane podlegające migracji

Zmigrowane zostaną conajmniej dane wg listy obiektów z punktu 5.1.4 Obiekty podlegające migracji z perspektywy organizatora i 5.1.5 Obiekty podlegające migracji z perspektywy pasażera.

Wybrane (wskazane w Planie Migracji, po uzgodnieniu z zamawiającym) z migrowanych obiekty, będą musiały być oznaczone tzw. Timestamp'em określających czas ich ostatniego zmigrowania / powstania / edycji dowolnego atrybutu dla danego obiektu.

Wybrane (wskazane w Planie Migracji, po uzgodnieniu z zamawiającym) obiekty zostaną zmigrowane w sposób pozwalający na jednoznaczne określenie ich pochodzenia. Przykładowa nomenklatura *GDN.BUSMAN.TPRZYSTANKI.1333* (system źródłowy.klasa/tabela.ID z systemu źródłowego).

Standaryzacja oznaczeń obiektów pozwalająca na odróżnienie obiektów w zależności od interesariusza (celem jest identyfikowalność obiektów przy migracji z wielu systemów do jednego).

### 5.1.3 Dane wyłączone z migracji

W ramach danych podlegających migracji wyłączone zostaną wybrane dane historyczne według zapisu w kolejnych podpunktach punktu 14.1 Wybór scenariusza migracji „Z bilansem otwarcia”.

### 5.1.4 Obiekty podlegające migracji z perspektywy organizatora

- organizator transportu,
- operator transportu,
- przystanki,
- pojazdy,
- linie komunikacyjne,
- rozkłady jazdy,
- taryfy,
- przeszłe przejazdy,
- inne obiekty niezbędne do organizacji, zarządzania, realizacji etc. usług transportowych.

### 5.1.5 Obiekty podlegające migracji z perspektywy pasażera

- użytkownik,
- karta (oraz inne nośniki identyfikacji),
- podróż,
- płatności,

- posiadane środki,
- dostępne usługi,
- przystanki – dane podlegające migracji (PKP),
- przystanki – dane podlegające migracji (inne niż PKP),
- inne obiekty niezbędne do zaplanowania, korzystania, rozliczania etc. usług transportowych.

## **5.2 ZAKRES ŹRÓDŁOWY I DOCELOWY MIGROWANYCH DANYCH**

### **5.2.1 Zakres danych migrowanych z systemów źródłowych do systemu docelowego**

#### **5.2.1.1 Dostępne usługi - dane podlegające migracji (taryfy)**

W ramach taryf, zostaną zmigrowane aktualne taryfy.

#### **5.2.1.2 Przystanki - dane podlegające migracji (PKP)**

Przystanki do migracji z zakresie kolejowym zostały wskazane w Opisie przedmiotu zamówienia.

#### **5.2.1.3 Przystanki - dane podlegające migracji (inne niż PKP)**

Przystanki drogowe zostaną pozyskane z systemów źródłowych, zostaną uzupełnione o pliki płaskie, a ich poprawność zostanie zweryfikowana z uchwałami organizatorów transportu i rejestrem przystanków InnoBaltica.

#### **5.2.1.4 Trasy - dane podlegające migracji**

Informacje o trasach zostaną zmigrowane.

#### **5.2.1.5 Rozkłady jazdy - dane podlegające migracji**

Rozkłady jazdy zostaną zmigrowane.

## 6 ŚRODOWISKA MIGRACJI DANYCH (LOGICZNE)

Poniżej przedstawiono opis ogólny tego środowiska i jego otoczenia.

Środowisko Migracji Danych jest to komponent sprzętowo-programowy służący do przygotowania i przeprowadzenia kolejnych iteracji procesu Migracji Danych.

Środowisko migracji obrazuje Diagram nr 1 – Architektura fizyczna Środowiska Migracji Danych i jego otoczenia – struktury bazodanowe.

Iteracji (ETL) w ramach Migracji będzie wiele (kilkadziesiąt); większość z nich będzie iteracjami testowymi, natomiast ostatnie dwie – to migracja dla testów odbiorczych (UAT) oraz migracja produkcyjna.

Nie wszystkie testowe ETL będą pełne. Na jednokrotnie dostarczonych danych źródłowych może zostać wykonanych pula transformacji i procedur loadu, oraz raportów kontrolnych.

Środowisko Migracji Danych posiada wewnętrzną architekturę, która pozwala na niezależne, iteracyjne tworzenie Skryptów Migracji Danych dla różnych zakresów (systemów źródłowych) migracji.

Środowisko to potrzebne będzie do czasu przeprowadzenia ostatniej produkcyjnej Migracji Danych systemu.

Na poniższym diagramie przedstawiono koncepcję architektury Środowiska Migracji oraz jego Otoczenia.

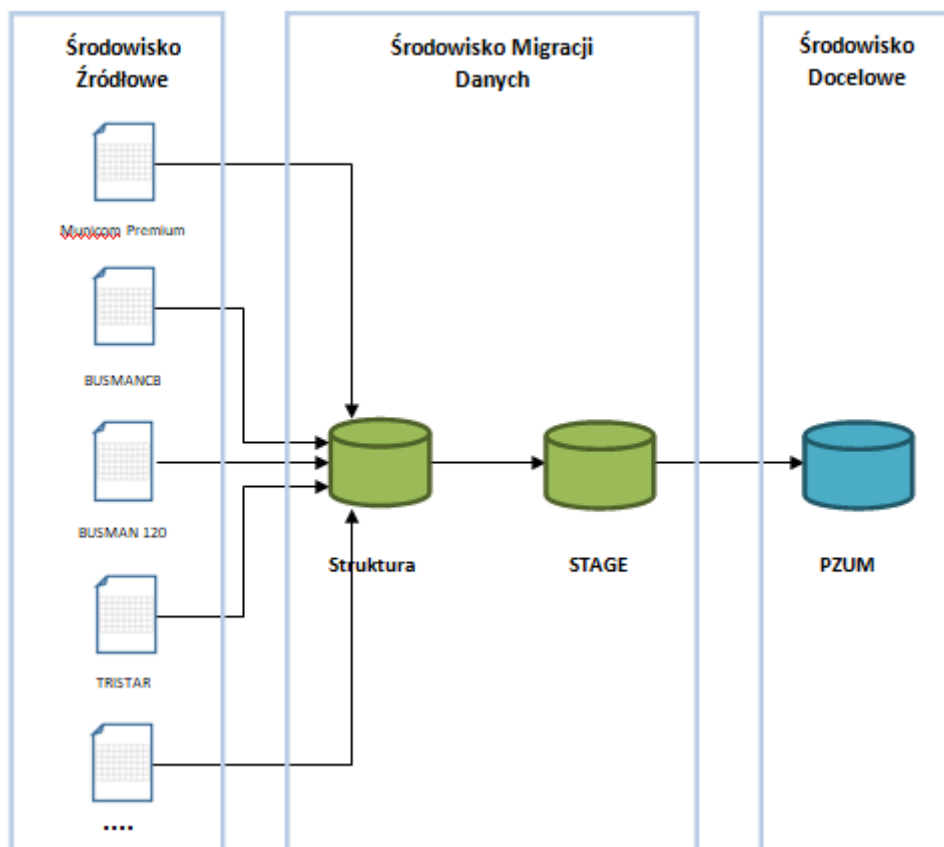


Diagram nr 2 – Architektura fizyczna Środowiska Migracji Danych i jego otoczenia – struktury bazodanowe

Na otoczenie składają się:

- Środowisko Udostępniające Dane Źródłowe (w skrócie – Środowisko Źródłowe).

Dane źródłowe dla danego systemu źródłowego zostaną przekazane w jednej z ustalonych i przedstawionych poniżej postaci:

- Plików płaskich,
- Pliki JSON,
- Raportów wygenerowanych z systemu źródłowego,
- Kopii bazy danych systemu źródłowego
- Użytkownika z uprawnieniami odczytu
- Inne.
- Środowisko Migracji Danych - Środowisko przejściowe zawierające pośrednie struktury danych



- Środowisko Docelowe - Środowisko schematów baz danych bezpośrednio wykorzystywane w operacyjnej pracy systemu.

## 6.1 STRUKTURA DANYCH W OBIEKTACH

Struktura danych w ekstraktach szczegółowo zostanie opisana w dokumencie Plan Migracji. Poglądowo została opisana w Z.02 - Wybrane informacje na potrzeby migracji danych

## 6.2 ŚRODOWISKO DOCELOWE

Środowiska Docelowe (Produkcyjne, Testowe lub UAT) rozumiane jako chmura, na której będzie funkcjonować system produkcyjny lub jego odpowiednik dla potrzeb testów. Elementami tych środowisk są m.in. schematy bazodanowe posadowione w chmurze, które przechowują dane używane przez system.

Środowisko Docelowe może być określane konfiguracyjnie - poprzez skierowanie danych wyjściowych migracji na wybrane środowiska systemu (Środowiska Testowe, Środowiska UAT, Środowiska Produkcyjne).

## 6.3 ŚRODOWISKO DEVELOPERSKIE

Środowisko developerskie może zostać wykorzystane przy migracjach testowych.

## 6.4 ŚRODOWISKA TESTOWE

Środowisko lub środowiska testowe, w zależności od cyklu wytwarzania wykonawcy, będą wykorzystywane przy Migracjach Testowych.

## 6.5 ŚRODOWISKO UAT

Środowisko UAT zostaną użyte przy odbiorze produktu przez Zamawiającego.

Aby uniknąć konfiguracji dodatkowego środowiska, podczas UAT może zostać wykorzystane jedno ze środowisk testowych.

## 6.6 ŚRODOWISKO PRODUKCYJNE

Środowisko Produkcyjne zostanie użyte przy Migracji Produkcyjnej.

Tym samym, zanim zostanie uruchomiony proces Migracji UAT czy Migracji Produkcyjnej, dane migrowane mogą być skierowywane na dowolne Środowiska Testowe i tam po każdej testowej iteracji migracji podlegać testom funkcjonalnym.



Samo Środowisko Migracji Danych są to zasoby w chmurze, których głównym i najważniejszym komponentem są struktury w postaci schematów bazodanowych, wykorzystywanymi w trakcie Migracji Danych. W celu migracji do systemu docelowego będą tu wykorzystywane następujące struktury (schematy bazodanowe):

- **Baza Przejściowa.**

Schemat z tablicami w bazie danych, zawierający zaczytane dane, dostarczone z systemów źródłowych.

Dane źródłowe będą podlegały weryfikacji wstępnej podczas zaczytania (pod kontem struktury), oraz weryfikacji zgodności.

- **Stage dla migrowanych systemów.**

Struktura pośrednia wykorzystywana w trakcie Migracji Danych do systemu docelowego. Do schematu Stage transferowane będą po przekształceniu dane pochodzące z Bazy Przejściowej zawierającej dane ekstraktu. Dane te docelowo mają trafić do schematu produkcyjnego systemu PZUM.

- **Instancja systemu docelowego**

Jest to dedykowana do celów migracyjnych instancja systemu PZUM. Pozwala na bieżące sprawdzanie efektów działania skryptów migracyjnych. Instancja ta będzie odpowiadała aktualnie rozwijanej wersji Systemu PZUM.

- **Framework**

Komponent własny Wykonawcy pozwalający na uruchamianie Skryptów Migracji Danych. Komponent pozwala m.in. na:

- kontrolę i monitorowanie poszczególnych podprocesów,
- rejestrację czasów wykonania poszczególnych podprocesów,
- zrównoleglenie przetwarzania (o ile jest to logicznie możliwe),
- rejestrację błędów i wznawianie wykonania.

- **Skrypty Migracji Danych**

Oprogramowanie wymagane do przeprowadzenia procesu Migracji i uruchamiane za pomocą komponentu Framework. Najczęściej skrypty migracji realizowane są jako procedury bazodanowe. Podstawowy podział skryptów Migracji jest następujący:

- Skrypty Ekstrakcji – do pobrania danych z systemów wejściowych (wykonywane po stronie systemów źródłowych),
- Skrypty Ładowania ekstraktów do bazy przejściowej
- Skrypty Transformacji – do przekształcenia danych z bazy przejściowej do postaci wymaganej przez Stage/system docelowy, (czyli do postaci istniejącej w strukturze systemu docelowego),
- Skrypty Ładowania – do załadowania danych ze struktury Stage do Systemu PZUM.

- **Narzędzia**

Narzędzia wykorzystywane przez Skrypty Ekstrakcji do podłączenia się do baz danych Środowiska Udostępniającego Dane Źródłowe lub do Bazy Przejściowej i „wyciągnięcia” danych do celów migracji.

INNOBALTICA SP. Z O.O. 80-067 GDAŃSK, UL. RÓWNA 19/21  
NIP 957-10-03-404, REGON 220639884, BANK PKO BP SA 63 1020 1811 0000 0102 0311 6969  
KRS 0000311943, VII WYDZIAŁ GOSPODARCZY KRAJOWEGO REJESTRU SĄDOWEGO KAPITAŁ ZAKŁADOWY 21 305 000,00 PLN



Rzeczpospolita  
Polska

Unia Europejska  
Fundusz Spójności



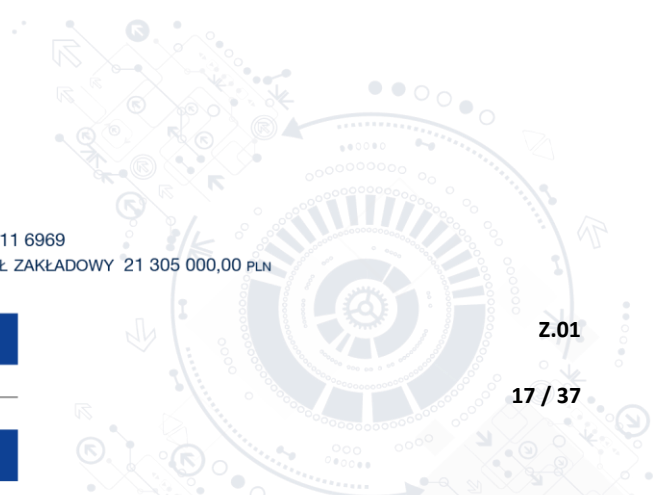
Wylączna odpowiedzialność za treść niniejszej publikacji leży po stronie  
jej autorów. Jej treść nie musi odzwierciedlać opinii Unii Europejskiej.  
Ani Europejski Bank Inwestycyjny ani Komisja Europejska  
nie są odpowiedzialne za jakiegokolwiek wykorzystanie zawartych tu informacji.

Współfinansowane przez program ramowy  
Unii Europejskiej „Horyzont 2020”



Z.01

17 / 37



## 7 KONCEPCJA TESTÓW MIGRACJI (PRÓBKOWANIE)

W każdej iteracji migracji testowych, po wykonaniu raportów walidacji i raportów kontrolnych i stwierdzeniu ich poprawności, nastąpią testy poprawności wykonanej migracji.

Testy te polegać będą na sprawdzeniu stanu zmigrowania próbki kilku – kilkunastu obiektów każdej klasy, z każdego systemu źródłowego, poprzez sprawdzenie działania systemu docelowego z danymi tych obiektów, zgodnie z zatwierdzonym scenariuszem. Lista scenariuszy zaproponuje zamawiający.

Minimalna lista testów:

- Na poziomie klasy
  - Zgodność definicji obiektów / klas z założeniami / wymaganiami
- Na poziomie obiektu
  - Możliwość identyfikacji systemu źródłowego dla każdego z obiektów
  - Wybór losowej części populacji rekordów źródłowych, pozwalających na wiarygodną walidację w systemie docelowym
- Na poziomie atrybutu
  - Zgodność formatów zapisów atrybutu w ramach poszczególnych klas
  - Kompletność obowiązkowych atrybutów
  - Zgodność poszczególnych atrybutów w systemie źródłowym, systemie docelowym

Dokładny przebieg i zakres testów zostanie zdefiniowany w Planie migracji przy akceptacji zamawiającego.

## **8 DOSTĘPY DO DANYCH ŹRÓDŁOWYCH (M.IN. ZIDENTYFIKOWANE PROBLEMY PRAWNE Z DOSTĘPEM DO DANYCH) OKREŚLONE W ZAŁOŻENIACH DO KONCEPCJI MIGRACJI**

Do wykonania migracji próbnych należy uzyskać zgody do przetwarzania. Lista uczestników projektu wraz z informacjami o rodzaju, roli, uczestnictwie w projekcie została wskazana w OPZ.

INNOBALTICA SP. Z O.O. 80-067 GDAŃSK, UL. RÓWNA 19/21  
NIP 957-10-03-404, REGON 220639884, BANK PKO BP SA 63 1020 1811 0000 0102 0311 6969  
KRS 0000311943, VII WYDZIAŁ GOSPODARCTWA KRAJOWEGO REJESTRU SĄDOWEGO KAPITAŁ ZAKŁADOWY 21 305 000,00 PLN



Rzeczpospolita  
Polska

Unia Europejska  
Fundusz Spójności



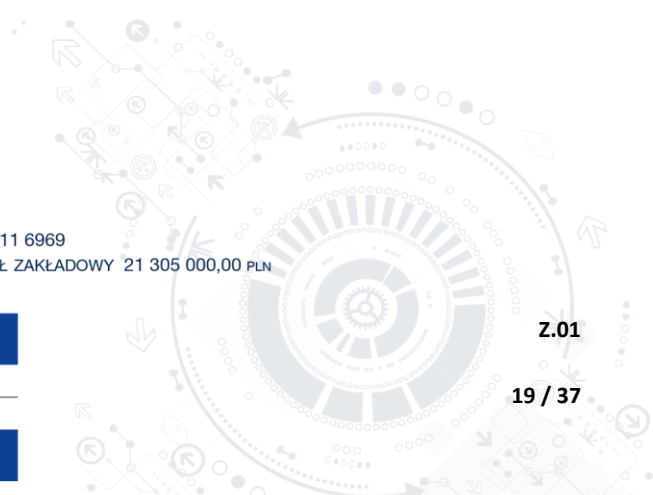
Wylączna odpowiedzialność za treść niniejszej publikacji leży po stronie jej autorów. Jej treść nie musi odzwierciedlać opinii Unii Europejskiej. Ani Europejski Bank Inwestycyjny ani Komisja Europejska nie są odpowiedzialne za jakiegokolwiek wykorzystanie zawartych tu informacji.

Współfinansowane przez program ramowy Unii Europejskiej „Horyzont 2020”



Z.01

19 / 37



## 9 ZASADY PODZIAŁU ZADAŃ W TRAKCIE MIGRACJI DANYCH

Wykonawca i Zamawiający są odpowiedzialni za stworzenie odrębnych produktów – opisano to w rozdziale 12 Produkty strumienia Migracji danych (z podziałem i powiązaniem).

**Po stronie Wykonawcy** leży przeprowadzenie całego procesu Migracji Danych, począwszy od pobrania danych z systemów Źródłowych lub zaczytania dostarczonych ekstraktów, przez ich transformację i rekomendowanie czyszczenia w zakresie niezbędnym do migracji (tam gdzie jest to możliwe i wskazane), kończąc na załadowaniu do systemu Docelowego.

Rekomendowanie czyszczenia w zakresie niezbędnym do migracji to wykrywanie braków i niespójności uniemożliwiających migrację danego obiektu i sygnalizowanie ich Zamawiającemu. Za czyszczenie danych w bazie Źródłowej odpowiedzialny jest Zamawiający. Zamawiający może wskazać konieczność zakodowania przez Wykonawcę automatycznego uzupełniania złych lub brakujących danych.

**Po stronie Zamawiającego** są następujące zadania:

- umożliwienie Wykonawcy pobrania danych pochodzących z systemów źródłowych,
- odpowiedzialność za czyszczenie biznesowe danych, w szczególności udzielanie odpowiedzi umożliwiających zaprojektowanie transformacji służących poprawie danych wejściowych,
- wykonanie testów akceptacyjnych,
- akceptacja wyników raportów kontrolnych oraz podejmowanie decyzji o uruchomieniu produkcyjnym.

## 10 ORGANIZACJA ZESPOŁU MIGRACJI DANYCH

### 10.1 ZESPOŁY ZAMAWIAJĄCEGO I WYKONAWCY

W celu realizacji projektu migracji w ramach wdrożenia systemu proponuje się utworzenie zespołów zgodnych z metodyką prowadzenia prac wykonawcy, zatwierdzoną przez zamawiającego.

Zespoły zostaną sformowane w trakcie prac projektowych. Szczegóły pojawiają się w kolejnych dokumentach (Plan Migracji)

INNOBALTICA SP. Z O.O. 80-067 GDAŃSK, UL. RÓWNA 19/21  
NIP 957-10-03-404, REGON 220639884, BANK PKO BP SA 63 1020 1811 0000 0102 0311 6969  
KRS 0000311943, VII WYDZIAŁ GOSPODARCZY KRAJOWEGO REJESTRU SĄDOWEGO KAPITAŁ ZAKŁADOWY 21 305 000,00 PLN



Rzeczpospolita  
Polska

Unia Europejska  
Fundusz Spójności



## 11 OPIS ZAKRESU MIGROWANYCH DANYCH

Wolumeny danych dla wydania

Poniższa tabela stanowi ramy do późniejszego określenia wolumenów danych migrowanych do DOCELOWEGO w podziale na Operatora Transportu.

Wolumeny danych:

- Ilość przystanków w Województwie Pomorskim (wg. Załącznik do Uchwały Nr 210/XVIII/16 Sejmiku Województwa Pomorskiego z dnia 29 lutego 2016 roku) – 1926
- Ilość taryf – są to zestawy danych rzędu kilkudziesięciu taryf (z uwzględnieniem ulg i wariantów)

Tabela jest wytyczną do zebrania danych (na dzień powstania niniejszego dokumentu wolumen danych nie jest znany – będzie to ustalone w Planie Migracji):

Lp.	Operator Transportu
1	BP Tour
2	GAIT Sp. z o.o.
3	IREX 3 Sp. z o.o.
4	Miejski Zakład Komunikacji Sp. z o. o. w Słupsku
5	Miejski Zakład Komunikacji spółka z ograniczona odpowiedzialnością w Chojnicach
6	Miejski Zakład Komunikacji w Malborku
7	Miejski Zakład Komunikacji w Starogardzie Gdańskim
8	Miejski Zakład Komunikacji Wejherowo Sp. z o.o.
9	Obszar Metropolitalny Gdańsk – Gdynia - Sopot
10	PKS Gdynia S.A.
11	Przedsiębiorstwo Komunikacji Autobusowej w Gdyni Sp. z o.o.
12	Przedsiębiorstwo Komunikacji Miejskiej w Gdyni Sp. z o.o.
13	Przedsiębiorstwo Komunikacji Samochodowej Gdańsk sp. z o.o.
14	Przedsiębiorstwo Komunikacji Trolejbusowej w Gdyni Sp. z o.o.
15	Przewozy Autobusowe GRYF Sp. z o.o. Sp. k.
16	Warbus Sp. z o.o. Warszawa
17	Zakład Komunikacji Miejskiej w Lęborku Spółka z o.o.
18	PKP Szybka Kolej Miejska w Trójmieście Sp. z o.o.
19	Pomorskie Towarzystwo Miłośników Kolei Żelaznych
20	Przewozy Regionalne Sp. z o.o. (marka POLREGIO)
21	Metropolitalny Związek Komunikacyjny Zatoki Gdańskiej

Tabela nr 1 – Lista do docelowego oszacowania wolumenu migrowanych danych

INNOBALTICA SP. Z O.O. 80-067 GDAŃSK, UL. RÓWNA 19/21

NIP 957-10-03-404, REGON 220639884, BANK PKO BP SA 63 1020 1811 0000 0102 0311 6969

KRS 0000311943, VII WYDZIAŁ GOSPODARCZY KRAJOWEGO REJESTRU SĄDOWEGO KAPITAŁ ZAKŁADOWY 21 305 000,00 PLN



Rzeczpospolita  
Polska

Unia Europejska  
Fundusz Spójności



Z.01

22 / 37





## 12 PRODUKTY STRUMIENIA MIGRACJI DANYCH (Z PODZIAŁEM I POWIĄZANIAM)

W procesie migracji Zespoły Wykonawcy i Zamawiającego współpracują ze sobą.

Wykonawca i Zamawiający mają do wykonania własne, określone produkty.

Produkty powinny zostać stworzone zgodnie z uzgodnioną metodyką realizacji projektu, powinien on jednak pokrywać merytorycznie, co najmniej, poniższe obszary.

### 12.1 PRODUKTY MIGRACJI DANYCH DOSTARCZANE PRZEZ WYKONAWCĘ

Kluczowymi produktami będą, (choć możliwa jest jeszcze zmiana tej listy):

- W.01.v1 – Koncepcja realizacji Migracji Danych
- W.02.v1 - Analiza, jakości danych i rekomendacja dotycząca ich czyszczenia
- W.03.v1 - Techniczna podejście do migracji
- W.04.v1 - Scenariusze testów migracji
- W.05.v1 - Plan Migracji
- W.06.v1 - Migracja produkcyjna wraz z raportami rekonyliacji danych
- W.07.v1 - Rekomendacja dotycząca migracji produkcyjnej
- W.08.v1 - Metoda lub narzędzie weryfikacji migracji danych zdefiniowane

### 12.2 PRODUKTY MIGRACJI DANYCH DOSTARCZANE PRZEZ ZAMAWIAJĄCEGO

Będą to poniższe produkty, (choć możliwa jest jeszcze zmiana tej listy):

- Z.01.v1 – Założenia do Koncepcji Migracji (niniejszy dokument)
- Z.02.v1 - Wybrane informacje na potrzeby migracji danych
- Z.03.v1 - Lista raportów na potrzeby weryfikacji migracji
- Z.04.v1 - Wskaźniki efektu migracji w systemach migrowanych
- Z.05.v1 - Wyczyszczone dane źródłowe do migracji w zakresie decyzji biznesowych
- Z.06.v1 - Potwierdzona poprawność migracji produkcyjnej
- Z.07.v1 – Wskazanie sposobów czyszczenia danych w ramach kart jakości danych

## 13 SCHEMAT WYKONANIA MIGRACJI DANYCH

Iteracyjny przebieg przygotowania skryptów migracyjnych w fazie Realizacja.

Szczegółowy opis sposobu wykonania Migracji Danych znajdzie się w produkcie W3.v1 – Techniczne podejście do migracji.

Poniżej schemat wykonania Migracji Danych opisano w sposób ogólny.

W ramach procesu odbędą się kolejno trzy rodzaje Migracji Danych:

- Migracja Testowa,
- Migracja UAT (User Acceptance Tests),
- Migracja Produkcyjna.

Uwaga: w ramach Migracji Testowej mogą też wystąpić iteracje Migracji Dodatkowych np. na dodatkowe środowisko migracyjne na potrzeby weryfikacji wskaźników efektów migracji. Sugerowane przez zamawiającego migracje takie planowane są do realizacji w cyklach miesięcznych lub innych zgodnych z metodyką Wykonawcy do momentu migracji produkcyjnej.

Takie migracje będą realizowane w miarę możliwości, a ich szczegóły opisane zostaną w produktach: W05.v1 Plan Migracji(w tym specyfikacja migracji danych).

### 13.1 MIGRACJA TESTOWA

Docelowym schematem bazodanowym Migracji Testowej jest dedykowane Środowisko Testowe lub Developerskie.

Migracja Testowa składa się z wielu Iteracji Migracji. Budowa i optymalizacja skryptów migracyjnych wymaga wielokrotnych przebiegów na środowisku developerskim. Ilość migracji na potrzeby testów będzie zależna min. od jakości danych.

Każda z kolejnych Iteracji Migracji Testowej ma 2 fazy:

- **Przygotowanie**
- **Przeprowadzenie**

#### 13.1.1 Przygotowanie

Iteracji Migracji Danych polega na:

- wzięciu pod uwagę wyników poprzedniej iteracji (m.in. Raportów Kontrolnych) **przez Wykonawcę i przez Zamawiającego**
- zmodyfikowaniu produktu **W05.v1 Plan Migracji(w tym specyfikacja migracji danych) przez Wykonawcę**

- czyszczeniu danych w Systemie Źródłowym **odpowiedzialność Zamawiającego**

Wszystkie elementy muszą zostać opracowane i wykonane tak, aby w nowej Iteracji wynik przeprowadzenia migracji był lepszy niż w poprzedniej iteracji. Same dane oraz skrypty migracyjne będą podlegały ewolucyjnemu rozwojowi. Jeśli w n-tej Iteracji wynik będzie słabszy, niż w Iteracji poprzedniej elementy mające wpływ na migrację zostaną zweryfikowane i zmiany mające wpływ na pogorszenie wyniku zostaną wycofane, lub zostaną wykonane odpowiednie modyfikacje.

Na wynik migracji będą składały się elementy

- Ilość znalezionych błędów (niepoprawnych transformacji, zapisów odrzuconych itp.)
- Migrowalność ilość zapisanych w systemie docelowym danych w odniesieniu do ilości odczytanych z systemu źródłowego

Modyfikowane są następujące składniki produktu W05.v1 Plan Migracji (w tym specyfikacja migracji danych):

- Specyfikacja migracji danych zawierająca m.in. opis reguł walidacji i transformacji,
- Skrypty migracji danych - czyli kod, którego egzekucja realizuje walidacje i transformacje danych,
- Słowniki - wspomagające transformację danych.

### 13.1.2 Przeprowadzenie

Iteracji Migracji Danych obejmuje zadania organizacyjne i techniczne wykonywane z użyciem przygotowanych Specyfikacji, Skryptów i Słowników. Następuje egzekucja Skryptów oraz generowane są Raporty Kontrolne. Sprawdzane jest również działanie aplikacji docelowej według scenariuszy testowych. Sprawdzane jest również inne aspekty działania skryptów takie jak wydajność.

## 13.2 MIGRACJA UAT

Po zakończeniu wielu iteracji Migracji Testowej nastąpi Migracja UAT (User Acceptance Tests).

Bazą docelową jest tu - podobnie jak przy Migracji Testowej - Baza Testowa.

Migracja UAT pozwoli na ostateczne sprawdzenie przez Zamawiającego poprawności przygotowanych produktów Migracji i ich zaakceptowanie do realizacji produkcyjnej.

### 13.3 MIGRACJA PRODUKCYJNA

Końcowy etap to Migracja Produkcyjna, dla której bazą docelową będzie Baza Produkcyjna (schemat bazodanowy w chmurze).

Migracja Produkcyjna będzie realizowana zgodnie z tym samym scenariuszem, co ostateczna Migracja Testowa i Migracja UAT. Różnica (oprócz systemu docelowego) polega na tym, że w Migracji Produkcyjnej nie dokonuje się już zmian Specyfikacji, Skryptów i Słowników.

INNOBALTICA SP. Z O.O. 80-067 GDAŃSK, UL. RÓWNA 19/21  
NIP 957-10-03-404, REGON 220639884, BANK PKO BP SA 63 1020 1811 0000 0102 0311 6969  
KRS 0000311943, VII WYDZIAŁ GOSPODARCTWA KRAJOWEGO REJESTRU SĄDOWEGO KAPITAŁ ZAKŁADOWY 21 305 000,00 PLN



Rzeczpospolita  
Polska

Unia Europejska  
Fundusz Spójności



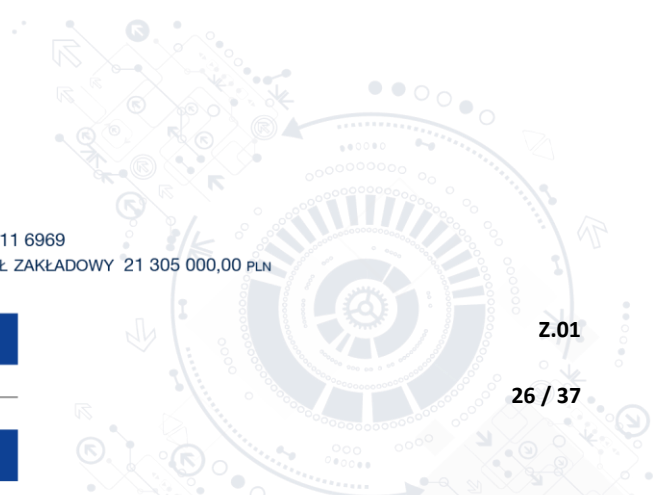
Wyłączna odpowiedzialność za treść niniejszej publikacji leży po stronie jej autorów. Jej treść nie musi odzwierciedlać opinii Unii Europejskiej. Ani Europejski Bank Inwestycyjny ani Komisja Europejska nie są odpowiedzialne za jakiegokolwiek wykorzystanie zawartych tu informacji.

Współfinansowane przez program ramowy  
Unii Europejskiej „Horyzont 2020”



Z.01

26 / 37



## 14 ALGORYTM ODCIĘCIA DANYCH (DANE HISTORYCZNE)

### 14.1 WYBÓR SCENARIUSZA MIGRACJI „Z BILANSEM OTWARCIA”

Wybrano do realizacji scenariusz migracji „Z bilansem otwarcia”.

Co do zasady, polega on na migracji historii jedynie dla otwartych pozycji, bez migracji historii poza pozycjami otwartymi; przy tym pozycje otwarte najlepiej byłoby zamknąć w celu uniknięcia migrowania jakiejkolwiek historii.

Jednak dla szczególnych przypadków może się okazać niezbędne zmigrowanie części danych historycznych.

- Linie - Wymagana jest migracja historii na dwa lata wstecz od daty migracji.
- Przystanki - Wymagana jest migracja historii na co najmniej dwa lata wstecz od daty migracji.
- Taryfy - Historia Taryf nie będzie migrowana.
- Konta Klientów - Migrowana będzie cała historia kont klientów.
- Usługi transportowe (dotyczy Tczewa) - Wymagana jest migracja historii na 1,5 lat wstecz od daty migracji.
- Tabor - Wymagana jest migracja historii na dwa lata wstecz od daty migracji.
- Kierowcy - Wymagana jest migracja historii na dwa lata wstecz od daty migracji.



## 15 REGUŁY TRANSFORMACJI I WALIDACJI DANYCH

### 15.1 GŁÓWNE ZASADY TRANSFORMACJI

Szczegółowy **TECHNICZNY OPIS MIGRACJI**, zawierający opis technologii i infrastruktury wykonania migracji, **ZASADY TRANSFORMACJI DANYCH** zostaną opisane zgodnie z uzgodnioną metodyką.

W szczególności, zostaną tam opisane w sposób szczegółowy:

- reguły transformacji i walidacji danych,
- reguły uzupełniania brakujących danych, niewymaganych w systemach źródłowych, a wymaganych w systemie docelowym,
- odstępstwa od standardowego procesu Migracji Danych,
- reguły tworzenia danych referencyjnych podczas migracji,
- reguły obsługi błędnych danych, znajdujących się w starych systemach i migrowanych do systemu docelowego,
- postępowanie w sytuacjach wyjątkowych.

Poniżej założenia do koncepcji zostały opisane w sposób ogólny.

Dane przewidziane do migracji znajdują się w plikach płaskich, dokumentach, arkuszach kalkulacyjnych, bazach danych różnych Systemów Źródłowych. Każdy system posiada swoją indywidualną reprezentację obiektów biznesowych i ich powiązań w strukturach danych. Ponadto sposób wypełnienia oraz jakość biznesowa danych zależna jest od konkretnej instancji systemu lub nawet od grupy użytkowników tego systemu.

System posiada swoją reprezentację obiektów biznesowych. Model danych systemu Docelowego uwzględnia wymagania postawione przed systemem w zapytaniu ofertowym i potwierdzone, bądź uszczegółowiona w trakcie wykonywania Analizy Biznesowej.

Podstawowym celem migracji jest umożliwienie prowadzenia kontynuacji rozliczeń Klientów Zamawiającego. Kluczowym i najbardziej skomplikowanym zadaniem procesu migracji jest transformacja modelu danych wejściowego na model danych systemu wyjściowego, czyli przeniesienie wymaganych obiektów z Systemu Źródłowego do obiektów systemu docelowego w taki sposób, aby procesy związane z obsługą planowania, realizacji podróży, obsługi nośników biletów mogły być kontynuowane, nie zostały naruszone i nie powstały straty finansowe.

Ze względu na różnice między systemami wejściowym i wyjściowym, migrowane dane zostaną poddane transformacji za pomocą dostosowanych do zakresu tych danych **Skryptów Migracji Danych** przygotowanych na podstawie zapisów w **Specyfikacji**

**Migracji Danych.** zapisy w **Specyfikacji Migracji Danych** będą uwzględniały podjęte decyzje dotyczące reguł transformacji danych.

Poniżej sformułowano w postaci punktów główne zasady transformacji, które Zespół Migracji Danych Wykonawcy będzie stosował w trakcie przygotowania produktów:

- reguły walidacji zakresu dostarczonych danych oraz reguły walidacji i transformacji danych,
- słowniki mapujące,
- unikanie tworzenia „sztucznych” obiektów,
- reguły uzupełniania brakujących danych,
- unikanie potencjalnie „niebezpiecznego” czyszczenia danych.

Poniżej jest krótkie omówienie głównych zasad transformacji.

### 15.1.1 Reguły wyodrębnienia zakresu danych oraz reguły walidacji i transformacji danych

W systemie docelowym główne obiekty biznesowe w modelu struktur danych są reprezentowane odmiennie niż w systemie źródłowym (różnice w nazewnictwie, granulacji, powiązaniach, etc.).

Niezbędne jest określenie sposobu wyodrębnienia zakresu danych do migracji oraz sposobu ich walidacji oraz transformacji z modelu źródłowego na docelowy.

Istnieją różne sposoby tego wyodrębniania/walidowania/transformowania, wpływające m.in. na:

- zakres migrowanych danych,
- dostępną funkcjonalność systemu docelowego,
- łatwość i intuicyjność eksploatacji systemu docelowego,
- konieczność podjęcia dodatkowych czynności biznesowych przy migracji i/lub eksploatacji systemu,
- komplikację (i ryzyko) wykonania procesu migracji.

W produkcie **W05.v1 Plan Migracji(w tym specyfikacja migracji danych)** Wykonawca przedstawi Zamawiającemu sposób transformowania głównych obiektów biznesowych. Podany zostanie zakres transformacji, który wynika z ustaleń i KWJD (karty wymagań jakości danych). W KWJD wskazywane jest, które naprawy danych będą realizowane w ramach Transformacji Danych. Lista taka z odniesieniem do KWJD powinna być wskazana w Planie Migracji.

Uzgodniony wybór zostanie potwierdzony i zrealizowany w **Skryptach Migracji Danych**.



### 15.1.2 Słowniki mapujące

W celu elastycznej możliwości mapowania identyfikatorów źródłowych z docelowymi, w środowisku migracyjnym utworzone zostaną w postaci tablic bazodanowych specjalne słowniki mapujące, zawierające „stare” i „nowe” wartości.

Słowniki mapujące będą wykorzystywane w Skryptach Migracji Danych. Mapowanie dotyczyć będzie przede wszystkim wartości słownikowych (taryfy, produkty, strefy czasowe, etc.).

### 15.1.3 Unikanie tworzenia „sztucznych” obiektów

Zamawiający nie zakłada konieczności tworzenia sztucznych obiektów tylko na potrzeby migracji. Każdy taki przypadek powinien zostać zaakceptowany przez Zamawiającego.

danych.

Lista „sztucznych” obiektów (jeśli nie uda się uniknąć konieczności ich utworzenia) wraz ze specyfikacją i opisem zostanie przedstawiona w dokumencie **W05.v1 Plan Migracji (w tym specyfikacja migracji danych)**.

### 15.1.4 Reguły uzupełniania brakujących danych

Proces migracji będzie dotyczył policzalnej, znacznej liczby obiektów.

Skrypty migracyjne będą zupełnie odmienne dla danych z różnych Systemów Źródłowych. Będą też odmienne dla poszczególnych typów obiektów wymagających odmiennego przetwarzania (trasy, przystanki, taryfy itp.).

Dla bardzo dużych grup Klientów, dla których jest możliwe automatyczne określenie brakujących danych, zostanie to zrealizowane poprzez odpowiedni kod w Skryptach Migracji danych.

### 15.1.5 Unikanie potencjalnie „niebezpiecznego” czyszczenia danych

Wiele algorytmów czyszczenia danych nie daje stuprocentowej pewności zamiany „złych” danych na „lepsze”. Istnieje potencjalne ryzyko „poprawy” na błędne dane. Nawet, jeśli prawdopodobieństwo takiego zdarzenia jest niewielkie, może spowodować poważne straty biznesowe. Należy, zatem wykluczyć takie potencjalne zagrożenie dla danych istotnych w ważnych procesach biznesowych, np. stosując prostsze metody czyszczenia danych.

Na etapie Transformacji mogą, na wniosek Zamawiającego, zostać zaprojektowane reguły poprawy danych.

## 15.2 GŁÓWNE ETAPY WALIDACJI DANYCH

Walidacja danych podczas przebiegu migracyjnego (ETL) będzie zaimplementowana na etapie:

### 15.2.1 Zaczytywania ekstraktów do bazy STAGE

Dostarczone ekstrakty mogą być obciążone błędami zwłaszcza formatów danych. Walidacja i ewentualna poprawa musi zostać wykonana, by utrzymać spójność formatów danych

### 15.2.2 Transformacji

Podczas transformacji na podstawie ewentualnych reguł konwersji (jeśli ich zastosowanie okaże się niezbędne) zostanie zaimplementowana kolejna część walidacji.

### 15.2.3 Loadu

Walidacja podczas loadu – może okazać się konieczna dla zapisania danych w schemacie bazodanowym. Kluczowe zwłaszcza podczas iteracji testowych migracji i testów funkcjonalności na danych migrowanych.

## 16 REGUŁY UZUPEŁNIANIA BRAKUJĄCYCH DANYCH, NIEWYMAGANYCH W SYSTEMACH ŹRÓDŁOWYCH, A WYMAGANYCH W SYSTEMIE PZU

### 16.1 ZAKRES DZIAŁAŃ ZWIĄZANYCH Z POPRAWĄ, JAKOŚCI DANYCH

Zgodnie z zapisami z Umowy, przy transformacji i czyszczeniu biznesowym danych powinny obowiązywać poniższe ustalenia.

Zakres działań związanych z poprawą, jakości danych (migrowanych z systemów dziedzinowych do Systemu DOCELOWEGO) obejmuje:

- weryfikację i ujednolicenie standardów przechowywania danych: kodów pocztowych, numerów NIP, dat, nazw jednostek administracyjnych, numerów PESEL, REGON, numerów telefonów, adresów email i innych,
- weryfikację wzajemnej zgodności kodów pocztowych, nazwy poczty i adresów,
- ujednolicenie danych słownikowych pod względem kompletności, poprawności i braku duplikatów, w obszarach: produkty, taryfy, przystanki, trasy, formy działalności, formy własnościowe, jednostki administracyjne (nazwy miast, gmin oraz ulic, województw, powiatów),
- identyfikacja zduplikowanych rekordów i opracowanie sposobu ich obsługi,
- Inne elementy, ustalone z wykonawcą i dwustronnie zaakceptowane rozwiązania.

**Szczegóły działań w zakresie Wykonawcy zostaną ustalone w trakcie dalszych analiz i opisane w Planie Migracji.**

### 16.2 CZYSZCZENIE BIZNESOWE DANYCH – POZA ZAKRESEM WYKONAWCY

Weryfikacja i poprawa biznesowej jakości danych w ramach Poprawy Jakości Danych w procesie Migracji Danych do Systemu DOCELOWEGO jest poza zakresem Wykonawcy.

Odpowiada za to Zamawiający, dostarczając produkt **Z05.v1 Wyczyszczone biznesowo dane do migracji**.

Czyszczenie danych realizowane jest za pomocą utworzonych Kart Wymagań na Jakość Danych, które opisane będą w produkcie **Z07.v1 Implementacja Kart Wymagań na Jakość Danych**.

## 16.3 POSTĘPOWANIE Z BRAKUJĄCYMI LUB BŁĘDNYMI DANYMI

W przypadku zidentyfikowania błędnych lub brakujących danych, zostaną przygotowywane walidatory (zapytania) zwracające, jako wynik rekordy, dla których stwierdzono braki lub niespójność zapisów w systemach źródłowych.

Na podstawie zgłoszonych walidatorów tworzone będą Karty Wymagań na Jakość Danych, w których określa się między innymi:

- Opis wymagania
- Opis niezgodności
- Ilość wystąpień niezgodności
- Priorytet z dostępnymi wartościami:
  - 0 – dla kart odrzuconych
  - 1 – dla kart, gdzie dane mogą zostać zmigrowane do systemu docelowego
  - 9999 – dla kart, w których brak naprawy niezgodności skutkuje niemożnością zmigrowania danych
- Zapytanie SQL identyfikujące niezgodność
- Preferowany sposób naprawy niezgodności
- Metodę naprawy niezgodności jako jedną z poniższych:
  - Ręcznie w aplikacji źródłowej,
  - W bazie danych aplikacji źródłowej,
  - Podczas migracji na podstawie dostarczonych danych,
  - W stage / bazie pośredniej,
  - w systemie docelowym po migracji

Niezgodności dla kart, dla których jako metodę niezgodności przyjęto „Ręcznie w aplikacji źródłowej” lub „W bazie danych aplikacji źródłowej” z priorytetem 9999 powinny zostać poprawione, aby możliwa była migracja danych do systemu docelowego.

Karty Wymagań na Jakość Danych przygotowywane są przez Zamawiającego na podstawie przesłanych walidatorów (zapytań SQL listujących rekordy o zbyt niskiej jakości).

## 17 REGUŁY TWORZENIA DANYCH REFERENCYJNYCH PODCZAS MIGRACJI

W rozdziale 15.1.2 Słowniki mapujące opisano mapowanie słowników.

Ponadto przy analizie ekstraktów systemów źródłowych zostaną określone zestawy kolumn stanowiące klucze pierwotne i obce, będą one używane początkowo do wiązania danych w systemie docelowym. Następnie za pomocą mechanizmów systemu docelowego zostaną one zastąpione przez klucze docelowe.

INNOBALTICA SP. Z O.O. 80-067 GDAŃSK, UL. RÓWNA 19/21  
NIP 957-10-03-404, REGON 220639884, BANK PKO BP SA 63 1020 1811 0000 0102 0311 6969  
KRS 0000311943, VII WYDZIAŁ GOSPODARCZY KRAJOWEGO REJESTRU SĄDOWEGO KAPITAŁ ZAKŁADOWY 21 305 000,00 PLN



Rzeczpospolita  
Polska

Unia Europejska  
Fundusz Spójności



Wylączna odpowiedzialność za treść niniejszej publikacji leży po stronie jej autorów. Jej treść nie musi odzwierciedlać opinii Unii Europejskiej. Ani Europejski Bank Inwestycyjny ani Komisja Europejska nie są odpowiedzialne za jakiegokolwiek wykorzystanie zawartych tu informacji.

Współfinansowane przez program ramowy Unii Europejskiej „Horyzont 2020”



Z.01

34 / 37

## 18 REGUŁY OBSŁUGI BŁĘDNYCH DANYCH, ZNAJDUJĄCYCH SIĘ W SYSTEMACH ŹRÓDŁOWYCH I MIGROWANYCH DO SYSTEMU DOCELOWEGO

Postępowanie z błędnymi (a także brakującymi) danymi, opisano w rozdziałach:

- 15.1.4 Reguły uzupełniania brakujących danych
- 15.1.5 Unikanie potencjalnie „niebezpiecznego” czyszczenia danych
- 16.3 Postępowanie z brakującymi lub błędnymi danymi

Reguły postępowania i skrypty zostaną opracowane w czasie testowych iteracji migracji. Będą wtedy wykrywane niespójne lub brakujące dane, oraz opracowywany sposób naprawy sytuacji.

Niektóre przypadki należy ręcznie wyczyścić i odpowiednio to udokumentować.

Błędne dane często można zastąpić danymi domyślnymi – należy tu jednak zachować szczególną ostrożność przy ważnych danych biznesowych.

Szczegółowy zestaw reguł zostanie przygotowany w trakcie iteracji testowych i przedstawiony za pomocą utworzonych Kart Wymagań na Jakość Danych.



## 19 KRYTERIA AKCEPTACJI POPRAWNOŚCI MIGRACJI

Kryteria akceptacji poprawności migracji, będą podstawą do opracowania planów testów akceptacyjnych (raport kontrolny) lub do wykonania migracji dla celów UAT (migracja próbki)

Kryteria te zostaną przedstawione w produkcie **W08.v1 Metoda lub narzędzie weryfikacji migracji danych zdefiniowane**.

### 19.1 SPRAWDZENIE KRYTERIÓW ROZPOCZĘCIA MIGRACJI

Sprawdzenie kryteriów rozpoczęcia Migracji Danych wykonywane jest zawsze jako pierwszy krok przeprowadzenia danej Iteracji Migracji Danych.

Czynności te należy traktować jako wykonanie zadań z listy kontrolnej (tzw. „check list”) niezbędnych do przeprowadzenia przed przystąpieniem do kolejnych kroków.

Brak spełnienia choćby jednego z kryteriów zamieszczonych na liście oznacza brak możliwości rozpoczęcia procesu.

Standardowa lista kontrolna (przygotowana w Planie Migracji) obejmuje co najmniej następujące czynności do sprawdzenia:

1. Czy jest przygotowany aktualny Plan Migracji z listą czynności i przypisanymi rolami?
2. Czy systemy źródłowe przygotowane są do wygenerowania ekstraktów?
3. Czy w gotowości i wymaganej dostępności są osoby wymagane do przeprowadzenia Migracji Danych?
4. Czy testowa Migracja Danych (UAT) została zaakceptowana? (punkt do spełnienia dla migracji produkcyjnej)
5. Czy jest przygotowane i dostępne Środowisko Migracji Danych?
6. Czy na Środowisku Migracji Danych zapewnione jest odbieranie danych z ekstraktów wejściowego produkcyjnego Systemu Źródłowego zgodnie z oczekiwanym stanem danych i wymaganiami przepustowości łącza?
7. Czy wszystkie niezbędne operacje w systemie wejściowym zostały wykonane?
8. Czy jest potwierdzenie, że produkcyjnie działający System DOCELOWY posiada funkcjonalność przewidzianą do uruchomienia w danym wydaniu?

Powyższa obligatoryjna lista może być rozszerzona w Planie Migracji Danych o dodatkowe pozycje do sprawdzenia.

Po pozytywnej weryfikacji kryteriów rozpoczęcia Migracji Danych nastąpi:

- Powiadomienie listy osób (zgodnie z Planem Migracji) o rozpoczęciu procesu Migracji Danych.



- Przystąpienie do kolejnego kroku procesu Migracji Danych.

W przypadku negatywnej weryfikacji przynajmniej jednego z kryteriów rozpoczęcia Migracji Danych nastąpi powiadomienie listy osób (zgodnie z Planem Migracji) o przyczynie odstąpienia od Migracji Danych.

Możliwe są następujące poziomy akceptacji Migracji Danych:

- akceptacja bez uwag,
- akceptacja z uwagami,
- brak akceptacji.

Akceptacja z uwagami oznacza przejście do kolejnego kroku Migracji Danych bez wstrzymywania procesu.

Brak akceptacji Migracji Danych oznacza przejście do kroku Wycofania Migracji.

Brak akceptacji Migracji Danych wymaga podania i uzasadnienia przyczyny.

## 19.2 WYCOFANIE MIGRACJI

Wycofanie Migracji Danych to zestaw czynności wykonywanych w celu przywrócenia działania systemów do stanu sprzed rozpoczęcia wykonania procesu Migracji Danych wówczas, gdy proces został awaryjnie - lub wskutek podjętej decyzji - przerwany. W zależności od momentu przerwania procesu Migracji Danych zestaw wymaganych czynności może się zdecydowanie różnić: od pustego zbioru do procesu przywracającego pierwotne działanie systemów.

Warunkiem wycofania danych jest posiadanie kopii bazy danych (lub baz danych) systemu DOCELOWEGO.

Wycofanie Migracji Danych jest czynnością administracyjną polegającą na przywróceniu bazy danych (lub baz danych) systemu DOCELOWEGO do stanu sprzed rozpoczęcia procesu Migracji.

Wycofanie Migracji Danych możliwe jest wyłącznie w oknie migracji, tzn. tylko wtedy, gdy zmianie uległy dane w wyniku migracji, a nie uruchomienia innych procesów lub działań użytkowników po migracji.

Szczegółowy opis wycofania migracji zostanie przedstawiony w dokumencie „**W05.v1 Plan Migracji dla Wydania.**”