Dokumentacja zbioru danych

Departament Analiz i Strategii

Narodowy Fundusz Zdrowia

Dane

Narodowy Fundusz Zdrowia dysponuje danymi sprawozdawczymi przekazywanymi przez podmioty lecznicze. Stanowią one między innymi podstawę do rozliczania świadczeń zdrowotnych. Przekazany w ramach wyzwania zbiór danych dotyczy świadczeń i leków refundowanych na rzecz pacjentów, spośród których u pewnego odsetka wystąpił udar niedokrwienny mózgu. Zbiór został przygotowany w taki sposób, że dane o świadczeniach i lekach pochodzą z okresu dwóch lat (*t*, *t+1*), natomiast informacja o wystąpieniu udaru dotyczy okresu kolejnych dwóch lat (*t+2*, *t+3*).

Przekazany zbiór danych składa się z 6 tabel zawierających rzeczywiste, zanonimizowane dane Narodowego Funduszu Zdrowia. Tabela 1 (pacjenci) zawiera dane definiujące czy u danego pacjenta wystąpił udar niedokrwienny w analizowanym okresie dwóch lat. Na potrzeby zapewnienia anonimizacji danych nie podano z jakiego okresu pochodzą dane. Pozostałe tabele zawierają informacje na temat historii świadczeń oraz realizacji recept dotyczących pacjentów z tabeli „pacjenci” z dwóch wcześniejszych lat .

Słownik pojęć:

1. ID\_EPIZODU - pole pozwalające na połączenie świadczenia z procedurami. W przypadku kiedy świadczenie było realizowane stacjonarnie (np. kilkudniowy pobyt w szpitalu) ID\_EPIZODU jest w relacji 1:1 z ID\_KSIEGI\_GLOWNEJ. W pozostałych przypadkach jest w relacji 1:1  
    z ID\_KONTAKTU.
2. ID\_KONTAKTU – pole pozwalające na połączenie świadczeń z rozpoznaniami. W przypadku świadczeń szpitalnych ID\_KONTAKTU odpowiada pobytowi na danym oddziale szpitalnym   
   w ramach jednej hospitalizacji identyfikowanej przez ID\_KSIEGI\_GLOWNEJ.
3. ID\_KSIEGI\_GLOWNEJ – identyfikator nadawany w sytuacji kiedy świadczenie było realizowane w trybie hospitalizacji. W ramach jednego identyfikatora hospitalizacji może wystąpić więcej niż 1 ID\_KONTAKTU. W takiej sytuacji ID\_KONTAKTU należy utożsamiać z pobytem na oddziale szpitalnym (tzn. hospitalizacja realizowana była na kilku oddziałach).
4. Zakres – grupa świadczeń wyodrębnionych ze względu na specjalność medyczną. W ramach zakresu sprawozdawane są produkty jednostkowe.
5. Produkt jednostkowy – produkt rozliczeniowy (kod) na podstawie którego świadczeniodawcy rozliczają się z płatnikiem. Jeden produkt może występować w wielu zakresach świadczeń[[1]](#footnote-1).   
   W polskim systemie ochrony zdrowia produkty zostały pogrupowane na katalogi. Do najważniejszych katalogów możemy zaliczyć:
   1. Katalog jednorodnych grup pacjentów (JGP) – jednorodna grupa pacjentów stanowi wyodrębnioną kategorię hospitalizacji posiadającą kod, nazwę, oraz charakterystykę, na którą składają się takie elementy, jak procedura medyczna, rozpoznanie (kod choroby), wiek pacjenta oraz czas pobytu. Produkty wchodzące w skład katalogu JGP stanowią najczęstszą podstawę do rozliczania hospitalizacji. Kody te rozpoczynają się od cyfr 5.51 np. 5.51.01.0001001 – *Zabiegi wewnątrzczaszkowe z powodu poważnego urazu*. Nadrzędną kategorią w strukturze systemu JGP jest sekcja stanowiąca zbiór jednorodnych grup pacjentów, występujących w układzie anatomiczno-fizjologicznym. Każda sekcja jest powiązana z obszarem klinicznym lub dziedziną medycyny   
      np.: Choroby układu nerwowego, Choroby narządu wzroku, Choroby układu mięśniowo-szkieletowego. Kody produktów wchodzące w skład danej sekcji można zidentyfikować na podstawie od 9 do 12 znaku kodu produktu np. w kodzie 5.51.01.0123018 będzie to „0123”[[2]](#footnote-2).
   2. Katalog produktów odrębnych – zawiera świadczenia, których nie można sparametryzować, np. Diagnostyka i leczenie chorego z niewydolnością wątroby   
      w okresie kwalifikacji do przeszczepu wątroby. Kody produktów wchodzące w skład katalogu rozpoczynają się od cyfr 5.52.
   3. Katalog produktów do sumowania - obejmuje świadczenia będące elementem kompleksowego postępowania, których nie można przyporządkować do odpowiedniej grupy ze względu na brak danych umożliwiających identyfikację grup posiadających oszacowane koszty, a których wartość należałoby zwiększyć o koszty dodatkowych czynności, np. żywienie dojelitowe, przetaczanie preparatów krwi, przeszczepy kostne allogeniczne sterylizowane radiacyjnie. Kody produktów wchodzące w skład katalogu produktów do sumowania rozpoczynają się od cyfr 5.53.
6. rodzaj świadczeń – kod wskazujący na miejsce udzielania świadczeń. Może on przyjąć wartości:
   1. PODSTAWOWA OPIEKA ZDROWOTNA;
   2. AMBULATORYJNE ŚWIADCZENIA SPECJALISTYCZNE;
   3. LECZENIE SZPITALNE;
   4. OPIEKA PSYCHIATRYCZNA I LECZENIE UZALEŻNIEŃ;
   5. REHABILITACJA LECZNICZA;
   6. OPIEKA DŁUGOTERMINOWA;
   7. LECZENIE STOMATOLOGICZNE;
   8. LECZNICTWO UZDROWISKOWE;
   9. POMOC DORAŹNA I TRANSPORT SANITARNY;
   10. PROFILAKTYCZNE PROGRAMY ZDROWOTNE.
7. typ komórki – kod wskazujący na rodzaj poradni/oddziału w którym udzielono dane świadczenie (VIII część kodu resortowego). Przykład przed procesem anonimizacji:   
   „4010 - ODDZIAŁ ALERGOLOGICZNY”
8. procedura – kod wg klasyfikacji ICD-9 wskazujący na to, jakich czynności udzielono pacjentowi

Opis poszczególnych tabel

1. pacjenci – informacja o wystąpieniu pierwszorazowego udaru niedokrwiennego   
   w analizowanym przedziale czasowym (2 lata). Pacjenci u których odnotowano udar niedokrwienny przed analizowanym przedziałem czasowym zostali wykluczeni z analizy. Składa się ona z kolumn:
   1. ID\_PACJENTA – unikalny identyfikator pacjenta.
   2. CZY\_UDAR – binarna informacja czy w przypadku osoby o danym identyfikatorze wystąpiła jednostka chorobowa podlegająca analizie.
2. swiadczenia – informacja o udzielonych świadczeniach. Składa się ona z kolumn:
   1. ID\_PACJENTA – unikalny identyfikator pacjenta
   2. ID\_EPIZODU – identyfikator epizodu pozwalający na połączenie świadczeń   
      z procedurami
   3. ID\_KONTAKTU – identyfikator pozwalający na połączenie świadczeń   
      z rozpoznaniami.
   4. ID\_KSIEGI\_GLOWNEJ – identyfikator księgi głównej. W skład jednego identyfikatora księgi głównej może wchodzić więcej niż jeden identyfikator kontaktu. Brakująca wartość (NA) oznacza, że dane świadczenie nie zostało rozliczone w trybie hospitalizacji
   5. ROK\_SWIADCZENIA – rok udzielenia świadczenia. Kolumna może przyjąć wartości:
      1. -1 – świadczenie udzielone w roku przed okresem w którym została zebrana informacja o pacjentach u których wystąpił incydent związany z udarem
      2. -2 – świadczenie udzielone w 2-gim roku przed okresem w którym została zebrana informacja o pacjentach u których wystąpił incydent związany   
         z udarem
   6. TYDZIEN\_POCZATKU\_KONTAKTU – tydzień roku (wartości od 1 do 53) rozpoczęcia kontaktu (ID\_KONTAKTU)
   7. TYDZIEN\_KONCA\_KONTAKTU - tydzień roku (wartości od 1 do 53) zakończenia kontaktu (ID\_KONTAKTU)
   8. KOD\_ZAKRESU – zanonimizowany kod zakresu świadczenia
   9. KOD\_PRODUKTU\_JEDNOSTKOWEGO – zanonimizowany kod produktu jednostkowego umożliwiający w przypadku katalogu JGP grupowanie kodów   
      do poszczególnych sekcji. Zaszyfrowany kod przyjmuje postać: A.AA.AA.BBBBCCC, gdzie:
      1. A.AA.AA – odpowiada pierwszej części kodu, która nie została zanonimizowana,
      2. BBBB – zanonimizowana, druga część kodu, która w przypadku katalogu JGP (produkty rozpoczynające się od 5.51) odpowiada za sekcję,
      3. CCC – zanonimizowana część kodu odpowiadająca trzem ostatnim znakom oryginalnego kodu.

Przykład anonimizacji:

1. kody oryginalne- 5.51.01.0003007, 5.51.01.0003012
2. kod zanonimizowany- 5.51.01.6321455, 5.51.01.6321698.

Podkreślony ciąg cyfr odpowiadający za sekcję kodów JGP (w tym przypadku choroby twarzy, jamy ustnej, gardła, krtani, nosa i uszu) po procesie anonimizacji nadal pozostaje unikalny w obrębie tej sekcji.

* 1. TYP\_KOMORKI – zanonimizowany kod specjalności komórki (VIII część kodu resortowego). Przykład przed procesem anonimizacji:  
     „4010 - ODDZIAŁ ALERGOLOGICZNY”
  2. RODZAJ\_SWIADCZENIA – zmienna tekstowa definiująca rodzaj miejsca udzielenia świadczenia (leczenie szpitalne, podstawowa opieka zdrowotna, ratownictwo medyczne itd.)
  3. KWOTA\_ROZLICZONA – kwota, jaką fundusz zapłacił podmiotowi za udzielone świadczenie. W przypadku świadczeń rozliczonych w ramach stawki kapitacyjnej lub ryczałtowo kwota ta wynosi 0

1. procedury – tabela zawierająca informacje o wykonanych procedurach medycznych. Składa się ona z kolumn:
   1. ID\_EPIZODU – Identyfikator epizodu pozwalający na połączenie świadczeń  
      z procedurami.
   2. KOD\_PROCEDURY – zanonimizowany kod procedury według klasyfikacji ICD-9. Zaszyfrowany kod przyjmuje postać: AAA.BCDE, gdzie:
      1. AA – odpowiada znakom znajdującym się przed kropką w oryginalnym kodzie,
      2. B – odpowiada pierwszemu znakowi po kropce,
      3. C – odpowiada drugiemu znakowi po kropce,
      4. D – odpowiada trzeciemu znakowi po kropce,
      5. E – odpowiada czwartemu znakowi po kropce,

Sprawozdawanie znaków po kropce jest opcjonalne. Oznacza to, że zaszyfrowane kody różnią się co do długości. Jeżeli zaszyfrowany kod składa się tylko z części AAA, oznacza to sytuację w której oryginalny kod składał się tylko z znaków znajdujących się przed kropką w oryginalnym kodzie. Sposób szyfrowania umożliwia grupowanie kodów do kolejnych części kodu. Przykład anonimizacji:

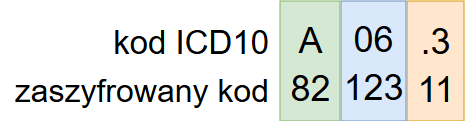
|  |  |
| --- | --- |
| Oryginalny kod | Zaszyfrowany kod |
| 88 | 534 |
| 88.55 | 534.32 |
| 88.59 | 534.37 |
| 88.592 | 534.379 |

Wszystkie z powyższych kodów należą do tej samej grupy (88 – „Inne diagnostyczne badania radiologiczne i pokrewne”). Istnieje możliwość uogólnienia wszystkich kodów do tej grupy poprzez ucięcie wszystkich znaków po kropce. Poprzez obcinanie kolejnych znaków możemy dobrać stopień szczegółowości kodów np. pozostawienie znaków AAA.B doprowadzi do uzyskania 2 grup kodów :

* „88 - Inne diagnostyczne badania radiologiczne i pokrewne” (534 dla zaszyfrowanego kodu) w przypadku kiedy sprawozdano kod 2 znakowy)
* „88.5 - Angiokardiografia przy użyciu kontrastu” (534.3 dla zaszyfrowanego kodu) w przypadku kiedy sprawozdano kod dłuższy niż 2 znakowy

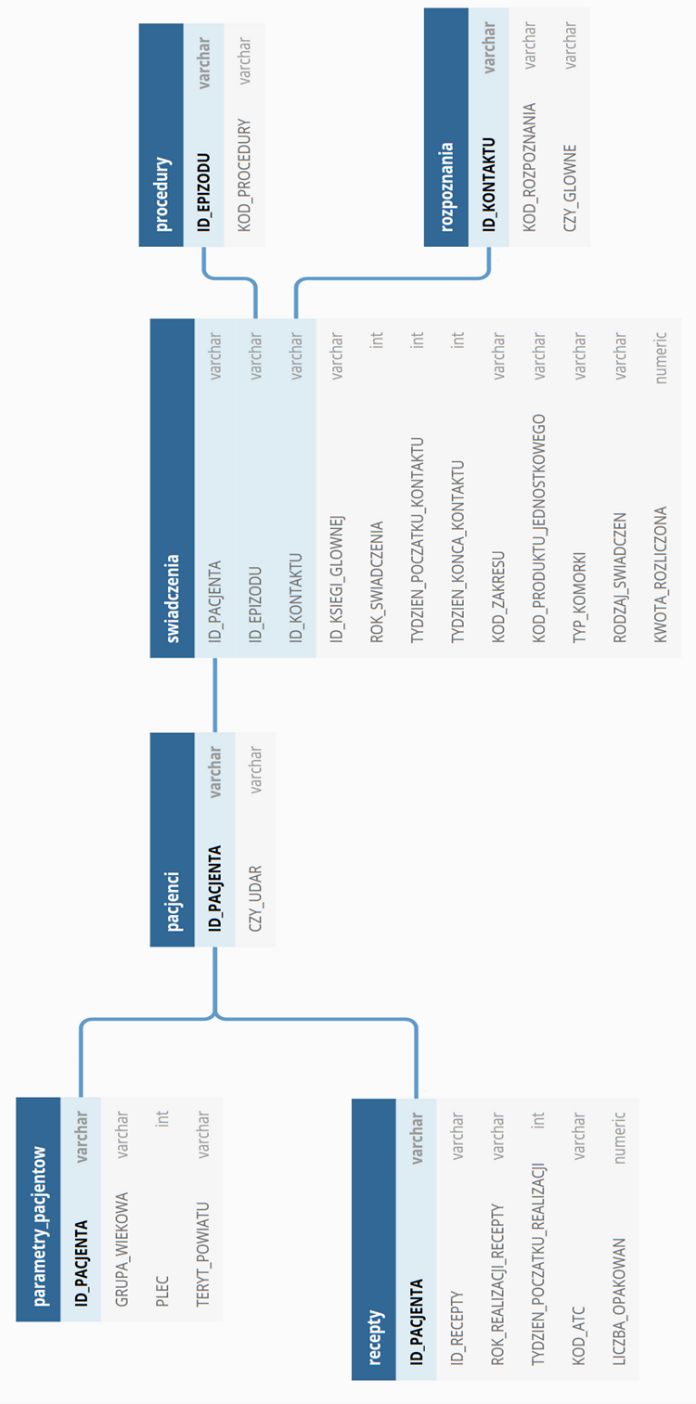
1. recepty – tabela zawiera informacje o zrealizowanych receptach na leki refundowane
   1. ID\_PACJENTA - unikalny identyfikator pacjenta.
   2. ID\_RECEPTY – identyfikator recepty
   3. ROK\_REALIZACJI\_RECEPTY - rok realizacji recepty. Kolumna może przyjąć wartości:
      1. -1 – recepta zrealizowana w roku przed okresem w którym została zebrana informacja o pacjentach u których wystąpił incydent związany z udarem
      2. -2 – recepta zrealizowana w 2-gim roku przed okresem w którym została zebrana informacja o pacjentach u których wystąpił incydent związany   
         z udarem
   4. TYDZIEN \_REALIZACJI – tydzień roku (wartości od 1 do 53) realizacji recepty
   5. KOD\_ATC – 3 znaki kodu wg klasyfikacji anatomiczno-terapeutyczno-chemicznej (ATC)
   6. LICZBA\_OPAKOWAN – wykupiona liczba opakowań leku o danym kodzie ATC
2. rozpoznania – tabela zawiera informacje na temat sprawozdanych rozpoznań według klasyfikacji ICD-10.
   1. ID\_KONTAKTU – identyfikator pozwalający na połączenie świadczeń   
      z rozpoznaniami
   2. KOD\_ROZPOZNANIA – zaszyfrowany kod rozpoznania wg Międzynarodowej Klasyfikacji Chorób i Problemów Zdrowotnych ICD10. Oryginalnie kod może składać się z 3 znaków (np. A06 Pełzakowica ) lub z 5 znaków, co doszczegółowia rozpoznanie (np. A06.3 - Guzowate nacieczenie ściany jelita). Zaszyfrowany kod przyjmuje postać: AABBBCC, gdzie:
      1. AA – odpowiada pierwszemu znakowi kodu,
      2. BBB – odpowiada drugiemu i trzeciemu znakowi,
      3. CC – odpowiada znakom znajdującym się po znaku kropki w oryginalnym kodzie.

Jeżeli zaszyfrowany kod składa się tylko z części AABBB, oznacza to sytuację w której oryginalny kod składał się tylko z pierwszych trzech znaków kodu. Sposób szyfrowania umożliwia grupowanie kodów do jednego oraz trzech znaków kodu ICD10. Świadczeniodawcy mają możliwość sprawozdania zarówno trzyznakowego kodu ICD10 np. A06 (po zaszyfrowaniu kod pięcioznakowy - 82123) jak i pięcioznakowych kodów ICD10 np. A06.3, A06.4 (po zaszyfrowaniu odpowiednio kody siedmioznakowe – 8212311, 8212364). Kody pięcioznakowe można doprowadzić do postaci A06 poprzez usunięcie dwóch ostatnich cyfr zaszyfrowanego kodu. Istnieje również możliwość zagregowania powyższych kodów do postaci jednoznakowej. Poniżej zaprezentowano sposób anonimizacji:



* 1. CZY\_GLOWNE – zmienna binarna o poziomach ‘N’ – rozpoznanie współistniejące oraz ‘T’ – rozpoznanie główne

1. parametry\_pacjentow – tabela zawiera informacje o parametrach pacjenta.
   1. ID\_PACJENTA – unikalny identyfikator pacjenta.
   2. GRUPA\_WIEKOWA – wiek pacjenta przedstawiony za pomocą 5. letnich przedziałów wiekowych
   3. PLEC – płeć pacjenta
   4. TERYT\_POWIATU – kod terytorialny miejsca zameldowania pacjenta przedstawiony do poziomu powiatu



1. Więcej o specyfice produktów rozliczeniowych można dowiedzieć się z materiałów: https://akademia.nfz.gov.pl/wp-content/uploads/2016/03/JGP\_2\_11\_07\_2010.pdf https://akademia.nfz.gov.pl/materialy-szkoleniowe-na-temat-jednorodnych-grup-pacjentow-jgp/ [↑](#footnote-ref-1)
2. Listę kodów produktów z podziałem na sekcje można znaleźć w załączniku 1a zarządzenia Prezesa

   Narodowego Funduszu Zdrowia w sprawie określenia warunków zawierania i realizacji umów w rodzaju leczenie szpitalne oraz leczenie szpitalne – świadczenia wysokospecjalistyczne: https://www.nfz.gov.pl/zarzadzenia-prezesa/zarzadzenia-prezesa-nfz/zarzadzenie-nr-382019dsoz,6906.html [↑](#footnote-ref-2)