



Opis przedmiotu zamówienia

Rozwiązanie technologiczne do automatycznej
ekstrakcji danych z dokumentów PDF
– ocena wykonalności i szacunkowa wycena

Sygnatura: ASOZ_P3_AOTMiT

Data dokumentu: 28.10.2025 r.

Wykaz wybranych skrótów

Agencja/AOTMiT	Agencja Oceny Technologii Medycznych i Taryfikacji
ATC	klasa anatomiczno-terapeutyczna produktu
ChPL	Charakterystyka Produktu Leczniczego, o której mowa w ...
digitalizacja	proces przekształcania informacji zapisanej w formie elektronicznych dokumentów PDF lub tekstowych w postać cyfrowego rekordu danych, możliwego do przechowywania, przetwarzania i przesyłania przy użyciu systemów informatycznych
EMA	Europejska Agencja Leków
GBD	projekt pod nazwą Global Burden of Disease
GTIN	numer jednostki handlowej, stosowany w systemie GS1, unikalny identyfikator przypisany do produktu, który umożliwia jednoznaczną identyfikację
HTA	ocena technologii medycznych
KE	Komisja Europejska
MZ	Ministerstwo Zdrowia
NFZ	Narodowy Fundusz Zdrowia
RPL	Rejestr Produktów Leczniczych
System/SCIoTM	System Centralnej Informacji o Technologiach Medycznych
UE	Unia Europejska
URPL	Urząd Rejestracji Produktów Leczniczych, Wyrobów Medycznych i Produktów Biobójczych, będący administratorem danych przetwarzanych w Rejestrze Produktów Leczniczych

Spis treści

Wykaz wybranych skrótów	2
Spis treści	3
1. Wprowadzenie	4
1.1. Charakterystyka zamawiającego	4
1.2. Zawartość opracowania	4
2. Podłoże i uzasadnienie biznesowe	5
2.1. Uzasadnienie	5
2.2. Problemy w zakresie źródeł danych	5
2.3. Zakres informacyjny Rozwiązania	5
3. Charakterystyka źródeł informacji	6
3.1. Rejestr Produktów Leczniczych	6
3.2. Komisja Europejska – Union Register of medicinal products	6
4. Charakterystyka produktów leczniczych	7
4.1. Struktura dokumentu	7
4.2. Charakterystyka atrybutów	10
4.3. Założenia i ograniczenia biznesowe	12
5. Ekstrakcja danych – opis procesu	13
5.1. Etapy	13
5.2. Identyfikacja danych	13
5.3. Przetwarzanie danych	14
5.4. Kontrola jakości	16
5.5. Aktualizacja danych	16

1. Wprowadzenie

1.1. Charakterystyka zamawiającego

Agencja Oceny Technologii Medycznych i Taryfikacji jest opiniodawczo-doradczą jednostką organizacyjną posiadającą osobowość prawną, nadzorowaną przez ministra właściwego ds. zdrowia. Rolą Agencji jest wspomaganie ministra właściwego ds. zdrowia w procesie podejmowania decyzji dotyczących finansowania świadczeń lekowych i nielekowych w systemie ochrony zdrowia.

Agencja Oceny Technologii Medycznych i Taryfikacji (AOTMiT) realizuje zadania wynikające ze zleceń Ministra Zdrowia, obejmujące m.in. przygotowywanie analiz weryfikacyjnych, analiz związanych z wykazami refundacyjnymi (m.in. 65 PLUS, 18 MINUS), analiz leków o ugruntowanej pozycji, analizy horyzontalne (horizon scanning) oraz oceny efektywności klinicznej technologii medycznych. Wykonywanie tych zadań wymaga dostępu do rzetelnych, aktualnych i wszechstronnych danych o produktach leczniczych.

1.2. Zawartość opracowania

Dokument zawiera:

- opis przedmiotu zamówienia stanowiącego: rozwiązanie technologiczne do automatycznej ekstrakcji danych z dokumentów PDF (Charakterystyk Produktów Leczniczych, dalej w skrócie ChPL¹) i konwersji ich do ustrukturyzowanych rekordów do zapisania ich w bazie danych zwane dalej Rozwiązaniem;
- podłoże i uzasadnienie biznesowe;
- określenie zakresu pozyskiwanych informacji z ChPL.

Przedstawiona treść ma umożliwić:

- **ocenie wykonalności oczekiwanego Rozwiązania,**
- **wstępnej wycenie wytworzenia Rozwiązania, z horyzontem realizacji maksymalnie 6 miesięcy.**

Dokument skierowany jest do potencjalnych wykonawców specjalizujących się w obszarze związanym z przedmiotowym Rozwiązaniem.

¹ patrz rozdz. 4. *Charakterystyka produktów leczniczych*

2. Podłoże i uzasadnienie biznesowe

2.1. Uzasadnienie

Realizacja ustawowych zadań AOTMiT obejmuje sporządzanie raportów oceny technologii medycznych dla terapii lekowych i nielekowych oraz wydawanie stanowisk i rekomendacji. Jednym z kluczowych etapów analiz jest ocena problemu decyzyjnego, która wymaga szybkiego dostępu do:

- aktualnych wskazań rejestracyjnych i ich zmian,
- statusu refundacyjnego produktu leczniczego,
- informacji o alternatywach terapeutycznych,
- zakresu wcześniejszych ocen OP/RP i rekomendacji,
- danych dotyczących bezpieczeństwa i skuteczności.

Obecny model pracy w Agencji zakłada **manualne pozyskiwanie tych danych** z wielu źródeł o niejednorodnej strukturze. Powoduje to:

- duże obciążenie analityków,
- wydłużenie czasu realizacji analiz,
- ryzyko niespójności danych przy różnych osobach pozyskujących dane,
- brak scentralizowanej bazy dla ponownego wykorzystania zebranych informacji.

W przypadku złożonych zleceń obejmujących wiele leków oraz wskazań konieczne jest zaangażowanie dodatkowych osób jedynie do pozyskania danych wejściowych. Stąd potrzeba wdrożenia algorytmu digitalizującego wynik bezpośrednio z dążenia do automatyzacji prac analitycznych i poprawy jakości danych.

2.2. Problemy w zakresie źródeł danych

Zidentyfikowane kluczowe kwestie problemowe:

- URPL nie posiada bazy danych zawierającej dane z ChPL w formacie umożliwiającym cyfrowe przeszukiwanie,
- Rejestr Produktów Leczniczych (RPL) nie udostępnia historii wersji ChPL w procedurach DCP/MRP/NAR,
- dokumenty udostępniane są wyłącznie w formacie PDF,
- dla produktów z importu równoległego nie publikuje się ChPL, a jedynie ulotki.
- Europejska Agencja Leków prowadzi obecnie projekt Electronic Product Information (ePI), (dostęp: <https://www.ema.europa.eu/en/human-regulatory-overview/marketingauthorisation/product-information-requirements/electronic-product-information-epi>), ePI zakłada standaryzację elektronicznego formatu ChPL opartą na FHIR,
- brak jest harmonogramu wdrożenia dla wszystkich produktów dostępnych w Polsce.

Powyższe potwierdza konieczność opracowania **narzędzia do digitalizacji ChPL** na poziomie krajowym, aby zapewnić aktualne dane na potrzeby analiz AOTMiT.

2.3. Zakres informacyjny Rozwiązania

Planowane Rozwiązanie obejmie co najmniej:

1. **Zakres produktów:** leki stosowane u ludzi (bez produktów leczniczych weterynaryjnych).
2. Identyfikacja źródła danych rejestracyjnych wg. procedury rejestracyjnej:
 - 1) **dla procedury scentralizowanej (CEN):** pobieranie ChPL z oficjalnych źródeł anglojęzycznej strony Komisji Europejskiej/EMA wraz z pełną historią wersji dokumentu w języku polskim,
 - 2) **dla procedur: narodowej (NAR), wzajemnego uznania (MRP), zdecentralizowanej (DCP):** pozyskiwanie aktualnych dokumentów z RPL/URPL (bez publicznej historii wersji; historia utrzymywana po stronie SCIoT na podstawie kolejnych pobrań).
3. **Ekstrakcja danych z ChPL w formacie PDF i konwersja do ustrukturyzowanych rekordów elektronicznych do docelowego zapisania w bazie danych** (zapis w bazie danych nie jest przedmiotem przedmiotowego zamówienia).

3. Charakterystyka źródeł informacji

Charakterystyki produktów leczniczych (ChPL) są oparte przede wszystkim na oficjalnych rejestrach i dokumentach wydawanych przez krajowe i europejskie organy regulacyjne. W Polsce kluczowym źródłem jest Rejestr Produktów Leczniczych Dopuszczonych do Obrotu. Do tego rejestru wpisane są:

- pozwolenia wydane w procedurach narodowych (NAR) – dokumentacja sporządzona i oceniona wyłącznie przez URPL (Urząd Rejestracji Produktów Leczniczych, Wyrobów Medycznych i Produktów Biobójczych),
- pozwolenia wzajemnego uznania ((MRP) i zdecentralizowanego DCP) – oparte na wspólnej ocenie co najmniej dwóch państw członkowskich UE z koordynacją Urzędu,
- pozwolenia scentralizowane (CEN), wydawane bezpośrednio przez Komisję Europejską w oparciu o opinię Europejskiej Agencji Leków (EMA),
- pozwolenia na import równoległy (IR), pozwalające na wprowadzenie do obrotu leków wytworzonych i zarejestrowanych w innych krajach UE.

Szczegółowe informacje w rozdz. 5.2 *Identyfikacja danych*.

3.1. Rejestr Produktów Leczniczych

Rejestr Produktów Leczniczych zawiera informacje o wszystkich produktach leczniczych przeznaczonych dla ludzi oraz weterynaryjnych dopuszczonych do obrotu na terytorium Rzeczypospolitej Polskiej. Gromadzi podstawowe informacje o produktach leczniczych, a także druki informacyjne.

Rejestr Produktów Leczniczych - Rejestry e-zdrowia

Podstawa prawna działania rejestru

- rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z 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 (ogólne rozporządzenie o ochronie danych) (RODO),
- ustawa z dnia 6 września 2001 r. – Prawo farmaceutyczne,
- rozporządzenie Ministra Zdrowia z dnia 17 kwietnia 2019 r. w sprawie sposobu i trybu prowadzenia oraz udostępniania Rejestru Produktów Leczniczych Dopuszczonych do Obrotu na terytorium Rzeczypospolitej Polskiej oraz struktury danych udostępnianych z tego rejestru,
- ustawa z dnia 28 kwietnia 2011 r. o systemie informacji w ochronie zdrowia,
- ustawa z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne.

Aktualny rejestr produktów leczniczych - wersja 6.0.0

<https://rejestry.ezdrowie.gov.pl/api/rpl/medicinal-products/public-pl-report/6.0.0/overall.xml>

<https://rejestry.ezdrowie.gov.pl/api/rpl/xsd/2.0.0/commonTypes.xsd>

3.2. Komisja Europejska – Union Register of medicinal products

W rejestrze unijnym znajduje się wykaz wszystkich produktów leczniczych stosowanych u ludzi i do celów weterynaryjnych, a także sierocych produktów leczniczych, które otrzymały pozwolenie na dopuszczenie do obrotu przez Komisję w ramach procedury scentralizowanej.

https://ec.europa.eu/health/documents/community-register/html/index_en.htm

Informacje gromadzone i przetwarzane w rejestrze obejmują: nazwę produktu leczniczego, numer rejestracyjny UE, nazwę i adres posiadacza pozwolenia na dopuszczenie do obrotu, substancję czynną, międzynarodową niezastrzeżoną prawnie nazwę, kod anatomiczno-terapeutyczno-chemiczny (ATC) oraz wskazanie terapeutyczne wraz z odpowiednimi dokumentami.

Oprócz produktów leczniczych, które znajdują się obecnie na rynku UE, w Rejestrze znajdują się informacje o:

- produktach leczniczych, które zostały zawieszone lub wycofane, lub w odniesieniu do których odmówiono wydania pozwolenia na dopuszczenie do obrotu;
- produktach leczniczych stosowane u ludzi i do celów weterynaryjnych przyjętych na szczeblu krajowym, w odniesieniu do których konieczna była decyzja Komisji;
- produktach leczniczych dopuszczonych do obrotu przez poszczególne kraje UE na podstawie art. 126a dyrektywy 2001/83;
- indeks substancji czynnych objętych decyzjami Komisji;
- spis produktów leczniczych wymienionych przy nazwie handlowej objętych decyzjami Komisji.

4. Charakterystyka produktów leczniczych

4.1. Struktura dokumentu

Zgodnie z definicją legalną określoną w ustawie z dnia 6 września 2001 r. Prawo farmaceutyczne (Dz. U. z 2025 r. poz. 750, 905, 924) produktem leczniczym – jest substancja lub mieszanina substancji, przedstawiana jako posiadająca właściwości zapobiegania lub leczenia chorób występujących u ludzi lub zwierząt lub podawana w celu postawienia diagnozy lub w celu przywrócenia, poprawienia lub modyfikacji fizjologicznych funkcji organizmu poprzez działanie farmakologiczne, immunologiczne lub metaboliczne. Produktem radiofarmaceutycznym – jest produkt leczniczy, z wyłączeniem produktu leczniczego weterynaryjnego, który zawiera jeden lub więcej izotopów radioaktywnych przeznaczonych dla celów medycznych.

Obligatoryjne elementy wniosku o dopuszczenie do obrotu produktu leczniczego zostały określone przez art. 10 tejże ustawy, w którym to ust. 2 pkt. 11 wskazuje na Charakterystykę Produktu Leczniczego.

Według ustawy Charakterystyka Produktu Leczniczego, o której mowa w art. 10 ust. 2 pkt 11, zawiera:

1. *nazwę produktu leczniczego, wraz z określeniem mocy produktu leczniczego i postaci farmaceutycznej;*
2. *skład jakościowy i ilościowy w odniesieniu do substancji czynnych oraz tych substancji pomocniczych, w przypadku których informacja ta jest istotna dla właściwego podawania danego produktu leczniczego, przy czym używa się nazwy powszechnie stosowanej lub nazwy chemicznej;*
3. *postać farmaceutyczną;*
4. *dane kliniczne obejmujące:*
 - a. *wskazania do stosowania,*
 - b. *dawkowanie i sposób podawania dorosłym oraz dzieciom, w przypadku stosowania produktu leczniczego u dzieci,*
 - c. *przeciwwskazania,*
 - d. *specjalne ostrzeżenia i środki ostrożności przy stosowaniu, a w przypadku immunologicznych produktów leczniczych – specjalne środki ostrożności podejmowane przez osoby mające styczność z takimi produktami oraz specjalne środki ostrożności przy podawaniu pacjentom, wraz ze środkami ostrożności, które powinny być podjęte przez pacjenta,*
 - e. *interakcje z innymi produktami leczniczymi lub inne formy interakcji,*
 - f. *stosowanie w czasie ciąży lub karmienia piersią,*
 - g. *wpływ na zdolność prowadzenia pojazdów i obsługiwanie urządzeń mechanicznych,*
 - h. *działania niepożądane,*
 - i. *przedawkowanie, w tym jego objawy, sposób postępowania przy udzielaniu natychmiastowej pomocy oraz antidota;*
5. *właściwości farmakologiczne obejmujące:*
 - a. *właściwości farmakodynamiczne,*
 - b. *właściwości farmakokinetyczne,*
 - c. *niekliniczne dane o bezpieczeństwie;*
6. *dane farmaceutyczne obejmujące:*
 - a. *wykaz substancji pomocniczych,*
 - b. *główne niezgodności farmaceutyczne,*
 - c. *okres ważności oraz jeżeli to konieczne, okres przydatności po rekonstytucji produktu leczniczego lub po pierwszym otwarciu jego opakowania bezpośredniego,*
 - d. *specjalne środki ostrożności przy przechowywaniu,*
 - e. *rodzaj opakowania i skład materiałów, z których je wykonano,*
 - f. *specjalne środki ostrożności przy unieszkodliwianiu zużytego produktu leczniczego lub odpadów powstających z produktu leczniczego, jeżeli ma to zastosowanie;*
7. *nazwę i adres podmiotu odpowiedzialnego;*
8. *numer pozwolenia na dopuszczenie do obrotu;*
9. *datę wydania oraz numer pierwszego pozwolenia na dopuszczenie do obrotu lub przedłużenia ważności pozwolenia;*
10. *datę ostatniej aktualizacji tekstu Charakterystyki Produktu Leczniczego;*

11. w przypadku produktów radiofarmaceutycznych także:
 - a. informacje na temat dozymetrii napromieniowania wewnętrznego,
 - b. wskazówki dotyczące bezpośredniego przygotowania i kontroli jakości produktu oraz jeżeli to konieczne, maksymalny okres przechowywania, w którym eluat lub produkt radiofarmaceutyczny gotowy do użytku zachowują swoje właściwości zgodnie ze specyfikacją;
12. informację dla osób wykonujących zawód medyczny o obowiązku albo uprawnieniu zgłaszania pojedynczych przypadków działań niepożądanych danego produktu leczniczego ze względu na konieczność nieprzerwanego monitorowania jego stosowania;
13. w przypadku produktów leczniczych podlegających dodatkowemu monitorowaniu:
 - a. wskazanie odpowiedniego czarnego symbolu, wybranego zgodnie z art. 23 rozporządzenia nr 726/2004/WE,
 - b. sformułowanie o następującej treści: „Ten produkt leczniczy podlega dodatkowemu monitorowaniu” uzupełnione o dodatkowe wyjaśnienia.

Stan faktyczny – struktura dokumentów EMA i URPL wraz z zawartością merytoryczną

1. Nazwa produktu leczniczego
 - Zawiera pełną nazwę handlową produktu.
 - Uwzględnia postać farmaceutyczną (np. roztwór do wstrzykiwań, tabletki powlekane) oraz stężenie substancji czynnej.
2. Skład jakościowy i ilościowy
 - Informuje o:
 - Substancji czynnej – jej nazwa i ilość wyrażona w odpowiednich jednostkach (np. mg, MBq, GBq/mL).
 - Substancjach pomocniczych o znanym działaniu – np. sól, laktoza, etanol.
 - Odniesieniu do pełnego wykazu pomocniczego w punkcie 6.1.
3. Postać farmaceutyczna
 - Opisuje wygląd produktu (np. „bezbardwy, przejrzysty roztwór”).
 - Wskazuje, czy produkt ma formę roztworu, proszku, tabletki, zawiesiny itp.
4. Szczegółowe dane kliniczne
 - 4.1. Wskazania do stosowania
 - Zawiera zatwierdzone wskazania terapeutyczne, czyli w jakich chorobach lub stanach klinicznych lek może być stosowany.
 - Może zawierać dodatkowe ograniczenia (np. wyłącznie dla dorosłych).
 - 4.2. Dawkowanie i sposób podawania
 - Dawkowanie dla różnych grup pacjentów, np. dorośli, dzieci, osoby starsze, pacjenci z niewydolnością nerek/wątroby.
 - Częstotliwość podania i zasady dostosowywania dawek.
 - Sposób podania (doustny, dożylny, wziewny itp.)
 - Szczególne środki ostrożności przed podaniem, np. pomiar aktywności radiofarmaceutyku, nawodnienie pacjenta.
 - 4.3. Przeciwwskazania
 - Lista sytuacji, w których lek nie może być stosowany, np. nadwrażliwość, ciąża, niewydolność wątroby.
 - 4.4. Specjalne ostrzeżenia i środki ostrożności
 - Wskazania dotyczące szczególnego nadzoru, środków przygotowawczych, możliwych powikłań.
 - Uwzględnia także zalecenia dotyczące ekspozycji na promieniowanie (dla radiofarmaceutyków).

4.5. Interakcje z innymi produktami leczniczymi i inne rodzaje interakcji

- Opisuje możliwe interakcje z:
 - innymi lekami,
 - jedzeniem lub napojami,
 - procedurami diagnostycznymi.

4.6. Wpływ na płodność, ciążę i laktację

- Zasady stosowania leku u kobiet w ciąży i karmiących piersią.
- Potencjalny wpływ na płodność.
- Czasowe zalecenia dotyczące przerywania karmienia piersią lub odroczenia terapii.

4.7. Wpływ na zdolność prowadzenia pojazdów i obsługiwanie maszyn

- Czy lek wpływa na zdolności psychofizyczne, np. poprzez senność, zawroty głowy, zaburzenia koordynacji.

4.8. Działania niepożądane

- Lista znanych działań niepożądanych, pogrupowana według częstości występowania.
- Uwagi o ekspozycji na promieniowanie, jeśli dotyczy.
- Informacje o systemie zgłaszania działań niepożądanych.

4.9. Przedawkowanie

- Informacje o objawach przedawkowania.
- Sugerowane metody postępowania (np. diureza, płukanie żołądka, leczenie objawowe).

5. Właściwości farmakologiczne

5.1. Właściwości farmakodynamiczne

- Grupa terapeutyczna i kod ATC.
- Mechanizm działania leku.
- Informacje o selektywności, powinowactwie, efekcie terapeutycznym.

5.2. Właściwości farmakokinetyczne

- Wchłanianie (biodostępność),
- Dystrybucja (np. wiązanie z białkami),
- Metabolizm (enzymy, przemiany w wątrobie),
- Eliminacja (np. przez nerki, z kałem),
- Okres półtrwania.

5.3. Przedkliniczne dane o bezpieczeństwie

- Dane z badań na zwierzętach (toksyczność ostra, przewlekła, mutagenność, rakotwórczość).
- Opis dawki bezpiecznej.

6. Dane farmaceutyczne

6.1. Wykaz substancji pomocniczych

- Wszystkie składniki niefarmakologicznie czynne zawarte w produkcie.

6.2. Niezgodności farmaceutyczne

- Informacja o lekach/substancjach, z którymi produkt nie powinien być mieszany.

6.3. Okres ważności

- Termin ważności gotowego produktu.
- Okres przydatności po otwarciu lub rozcieńczeniu.

- 6.4. Specjalne środki ostrożności podczas przechowywania
 - Zalecana temperatura przechowywania.
 - Informacje o konieczności ochrony przed światłem lub mrozem.
- 6.5. Rodzaj i zawartość opakowania
 - Opis fiolki, ampułki lub blistra.
 - Zakres objętości (np. 5–15 mL) lub ilość tabletek.
- 6.6. Specjalne środki ostrożności dotyczące usuwania
 - Instrukcje dla personelu odnośnie do bezpiecznego usuwania produktu, pozostałości i opakowań.
 - Dodatkowe zalecenia dla produktów promieniotwórczych.
7. Podmiot odpowiedzialny posiadający pozwolenie na dopuszczenie do obrotu
 - Pełna nazwa i adres firmy farmaceutycznej odpowiedzialnej za produkt.
8. Numer pozwolenia na dopuszczenie do obrotu
 - Numer nadany przez krajowy organ rejestracyjny.
9. Data wydania pierwszego pozwolenia / data przedłużenia pozwolenia
 - Data dopuszczenia do obrotu.
 - Ewentualnie: data ostatniego odnowienia.
10. Data zatwierdzenia lub częściowej zmiany tekstu
 - Ostatnia oficjalna aktualizacja dokumentu.
11. Dozymetria (*dla produktów radiofarmaceutycznych*)
 - Tabela dawek pochłoniętych przez poszczególne narządy (mGy/MBq).
 - Dawka efektywna w zależności od masy ciała i trybu obrazowania (2D/3D).
12. Instrukcja przygotowania radiofarmaceutyków (*dla produktów radiofarmaceutycznych*)
 - Procedura przygotowania leku do użycia.
 - Zasady rozcieńczania, pomiarów aktywności, zachowania aseptyki.
 - Wskazówki dotyczące postępowania z fiolkami, sprzętem i odpadami.

4.2. Charakterystyka atrybutów

Celem identyfikacji rzeczywistej charakterystyki atrybutów przeprowadzono przegląd dokumentów ChPL, na zasadzie manualnej digitalizacji charakterystyk w trzech losowo wybranych grupach leków:

- Grupa I – różnego rodzaju standardowe produkty lecznicze (jedna droga podania leku, produkt nie będący radiofarmaceutykiem, uwzględniono dodatkowo charakterystykę weterynaryjnego produktu leczniczego);
- Grupa II – jeden produkt, dwie drogi podania;
- Grupa III – radiofarmaceutyki.

Przeprowadzono analizę atrybutów występujących w charakterystykach produktów leczniczych (ChPL), uwzględniając ich rodzaje, częstotliwość występowania oraz rozbieżności w nazewnictwie. Zidentyfikowano różnice terminologiczne w obrębie atrybutów i dokonano ich ujednolicenia.

Poniżej przedstawiono zestawienie produktów leczniczych objętych analizą, podzielonych według grup roboczych zdefiniowanych powyżej. Dla każdego produktu wskazano liczbę zidentyfikowanych atrybutów, przypisanych numerów GTIN na podstawie stanu Rejestru Produktów Leczniczych na 30 czerwca 2025 r. oraz liczby zidentyfikowanych charakterystyk.

Różnice między atrybutami

W poniższym zestawieniu przedstawiono zidentyfikowane różnice w nazewnictwie atrybutów w formularzach ekstrakcji dla poszczególnych produktów leczniczych ujętych w grupach roboczych. W większości przypadków różnice te dotyczą wariantów zapisu, które nie wpływają na zakres ani sens merytoryczny atrybutu. Szczególnym przypadkiem jest produkt Butomidor (grupa robocza I), będący produktem weterynaryjnym. W jego formularzu występują atrybuty charakterystyczne. W grupach II i III występują drobne rozbieżności w nazewnictwie (np. warianty pisowni lub rozszerzenia opisu), które jednak nie wpływają na treść semantyczną atrybutu.

Standaryzowanie nazewnictwa atrybutów

W celu ujednolicenia struktury danych oraz poprawy identyfikowalności atrybutów w formularzach ekstrakcji, przeprowadzono standaryzację nazw atrybutów. Proces polegał na wyborze jednej, najbardziej reprezentatywnej wersji nazwy spośród wariantów występujących dla tego samego atrybutu. Jako standard przyjęto nazwę, która pojawiała się najczęściej, pod warunkiem, że zachowywała sens i logikę merytoryczną pierwotnych wersji.

Powtarzalność atrybutów w ChPL

Na tym etapie przeprowadzono analizę powtarzalności występowania atrybutów w charakterystykach produktów leczniczych (ChPL), po wcześniejszym ustandaryzowaniu ich nazw zgodnie z zakresem merytorycznym. Z analizy wykluczono produkt weterynaryjny Butomidor, którego struktura odbiega od standardów obowiązujących dla produktów leczniczych stosowanych u ludzi.

Po ujednoliceniu nazewnictwa stwierdzono, że

- atrybuty oznaczone numerami od 1 do 9 występują we wszystkich analizowanych charakterystykach produktów;
- atrybut 10 („Data zatwierdzenia lub częściowej zmiany tekstu”) nie pojawia się jedynie w formularzu dla produktu Orofar Max;
- atrybuty 11 („Dozymetria”) oraz 12 („Instrukcja przygotowania radiofarmaceutyków”) występują wyłącznie w przypadku produktów radiofarmaceutycznych i mają charakter specjalistyczny.

Struktura formatów danych

Przeanalizowano formaty danych, w jakich zapisane są poszczególne atrybuty w dokumentach ChPL. Format tekstowy jest rozumiany jako występowanie danych zapisanych wyłącznie w ogólnym formacie. Format tabelaryczny został przypisany dla danych przedstawionych w układzie typowym dla tabeli, czyli organizacji tekstu w kolumnach i wierszach. Format porządkowy został przypisany dla atrybutów, w których poza tytułem nie pojawia się żadna informacja. Format rycina został przyporządkowany dla informacji przedstawianych w formacie grafik, wykresów i rycin.

Struktura formatów danych w ChPL

Agregat	Format
1. Nazwa Produktu Leczniczego	tekst
2. Skład Jakościowy I Ilościowy	tekst tekst + tabela
3. Postać Farmaceutyczna	tekst
4. Szczegółowe Dane Kliniczne	porządkowy
4.1 Wskazania Do Stosowania	tekst
4.2 Dawkowanie I Sposób Podawania	tekst tekst + tabela tekst + tabela i rycina
4.3 Przeciwwskazania	tekst
4.4 Specjalne Ostrzeżenia I Środki Ostrożności Dotyczące Stosowania	tekst tekst + tabela
4.5 Interakcje Z Innymi Produktami Leczniczymi I Inne Rodzaje Interakcji	tekst tekst + tabela
4.6 Wpływ Na Płodność, Ciążę I Laktację	tekst tekst + tabela
4.7 Wpływ Na Zdolność Prowadzenia Pojazdów I Obsługiwanie Maszyn	tekst
4.8 Działania Niepożądane	tekst tekst + tabela
4.9 Przedawkowanie	tekst
5. Właściwości Farmakologiczne	porządkowy

Agregat	Format
5.1 Właściwości Farmakodynamiczne	tekst
	tekst + tabela
	tekst + tabela i rycina
5.2 Właściwości Farmakokinetyczne	tekst
	tekst + tabela
5.3 Przedkliniczne Dane O Bezpieczeństwie	tekst
6. Dane Farmaceutyczne	porządkowy
6.1 Wykaz Substancji Pomocniczych	tekst
6.2 Niezgodności Farmaceutyczne	tekst
6.3 Okres Ważności	tekst
6.4 Specjalne Środki Ostrożności Podczas Przechowywania	tekst
6.5 Rodzaj I Zawartość Opakowania	tekst
	tekst + tabela
	tekst + tabela i rycina
6.6 Specjalne Środki Ostrożności Dotyczące Usuwania I Przygotowania Produktu Leczniczego Do Stosowania	tekst
	tekst + tabela
	tekst + tabela i rycina
7. Podmiot Odpowiedzialny Posiadający Pozwolenie Na Dopuszczenie Do Obrotu	tekst
8. Numer Pozwolenia Na Dopuszczenie Do Obrotu	numeryczny
	tekst
9. Data Wydania Pierwszego Pozwolenia Na Dopuszczenie Do Obrotu I Data Przedłużenia Pozwolenia	data
	tekst
10. Data Zatwierdzenia Lub Częściowej Zmiany Tekstu	data
	tekst
11. Dozymetria	tekst + tabela
12. Instrukcja Przygotowania Radiofarmaceutyków	tekst
	tekst + rycina
	tekst + tabela

Źródło: Opracowanie własne

4.3. Założenia i ograniczenia biznesowe

- Źródła danych (Rejestr Produktów Leczniczych, EMA/KE) muszą być klasyfikowane na podstawie procedury rejestracyjnej produktu.
- Źródłem GTIN jest Rejestr Produktów Leczniczych,
- W ChPL brak jest informacji odnośnie do GTIN, co powoduje konieczność przypisania informacji na podstawie parametrów kontrolnych, opisanych poprzez atrybuty: moc, dawka, nazwa międzynarodowa substancji czynnej (INN),
- Jeden dokument ChPL może odnosić się do kilku produktów leczniczych identyfikowanych odrębnie poprzez numer GTIN, który jest wyznacznikiem wytworzenia ustrukturyzowanego rekordu dla konkretnego produktu leczniczego ze wszystkimi jego atrybutami (m.in. wskazania do stosowania, przeciwwskazania).
- Rozbieżności w zakresie nazwy substancji czynnej, klasy farmakoterapeutycznej w Rejestrze Produktów Leczniczych, a zawartością ChPL,
- Występuje różnorodność semantyczna nazw tych samych merytorycznie atrybutów, które wymagają standaryzacji nazewnictwa
- Format danych w ChPL w większości atrybutów jest tekstowy, jednak część atrybutów zawiera format combo (typu mieszanego), co wymaga opracowania procedury postępowania.
- Unijny rejestr jest zdecentralizowany – podział na produkty do stosowania u ludzi i produkty do stosowania u ludzi w chorobach rzadkich.
- Dane o kolejnych decyzjach w rejestrze unijnym nie precyzują, czy zmiana jest wynikiem procedury, czy jest to zmiana merytoryczna – wymagane jest utworzenie analizatora zmian do poziomu „modyfikacja brzmienia wskazania”; „nowe wskazanie”, „nowa postać leku”, „aktualizacja danych o skuteczności i/lub bezpieczeństwie”.

5. Ekstrakcja danych – opis procesu

5.1. Etapy

Proces został podzielony na cztery zasadnicze etapy:

1. Identyfikacja źródła danych – lokalizacja produktu w rejestrach oraz określenie procedury rejestracyjnej.
2. Uzyskanie dostępu do źródeł danych (ChPL).
3. Ekstrakcja danych i konwersja do ustrukturyzowanego rekordu elektronicznego.
4. Kontrola jakości – walidacja kompletności i poprawności danych po dokonanej konwersji.
5. Aktualizacja danych – bieżący monitoring i wersjonowanie w cyklu dziennym.

5.2. Identyfikacja danych

Proces ekstrakcji danych z Charakterystyk Produktów Leczniczych (ChPL) rozpoczyna się od prawidłowej identyfikacji produktu oraz ustalenia źródła dokumentu. Na tym etapie kluczowe jest powiązanie numeru GTIN z produktem leczniczym, określenie procedury rejestracyjnej oraz wybór odpowiedniej ścieżki pozyskania dokumentu ChPL.

Numer GTIN jako klucz główny

Numer GTIN (ang. Global Trade Item Number) stanowi podstawowy identyfikator w procesie digitalizacji. Jest to unikalny, międzynarodowy numer przypisany produktom leczniczym i ich wariantom (np. różnym mocom, postaciom farmaceutycznym, opakowaniom). GTIN pełni rolę klucza głównego, który umożliwia jednoznaczne powiązanie dokumentu ChPL z konkretnym produktem, pozwala rozróżniać warianty tego samego leku oraz zapewnia spójność i integralność danych między rejestrem źródłowym (URPL, KE/EMA) a bazą docelową.

Określenie procedury rejestracyjnej

Na podstawie uzyskanych informacji z RPL ustala się procedurę rejestracyjną produktu. Procedury krajowe: NAR, MRP, DCP, IR – źródłem dokumentu jest RPL prowadzony przez URPL. Procedura scentralizowana (CEN) – decyzję o dopuszczeniu do obrotu wydaje Komisja Europejska na podstawie oceny EMA. Źródłem dokumentów są rejestry KE: Community register dla produktów standardowych i Orphan register dla leków stosowanych w chorobach rzadkich.

Charakterystyka procedur rejestracyjnych w podziale na podmiot i źródło danych

Procedura	Opis	Podmiot źródłowy	Źródło ChPL
NAR	Rejestracja wyłącznie na poziomie krajowym	URPL	Rejestr Produktów Leczniczych URPL
MRP	Rejestracja w oparciu o ocenę innego państwa członkowskiego UE	URPL	Rejestr Produktów Leczniczych URPL
DCP	Jednoczesna rejestracja w kilku państwach UE z koordynacją	URPL	Rejestr Produktów Leczniczych URPL
IR	Dopuszczenie do obrotu leku sprowadzonego z innego kraju UE	URPL	Rejestr Produktów Leczniczych URPL
CEN	Rejestracja obowiązująca we wszystkich krajach UE	EMA, KE	Rejestr Komisji Europejskiej: Community register (standard), Orphan register (choroby rzadkie)

Źródło: Opracowanie własne

Wybór ścieżki do źródła ChPL

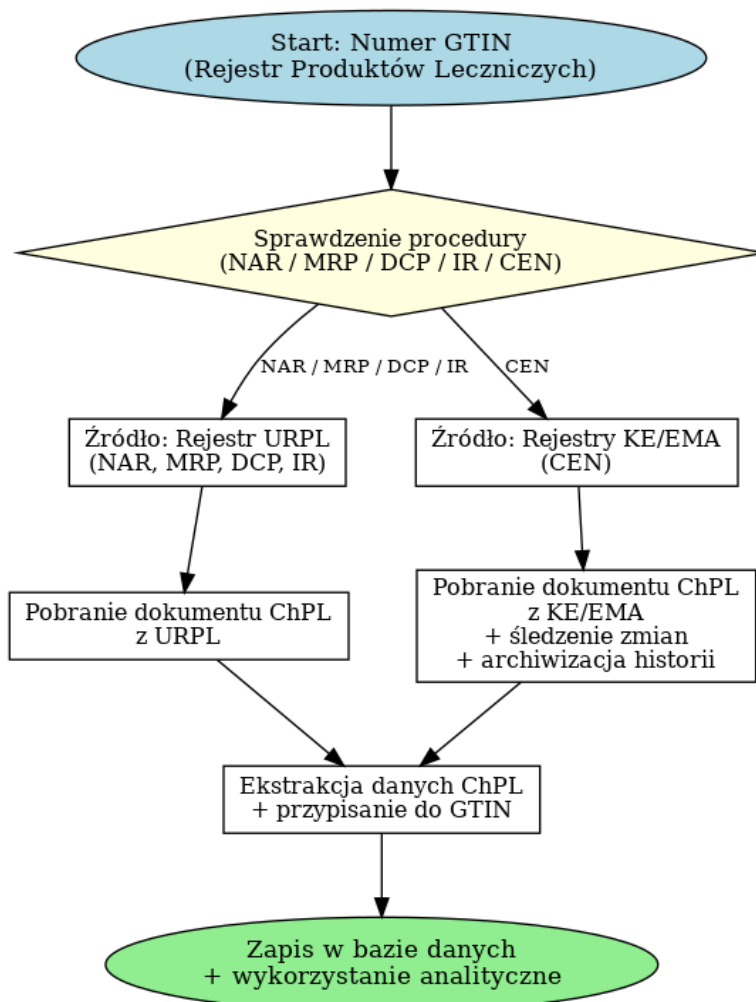
Po określeniu procedury wybierana jest odpowiednia ścieżka pozyskania dokumentu:

- Ścieżka krajowa – produkty z procedurami NAR, MRP, DCP, IR. Źródłem jest rejestr URPL. System pobiera aktualną wersję dokumentu ChPL, dokonuje ekstrakcji danych i przypisuje je do numeru GTIN.
- Ścieżka scentralizowana – produkty procedurami CEN. Źródłem są rejestry KE/EMA (Community register, Orphan register). W tym przypadku oprócz pobrania dokumentu należy zapewnić archiwizację historii zmian i oznaczenie wszystkich wersji datą decyzji.

Ścieżka krajowa obejmuje produkty dopuszczone do obrotu w procedurach krajowych: NAR, MRP, DCP, IR. We wszystkich tych przypadkach źródłem dokumentu ChPL jest Rejestr Produktów Leczniczych prowadzony przez URPL. Po identyfikacji produktu system pobiera dokument w aktualnej wersji, następnie dokonuje ekstrakcji danych i przypisuje je do numeru GTIN.

Ścieżka scentralizowana dotyczy produktów zarejestrowanych w procedurze scentralizowanej (CEN). W tym przypadku źródłem są rejestry Komisji Europejskiej i EMA, w tym Community register dla produktów standardowych oraz Orphan register dla leków na choroby rzadkie. W przypadku procedury CEN należy zidentyfikować produkt w Rejestrze KE/EMA, następnie uzyskać do właściwego dokumentu ChPL. W odróżnieniu od procedur krajowych, w przypadku procedury CEN zachodzi obowiązek śledzenia zmian i archiwizacji historii wszystkich wersji dokumentu wraz datą zmian.

Diagram decyzyjny w procesie digitalizacji danych z ChPL w zależności od procedury.



Źródło: Opracowanie własne

5.3. Przetwarzanie danych

Celem tego etapu jest dostarczenie danych w ustrukturyzowanej formie, gotowych do dalszego wykorzystania w procesach analitycznych, przy zachowaniu pełnej wartości informacyjnej zawartej w dokumentach źródłowych. Następuje etap ekstrakcji danych dokumentu ChPL do struktury bazodanowej. Proces obejmuje:

- Format danych – zapis danych we właściwym formacie, w szczególności dat i danych nienumerycznych
- Treści tekstowe – przypisanie zawartości poszczególnych sekcji ChPL do odpowiednich pól bazodanowych, z zachowaniem oryginalnej treści i układu.
- Dane tabelaryczne – konwersja tabel do struktury rekordów bazodanowych umożliwiających przetwarzanie danych i ich manipulacje.
- Ryciny i schematy – archiwizacja w oryginalnym formacie lub konwersja do standardowych formatów (np. PNG, PDF), uzupełniona o metadane i opis tekstowy.
- Powiązanie danych z GTIN – wszystkie atrybuty są przypisywane do właściwego numeru GTIN, który pełni rolę klucza głównego i zapewnia spójność między dokumentem a produktem.
- Standaryzacja nazewnictwa atrybutów.

Zakres danych do digitalizacji

Po określeniu źródła i pozyskaniu dokumentu ChPL konieczne jest zdefiniowanie zakresu danych, które podlegają digitalizacji, a następnie ich przetworzenie do formatu ustrukturyzowanego. Ten etap ma znaczenie dla zachowania kompletności informacji i ich dalszego wykorzystania w analizach

Zakres danych w rozróżnieniu na produkty i źródło informacji

Zakres danych	Standardowe produkty lecznicze	Produkty radiofarmaceutyczne	Metadane – produkty CEN
Nazwa produktu leczniczego	✓	✓	–
Skład jakościowy i ilościowy	✓	✓	–
Postać farmaceutyczna	✓	✓	–
Wskazania do stosowania	✓	✓	–
Dawkowanie i sposób podawania	✓	✓	–
Przeciwwskazania	✓	✓	–
Ostrzeżenia i środki ostrożności	✓	✓	–
Działania niepożądane	✓	✓	–
Właściwości farmakokinetyczne	✓	✓	–
Właściwości farmakodynamiczne	✓	✓	–
Dozymetria	–	✓	–
Instrukcja przygotowania produktów radiofarmaceutycznych	–	✓	–
Data aneksu decyzji	–	–	✓
Opis zmiany	–	–	✓
Numer wersji dokumentu	–	–	✓
Identyfikator dokumentu w rejestrze KE/EMA	–	–	✓

Źródło: Opracowanie własne

Zakres informacji

Dokumenty ChPL posiadają ujednoliconą strukturę sekcji wynikającą z wymogów regulacyjnych. Każda z nich zawiera informacje o innym aspekcie stosowania produktu.

Przykładowe etykiety

Kategoria	Zakres danych
Interwencja	Nazwa produktu leczniczego, skład jakościowy i ilościowy, postać farmaceutyczna
Populacja	Wskazania do stosowania, dawkowanie i sposób podawania
Bezpieczeństwo	Przeciwwskazania, ostrzeżenia i środki ostrożności, działania niepożądane
Farmakologia	Właściwości farmakokinetyczne, właściwości farmakodynamiczne
Radiofarmaceutyki	Dozymetria, instrukcja przygotowania produktów radiofarmaceutycznych
Metadane	Data aneksu decyzji, numer wersji dokumentu, identyfikator dokumentu w rejestrze KE/EMA

Źródło: Opracowanie własne

Przykładowe dane wyjściowe na podstawie danych ChPL

Zmienna	Opis	Źródło
GTIN	Numer GTIN, klucz główny	URPL, KE/EMA
PRODUCT_NAME	Nazwa produktu leczniczego	URPL, KE/EMA
COMPOSITION	Skład jakościowy i ilościowy	URPL, KE/EMA
FORM	Postać farmaceutyczna	URPL, KE/EMA
INDICATIONS	Wskazania do stosowania	URPL, KE/EMA
DOSE	Dawkowanie i sposób podawania	URPL, KE/EMA
CONTRAINDICATIONS	Przeciwwskazania	URPL, KE/EMA
WARNINGS	Ostrzeżenia i środki ostrożności	URPL, KE/EMA
ADVERSE_EFFECTS	Działania niepożądane	URPL, KE/EMA
PHARMACOKINETICS	Właściwości farmakokinetyczne	URPL, KE/EMA
PHARMACODYNAMICS	Właściwości farmakodynamiczne	URPL, KE/EMA
DOSIMETRY	Dozymetria	URPL (radiofarmaceutyki), KE/EMA (radiofarmaceutyki)
PREPARATION_INSTRUCTIONS	Instrukcja przygotowania produktów radiofarmaceutycznych	URPL (radiofarmaceutyki), KE/EMA (radiofarmaceutyki)
AMENDMENT_DATE	Data aneksu decyzji	KE/EMA (CEN)
VERSION_NUMBER	Numer wersji dokumentu	KE/EMA (CEN)
SOURCE_ID	Identyfikator dokumentu w rejestrze KE/EMA	KE/EMA (CEN)

Źródło: Opracowanie własne

Wymagania w obszarze przetwarzania danych

- Zakres digitalizacji musi obejmować wszystkie wymagane sekcje określone w zakresie digitalizacji
- W przypadku radiofarmaceutyków system musi dodatkowo obsługiwać sekcje: dozymetria i instrukcja przygotowania.
- Dla procedury CEN system musi zapisywać metadane dotyczące zmian dokumentu (data aneksu, numer wersji, identyfikator KE/EMA).
- Wszystkie dane muszą być przekształcane do formatu ustrukturyzowanego, zgodnego ze schematem bazodanowym.
- Dane tekstowe, tabelaryczne i graficzne muszą być poprawnie konwertowane i przypisywane do odpowiednich atrybutów.
- Wszystkie atrybuty muszą być jednoznacznie przypisane do numeru GTIN.
- Eksport danych musi być możliwy do formatów zewnętrznych, np. CSV, XLSX, JSON, SQL, PDF lub poprzez API.

5.4. Kontrola jakości

Po przetworzeniu informacji z dokumentów ChPL konieczne jest zapewnienie ich poprawności, spójności i zgodności ze źródłami. Dopiero dane, które przejdą pełną walidację, mogą zostać wprowadzone do systemu docelowego i stanowić podstawę dalszych analiz oraz raportów.

Walidacja

Celem jest zminimalizowanie ryzyka błędów wpływających na rzetelność analiz. Obejmuje ona: weryfikację numeru GTIN – sprawdzenie poprawności numeru i powiązanie go z właściwym produktem, kontrolę kompletności – upewnienie się, że wszystkie wymagane sekcje dokumentu zostały wyodrębnione i zapisane, porównanie ze źródłem – weryfikacja treści względem dokumentów z URPL lub KE/EMA, identyfikację błędów – wykrywanie braków, niespójności, błędnej konwersji tabel czy nieczytelnych elementów graficznych, oznaczenie wersji – przypisanie daty aneksu decyzyj i numeru wersji, co pozwala śledzić historię zmian.

Wymagania w obszarze kontroli jakości danych

System musi weryfikować poprawność numeru GTIN zgodnie ze standardem. Walidacja danych musi obejmować: kompletność, spójność, zgodność z dokumentem źródłowym oraz poprawność danych z oryginałem. Każdy dokument musi być oznaczony numerem wersji oraz datą.

5.5. Aktualizacja danych

Utrzymanie aktualności bazy danych wymaga stałego monitorowania rejestrów źródłowych (RPL, KE/EMA) oraz zapewnienia elastycznych mechanizmów eksportu i udostępniania informacji.

Aby Rozwiązanie odzwierciedlało najnowsze informacje, konieczne jest wdrożenie automatycznych mechanizmów monitorowania dokumentów ChPL. Rozwiązanie powinno: cyklicznie sprawdzać rejestry URPL oraz KE/EMA pod kątem dostępności nowych wersji dokumentów, pobierać i przypisywać aktualne dokumenty do właściwego numeru GTIN, archiwizować poprzednie wersje wraz z oznaczeniem daty aneksu decyzyj i numeru wersji, generować powiadomienia o zmianach, w szczególności dotyczących wskazań, dawkowania, przeciwwskazań czy ostrzeżeń.