

ZPPZ.510.1.2025  
(151659)

Pytania wspólne:

1. Wszystkie 4 części zapytania dotyczą rozbudowy analitycznej hurtowni danych. Jeżeli dostarczone rozwiązania mają rozbudować istniejącą hurtownię to prosimy o podesłanie podstawowych informacji – np. w jakiej technologii jest zbudowana? Z jakich źródeł się integruje? Jaki jest obecny rozmiar hurtowni? Jakie są miesięczne przyrosty?

Odpowiedź: Projekt rozbudowy analitycznej hurtowni danych obejmuje rozszerzenie zakresu projektu o nowe komponenty i grupy usług. Całość rozwiązania projektowana jest w architekturze Data Mesh, pozwalającej na korzystanie zarówno z obecnie funkcjonujących komponentów jak i nowo dodanych. W świetle powyższego dane wolumetryczne, technologiczne dotyczące obecnego rozwiązania Zamawiający uznaje za nieistotne z perspektywy procesu konsultacji rynkowych.

2. Czy poszczególne części już istnieją w ramach obecnej hurtowni? Np. zbudowanie platformy ładowania i przetwarzania danych (ETL) oznacza budowę ETL od zera? Przepisanie obecnego rozwiązania ETL (procesów)? Rozbudowę obecnego rozwiązania?

Odpowiedź: Projekt rozbudowy analitycznej hurtowni danych obejmuje rozszerzenie zakresu projektu o nowe komponenty i grupy usług w tym komponenty i usługi ETL, uzupełniające istniejące funkcjonalności. Zamawiający nie planuje rozbudowy istniejących komponentów ETL a równoległe uruchomienie dodatkowych komponentów. W świetle powyższego dane wolumetryczne, technologiczne dotyczące obecnego komponentu ETL Zamawiający uznaje za nieistotne z perspektywy procesu konsultacji rynkowych.

3. Załącznik 4. W załączniku pojawiają się pytania o integrację z systemami Zamawiającego (np. część ETL: „W jaki sposób planowana i przeprowadzana jest integracja systemów Zamawiającego z dostarczonym rozwiązaniem?”). Czy w ramach prezentacji oczekiwane jest zaprezentowanie ogólnych możliwości integracji? Odpowiedź dotycząca konkretnie integracji z systemami Zamawiającego – wymaga informacji jakie to są systemy, kto będzie odpowiedzialny za wystawienie danych, w jakim formacie dane będą dostępne, czy dane będą dostarczane przyrostowo, czy są to tylko ustrukturyzowane dane, itd.

Odpowiedź: W ramach prezentacji należy przedstawić ogólne możliwości integracyjne narzędzia, dane dotyczące możliwości skalowania pionowego poziomego (z wyróżnieniem propozycji dla budowy rozwiązania „semi-online” dla systemu krytycznego), ogólne parametry wydajnościowe.

4. Czy do rozwiązania (np. warstwa ODS) trafiać będą tylko dane ustrukturyzowane? Ustrukturyzowane + nieustrukturyzowane (w jakich proporcjach)?

Odpowiedź: W Operational Data Store (ODS) przechowywane będą dane ustrukturyzowane, ale w formie zbliżonej do danych operacyjnych (transakcyjnych) w systemach dziedzinowych:

- Struktura danych zbliżona będzie do struktury danych transakcyjnych

- Brak wymiaru czasu
  - Wysoka aktualność danych - zakładane odświeżanie online lub w trybie zbliżonym do online
  - Struktura danych zoptymalizowana jest pod szybki dostęp do bieżących danych operacyjnych
5. Czy wymagana jest anonimizacja danych w ramach rozwiązania?

Odpowiedź: Za anonimizację danych odpowiadać będą usługi po stronie CeZ

6. Czy anonimizacja będzie dotyczyć tylko danych strukturalnych?

Odpowiedź: patrz Część 1, pytanie 5

7. Załącznik 4. Dla części 1. Harmonogram powinien wynikać z zakresu, np. liczba modeli do dostarczenia, wymagania dotyczące modelu, zakres. Na tym etapie brak informacji dotyczących zakresu. Czy chodzi o jakiś przykładowy harmonogram?

Odpowiedź: Tak

8. Odnośnie zakresu prezentacji dla części 1 – prosimy o wyjaśnienie oczekiwania Zamawiającego co do przedstawienia Harmonogramu realizacji pkt 3 podpunkt 3) oraz jaka jest różnica pomiędzy harmonogramem, który należy przedstawić w pkt 4 podpunkt 1)

Odpowiedź: Przytoczone harmonogramy dotyczą tego samego. Różnica jest w szczegółowości prezentowanego materiału, w pkt4.1 oczekujemy więcej szczegółów naniesionych na diagram Gantta (np. Zależności, jeśli istnieją).

9. Prosimy o informację czy akapit zaczynający się od „Proszę w punktach określić i krótko opisać poniżej podane funkcjonalności rozwiązania” należy rozumieć jako osobny podpunkt – czy przypadkowo pominięto go jako osobny punkt? (dotyczy wszystkich części postępowania)

Odpowiedź: Powyższy akapit jest rozwinięciem tego punktu głównego, informującym co należy uwzględnić.

10. Jak należy rozumieć sformułowanie „poziom dostępności rozwiązania na rynku”? Chodzi o liczbę wdrożeń danego rozwiązania w skali kraju/świata?

Odpowiedź: Intencją Zamawiającego jest poznanie wolumetrii wdrożeń rozwiązania na obszarze UE, dostępności partnerów wdrożeniowych, integratorów.

11. Proszę wskazać rodzaje oraz typy i interfejsy (SQL/ODBC/JDBC/Logi/pliki/API/FHIR/CDA/HL7/etc.) źródeł danych oraz ich ilości na sztuki per typ/rodzaj/interfejs dla wszystkich części (1,2,3,4) systemu.

Odpowiedź: Zamawiający na tym etapie konsultacji nie wskazuje ani nie ogranicza interfejsów i technologii możliwych do wykorzystania w projekcie HDC.

12. Proszę wskazać czy w jakiej formie lub interfejsie oczekiwane jest udostępnianie danych oraz wyników poszczególnych części systemu (GUI/SQL/ODBC/JDBC/pliki/API/FHIR/CDA/HL7/etc.) oraz dla jakich populacji i obciążeń.

Odpowiedź: Zamawiający na tym etapie konsultacji nie wskazuje ani nie ogranicza interfejsów i technologii możliwych do wykorzystania w projekcie HDC.

13. Proszę wskazać główne oczekiwania kluczowych użytkowników poszczególnych części systemu.

Odpowiedź: Główne oczekiwania kluczowych użytkowników to:

- a. stworzenie repozytorium dla kanonicznego modelu danych eZdrowia a w szczególności dla systemów CeZ z propagacją modelu do poszczególnych systemów
  - b. dostarczenie platformy narzędziowej umożliwiającej realizację zadań raportowania operacyjnego (zintegrowanego z systemami) jak zadań raportowania użytkownika oraz analityki
  - c. odciążenie systemów dziedziny (transakcyjnych) od zadań przetwarzania wtórnego (patrz podpunkt b) z zapewnieniem bezpieczeństwa i wydajności zadań przetwarzania
  - d. wprowadzenie platformy narzędziowej umożliwiającej tworzenie, uczenie, rozwój i uruchamianie własnych algorytmów AI
14. Proszę podać listę kluczowych produktów całego projektu lub chociaż wstępnie zakres wdrożenia pozwalający na oszacowanie rozmiarów usług po zainstalowaniu komponentów składowych, aby lepiej określić możliwości ich harmonogramowania.

Odpowiedź: Intencją Zamawiającego jest określenie katalogu dostępnych technologii, rozwiązań i możliwości ich implementacji i dostarczenia w założonym harmonogramie realizacji projektu.

#### Część I:

Dostawa sprzętu lub wdrożenie środowisk i narzędzi dla platformy uczenia maszynowego i sztucznej inteligencji (AI/ML) uruchamianych on-premise:

1. Jakiego rodzaju zastosowania przewidują Państwo dla platformy uczenia maszynowego?

Odpowiedź: Platforma przeznaczona będzie zarówno do wypracowywania i trenowania własnych, dziedziny modeli / algorytmów jak i do inferencji z wytrenowanymi oraz gotowymi modelami / algorytmami. Algorytmy i usługi mają umożliwić właściwą prezentację danych i informacji, wspierać podejmowanie decyzji, umożliwiać przekazywanie właściwych informacji użytkownikom, a także wspierać komunikację z nimi.

2. Czy platforma ma służyć do trenowania modeli, czy wyłącznie do inferencji?

Odpowiedź: ma służyć do obydwu celów

3. Jeżeli platforma ma służyć do trenowania modeli, kto będzie odpowiedzialny za ich trenowanie – dostawca czy zespół Data Science po stronie Klienta?

Odpowiedź: Na etapie konsultacji rynkowych określone są wymagania funkcjonalne, niefunkcjonalne oraz stosy technologiczne projektowanych rozwiązań bez wskazywania na odpowiedzialności i usługi dodatkowe związane z wytwarzaniem i utrzymaniem rozwiązań, jednak już dziś można stwierdzić, że przewidujemy możliwość trenowania modeli przez zespół CeZ

4. Ile będzie modeli? Jakiego rodzaju?

Odpowiedź: Zamawiający na etapie doboru technologii nie określa docelowych parametrów wolumetrycznych rozwiązania ani też rodzajów modeli. Platforma ma być podatna na rozszerzenia i umożliwiać uruchamianie zarówno obecnie dostępnych modeli jak i tych, które pojawią się w zakładanym okresie trwałości projektu.

5. Z jakimi regulacjami formalno-prawnymi muszą być kompatybilne modele na platformie?

Odpowiedź: platforma musi spełniać wszystkie aktualne regulacje formalno-prawne w zakresie sztucznej inteligencji oraz przetwarzania danych dotyczące tego typu platform obowiązujące w Polsce i Unii Europejskiej

6. Czy dane przetwarzane przez modele powinny być utrwalane? Jeżeli tak:

Odpowiedź: Zamawiający wymaga, aby modele, parametry, zbiory testowe mogłyby być przechowywane w pamięci trwałej, posiadały możliwość tworzenia kopii zapasowych oraz wersjonowania.

- a. Jaki będzie wolumen tych danych?

Odpowiedź: Zamawiający na tym etapie konsultacji nie jest w stanie określić wolumenu danych.

- b. Jaki jest szacowany średni przyrost dzienny/miesięczny?

Odpowiedź: Zamawiający na tym etapie konsultacji nie jest w stanie określić przyrostu wolumenu danych.

- c. Jaki jest okres retencji danych?

Odpowiedź: na tym etapie brak takich danych, jednak musi być możliwość stosowania różnych polityk zw. z retencją danych

- d. Jakie są wymagania do anonimizacji/pseudonimizacji danych?

Odpowiedź: na tym etapie zakłada się, że dane będą anonimizowane z wykorzystaniem planowanego komponentu zewnętrznego w stosunku do platformy ML/AI jeszcze przed przekazaniem ich do platformy AI/ML

7. Czy dane przetwarzane przez modele będą służyły do douczania modeli?

Odpowiedź: Może wystąpić taka potrzeba.

8. Jakiego rodzaju dane będą przetwarzane przez modele?

Odpowiedź: W przypadku algorytmów własnych modele będą przetwarzać uporządkowane dane o średniej krotności, pochodzące z różnych systemów dziedzinowych i występujące w różnych formatach. W przypadku algorytmów ogólnych przetwarzane będą również dane graficzne oraz dane głosowe (przekształcanie mowy na tekst). Nie zakładamy wykorzystania funkcjonalności przetwarzania medycznych danych obrazowych, jednak nie wykluczamy takiej możliwości.

9. Jakiego rodzaju inferencje są przewidziane (realtime/batch)

Odpowiedź: Realtime i batch.

10. Ile inferencji powinna obsłużyć platforma średnio/maksymalnie na godzinę.

Odpowiedź: Zamawiający na tym etapie konsultacji nie określa wolumenu inferencji na godzinę.

11. Jakie konkretne problemy i wyzwania w przetwarzaniu danych medycznych ma rozwiązać wdrożenie platformy AI/ML?

Odpowiedź:

- a. Nie ma założeń ograniczenia działania platformy wyłącznie do przetwarzania danych medycznych – platforma ma wspierać wytypowane zadania w ramach eZdrowia i wsparcia planów jego rozwoju.
  - b. Cele merytoryczne nie dotyczą omawianych kwestii technologicznych
  - c. W zakresie obszaru przetwarzania danych medycznych zakładamy przetworzenie danych związanych z realizacją danej aktywności biznesowej by wsparcie pracy użytkownika korzystającego z tej usługi było jak najbardziej optymalne. Przykładowo działanie takie może być wprowadzone w zakresie kwalifikacji biorcy na donacje czy wsparcie pracy lekarza POZ w zakresie potencjalnego wystąpienia choroby rzadkiej u pacjenta. W innych obszarach mogą to być analizy zachowania się ekosystemu eZdrowia w części finansowanej poprzez wprowadzenia kierunkowych kompletnych terapii leczenia.
12. Czy kluczowym celem jest wsparcie podejmowania decyzji medycznych, optymalizacja operacji szpitalnych, predykcja zdarzeń medycznych czy inny zakres funkcjonalny? Czy wymagane jest wsparcie dla modelowania predykcyjnego, analizy obrazów lub NLP?
- Odpowiedź: Celem platformy jest wsparcie w podejmowaniu decyzji w ramach różnych systemów i usług (interoperacyjność) również w zakresie decyzji medycznych/klinicznych, a także wsparcie analityki predykcyjnej i preskryptywnej. Obecnie nie zakładamy wykorzystywania funkcjonalności przetwarzania medycznych danych obrazowych, jednak nie wykluczamy takich funkcjonalności.
13. Jakie źródła danych będą zasilaty platformę AI/ML (systemy szpitalne HIS, LIMS, EHR, bazy kliniczne, urządzenia IoT, dane z obrazowania medycznego)?
- Odpowiedź: Dane do platformy AI/ML będą dostarczane z ODS, jak i potencjalnie z zewnętrznych źródeł. Na obecnym etapie nie zakładamy zasilania danymi z urządzeń IoT ani obrazowania medycznego.
14. Czy platforma AI/ML ma przetwarzać dane w czasie rzeczywistym, czy w modelu wsadowym?
- Odpowiedź: Zakładamy oba
- a. budowanie modeli AI/ML – model wsadowy
  - b. implementacja modeli AI/ML i ich wykorzystanie – model rzeczywisty / model wsadowy
15. Jakie są oczekiwane wolumeny danych do przetworzenia? Czy dane będą przechowywane lokalnie, w hurtowni danych czy w systemie plików?

Odpowiedź: Rząd >10 TB. Zamawiający przewiduje zastosowanie odpowiednich technologii przechowywania danych dla wszystkich komponentów stanowych HDC, to znaczy utrwalających swój stan, dane w pamięci nieulotnej.

Wybór odpowiednich technologii, sposobów organizacji danych musi być przeprowadzony zarówno zgodnie z najlepszą wiedzą Wykonawcy jak stosując powszechnie obowiązujące standardy rynkowe IT.

16. Czy preferowana jest jedynie architektura on-premise, czy rozważana jest możliwość wdrożenia hybrydowego z wykorzystaniem chmury?

Odpowiedź: Proszę przedstawić obydwa warianty, choć na obecnym etapie nie zakładamy rozwiązań chmurowych.

17. Jakie są aktualne zasoby infrastrukturalne, które mogą być wykorzystane?

Odpowiedź: Zasoby x86, brak zasobów GPU

18. Czy platforma powinna posiadać otwarte interfejsy API umożliwiające integrację z istniejącymi systemami raportowania i analityki danych?

Odpowiedź: Tak.

19. Jakie są wymagania dotyczące interoperacyjności z innymi narzędziami analitycznymi i BI (np. Tableau, Power BI, Qlik Sense)?

Odpowiedź: Możliwość pełnego wykorzystania.

20. Jakie formaty wymiany danych powinny być obsługiwane (FHIR, HL7, JSON, XML, CSV, DICOM w przypadku obrazowania medycznego)?

Odpowiedź: na obecnym etapie nie przewidujemy analizy danych obrazowych.

## Część II:

### Zbudowanie platformy raportowania operacyjnego (ODS) i infrastruktury

1. Jaką rolę ma pełnić ODS w Hubie Danych CeZ?

Odpowiedź: repozytorium danych transakcyjnych odpersonalizowanych i świadczącym usługę przetwarzania danych na rzecz innych komponentów/systemów a w szczególności: dalszego „post-processingu” danych w celu aktualizacji modeli analitycznych, przygotowywania zestawów danych do uczenia maszynowego, dostarczania danych (zmienne) na potrzeby działania algorytmu AI, przetwarzania danych na potrzeby raportowania operacyjnego a wykorzystującego ODS jako zasób i usługi „bazodanowe”.

2. Jakiego typu dane mają być przechowywane w ODS (structured/unstructured/semi-structured)

Odpowiedź: Dopuszcza się wszystkie 3 warianty.

3. Ilu użytkowników/odbiorców będzie korzystać z ODS?

Odpowiedź: Liczba użytkowników ODS będzie ograniczona jedynie dostępnymi zasobami i możliwościami przetwarzania danych. Nie przewiduje się ograniczeń dla poszczególnych nazwanych użytkowników.

4. Jakie są planowane metody raportowania?

Odpowiedź: Raportowanie osadzone w aplikacji użytkownika, możliwość przekierowania użytkownika do systemu raportowego, korzystanie z platformy raportowej, raporty powinny móc być rozsyłane mailem lub odkładane w katalogu. Raporty predefiniowalne, definiowalne jak i własne. Możliwość/narzucenie wykorzystania szablonów. Projekcje typu „lista”, tabela



przystępna. Raporty interaktywne oraz z podziałem na strony. Raporty powinny móc być parametryzowane. Raporty powinny mieć możliwość ograniczenia dostępu na poziomie raportu. Raporty powinny mieć możliwość ograniczenia dostępu na poziomie danych (np.: powinien być realizowalny scenariusz, w którym jednostka organizacyjna widzi tylko te rekordy lub agregacje, które jej dotyczą i nie widzi danych żadnych innych jednostek organizacyjnych). Ważna jest możliwość zabezpieczonego wystawiania niektórych raportów poza CEZ i ograniczenie dostępu do nich wybranym innym instytucjom. Dodatkowe wizualizacje z wsparciem wykresów. Filtracja. Funkcje grupujące. Raporty powinny śledzić własne użycie, tzn. powinna być odkładana informacja kto, kiedy i co oglądał. To jest informacja potrzebna do śledzenia czy raport jest używany, jak często, przez kogo i czego użytkownicy w nim najczęściej poszukują. Dane o użyciu raportów powinna móc być źródłem do wyraportowania metryk użycia wskazanej osobie. Wybrane raporty powinny posiadać możliwość przekazania prostej informacji zwrotnej.

5. Jaki jest szacunkowy wolumen danych przechowywanych w ODS?

Odpowiedź: Więcej niż 100TB

6. Jaki jest szacunkowy przyrost danych przechowywanych w ODS?

Odpowiedź: Nie posiadamy w tym momencie takich szacunków.

7. Czy istnieją jakieś wymagania odnośnie retencji danych?

Odpowiedź: Musi być możliwość tworzenia różnych polityk zw. z retencją różnych zbiorów danych.

8. Czy w ODS będą przechowywane dane objęte regulacjami (osobowe, medyczne itp.)?

Odpowiedź: Tak, ale założeniem jest ich wstępna depersonalizacja przed umieszczeniem w ODS.

9. Czy będzie wymagana anonimizacja lub pseudonimizacja danych?

Odpowiedź: Będzie realizowana przez odrębny komponent CeZ, niezależny od ODS.

10. Z jakimi dostawcami i konsumentami danych będzie wymagana integracja? Jakiego rodzaju interfejsy wymiany danych mają być wspierane?

Odpowiedź: Na etapie konsultacji oczekujemy informacji o potencjalnych możliwościach integracji.

11. Tryb zasilania ODS (batch, real-time, near real-time, event driven)?

Odpowiedź: Wszystkie wymienione w pytaniu. Zależnie od wymagań dla systemu źródłowego.

12. Czy model danych w ODS powinien być zgodny z modelami komercyjnymi lub innymi standardami branżowymi?

Odpowiedź: W części będzie pokrycie dla konkretnych archetypów przemysłowych.

13. Czy w ODS ma być zawarta logika raportowa, czy też będzie ona implementowana w narzędziach raportowych korzystających z ODS?

Odpowiedź: Będzie ona implementowana głównie po stronie ODS, ponieważ wymaga przetwarzania danych wysoko wolumenowych. Może się zdarzyć, że na ostatnim etapie, w narzędziu raportowym zaszyta zostanie logika operująca na danych nisko wolumenowych. Jeżeli sugerują Państwo przetwarzanie logiki biznesowej na wysokim wolumenie po stronie narzędzia raportowego, proszę o taką informację w prezentacji.

14. Czy rozważają Państwo integrację z technologiami chmurowymi? Jeśli tak, to czy planowane jest przejście w pełni do chmury, czy wdrożenie modelu hybrydowego? Jakie są główne powody rozważania tej integracji?

Odpowiedź: Na tym etapie nie jest to przewidywane ze względu bezpieczeństwa przetwarzania danych.

15. Czy w kontekście wdrożenia technologii chmurowych analizowane są wyłącznie rozwiązania OCHK (Chmura Krajowa), czy również inne platformy, takie jak AWS, Azure czy Google Cloud? Jakie kryteria będą decydować o wyborze dostawcy?

Odpowiedź: Rozwiązanie chmurowe nie jest brane pod uwagę na tym etapie.

16. Jakie systemy wymagają integracji? Z jakich źródeł pochodzą dane, gdzie są obecnie przechowywane i w jakim formacie (np. pliki płaskie, API, bazy danych relacyjne, NoSQL)?

Odpowiedź: Docelowo należy założyć, że integrowane będą wszystkie systemy dziedziczne CeZ. Źródła mogą być dowolne m.in. wymienione w pytaniu.

17. Z jaką częstotliwością dane powinny być przesyłane do różnych systemów? Czy wymagane jest przetwarzanie w trybie rzeczywistym, czy w określonych interwałach (np. co godzinę, dziennie, tygodniowo)?

Odpowiedź: Musi to być definiowalne. W części będzie wymagalne przetwarzanie real-time.

18. Jakie są szacunkowe wolumeny danych, które będą przetwarzane? Czy istnieją prognozy dotyczące wzrostu ilości danych w kolejnych latach?

Odpowiedź: Na obecnym etapie brak takich danych.

19. Czy organizacja korzysta już z narzędzi do wizualizacji danych? Jeśli tak, to jakie (np. Power BI, Tableau, Qlik Sense)? Czy istnieją określone standardy raportowania i wizualizacji?

Odpowiedź: Tak, korzysta z rozwiązania Power BI. W części rozwiązań istnieją zdefiniowane standardy wizualizacyjne i raportowe, Brak jednak zdefiniowanych ogólnych, wspólnych standardów – w ramach realizacji projektu są one planowane do wypracowania.

20. Czy w organizacji dostępne są narzędzia do kontroli wersji, takie jak Git lub SVN? Jakie są obecne standardy zarządzania kodem i wersjonowania w projektach związanych z danymi?

Odpowiedź: Tak, CeZ wytwarza rozwiązania zgodnie z stosowanymi powszechnie praktykami i zasadami. W ramach zarządzania modelem danych planowane jest wdrożenie Data Governance.

21. Jakie środowiska są obecnie wykorzystywane w organizacji? Czy istnieją dedykowane środowiska dla deweloperów (Dev), testowe (Preprod, Sandbox) oraz produkcyjne (Prod)?

Odpowiedź: W organizacji powszechnie stosowane są środowiska developerskie, testowe oraz produkcyjne. W wybranych projektach są też dodatkowe środowiska

22. Jakie środowiska powinny zostać wdrożone w ramach projektu? Czy potrzeba tylko środowisk testowych i produkcyjnych, czy także np. staging, UAT (User Acceptance Testing) lub dedykowanych środowisk dla integracji?



Odpowiedź: Oczekujemy rekomendacji w zakresie środowisk.

23. Czy konieczne jest zaprojektowanie i wdrożenie nowego procesu CI/CD? Czy w organizacji istnieją już rozwiązania do automatyzacji wdrożeń, które należy uwzględnić, a jeśli tak, to jakie narzędzia są stosowane (np. Jenkins, GitLab CI/CD, Azure DevOps)?

Odpowiedź: Rekomenduje się zaproponowanie optymalnego procesu CI/CD umożliwiającego maksymalizację realizacji parametru dotyczącego dostępności i ciągłości.

24. Część 2/ODS – proszę wskazać listę konkretnych systemów zewnętrznych do integracji z ODS.

- Odpowiedź: projekt nie jest realizowany pod kątem zrealizowania zamkniętej liczby konkretnych integracji, a wprowadzenie klasy rozwiązań umożliwiających ich stopniowe rozszerzanie. Inicjalnie przewiduje się rozbudowę / stworzenie integracji w kooperacji z systemami/dostępnymi interfejsami: PZH, GUS, NFZ, ZUS.

25. Część 2/ODS – jaki sumarycznie wolumen danych w OSD do raportowania przewiduje się docelowo.

- Odpowiedź: Więcej niż 1 PB.

26. Część 2/ODS – jaką całkowitą populację użytkowników raportów i kokpitów Państwo przewidują w sumie oraz czy każdy ma mieć możliwość ich tworzenia lub modyfikacji wg. własnych preferencji.

- Odpowiedź: w zakresie raportów operacyjnych przewidujemy uśredniony wolumen do 10000 (prosimy nie wiązać tego z wolumenem użytkowników korzystających z rozwiązania w perspektywie: minuta, godzina, dzień) aktywnych użytkowników (żądanie przetworzenia i wygenerowania raportu). W zakresie „kokpitów” wolumen ten będzie ograniczony wyłącznie do osób realizujących kierunkowe zadania przetwarzania danych (nie przewiduje się więcej niż 500 sumarycznych użytkowników).

27. Część 2/ODS – jakie są główne cele użytkowe i biznesowe stawiane przez Państwa dla warstwy ODS.

- Odpowiedź: Zostało to opisane w punkcie 1 w niniejszej części.

28. Proszę podać główne role biznesowe i przykłady użycia warstwy raportowej w ODS – co kto w jakim celu i charakterze będzie oczekiwał i potrzebował najczęściej  
Odpowiedź: bezpośredni użytkownik aplikacji albo interesariusz (zarówno bezpośrednio korzystający z dostępnych zasobów przetwarzania danych jak i pośrednio – zlecający ich opracowania) dla celów przetwarzania zgodnych z obowiązującą legislacją dla konkretnych typów danych. Nie określimy w tym momencie, który dokładnie strumień biznesowy będzie najczęściej wykorzystywać platformę.

29. Część 2/ODS – ilu średnio użytkowników współbieżnych będzie raportowało z ODS na minutę – wystarczą przedziały/rząd wielkości, np. 10, 100, 1 000, 10 000, 100 000, 1 000 000, itd.  
Odpowiedź: docelowo 10 000.

30. Część 2/ODS – czy rozwiązanie raportowe nad OSD ma umożliwiać również zastosowanie metod statystycznych lub klasycznej analityki predykcyjnej np. w postaci Python/JupyterNotebook, a jeśli tak to czy dla całej populacji użytkowników lub też dla jakiego ich procenta  
Odpowiedź. Przewiduje się klasyczne typy przetwarzania analitycznego: analityka opisowa,

analityka predykcyjna i analityka preskryptywna. Nie mamy KPI wskazującego na konkretny podział %.

31. Część 2/ODS – czy ODS ma być zorganizowany tradycyjnie w postaci struktur relacyjnych zbliżonych do formatów źródłowych czy też oczekiwane są również jakieś formy przetworzonego, wielowymiarowego ujęcia danych – np. ROLAP czy też kostki OLAP już na tym poziomie systemu? Odpowiedź: modele wtórne (np.: kostki) budowane w modelu kaskadowym. Wstępne założenie – ODS ma być źródłem do ich wytworzenia z możliwości spięcia się z innymi posiadanymi rozwiązaniami typu HD w trybie Mesh.
  32. Część 2/ODS – czy rozwiązanie raportowe ma również zapewniać jakieś mechanizmy monitorowania i alertowania użytkowników biznesowych o przekroczeniach oczekiwanych wartości danych z informacji w ODS? Odpowiedź: jeżeli prezentowane rozwiązanie ma taką cechę – prosimy o prezentację.
  33. Część 2/ODS – jak aktualne mają być informacje w ODS w stosunku do zmian systemów źródłowych – oczekiwane opóźnienia w danych – np. 1 tydzień, 1 dzień, 1 godzina, 1 minuta...? Odpowiedź: zależne od strumienia i zakresu stosowania danych w ODS. W wyróżnionych zakresach opóźnienie ograniczone będzie do pojedynczej transakcji w systemie źródłowym
  34. Część 2/ODS – czy mają Państwo preferencje co do technicznego sposobu realizacji warstwy danych ODS – np. RDBMS vs DWH vs Datalakehouse vs Appliances, etc?
- Odpowiedź: jesteśmy otwarci na przedstawienie koncepcji.

### Część III

Zbudowanie platformy ładowania i przetwarzania danych (ETL) i infrastruktury.

1. Jakie typy źródeł danych musi wspierać platforma?

Odpowiedź: Bazy danych (relacyjne, nie relacyjne), zasoby plikowe, API, pliki wstępnie przetworzone.

2. Ile systemów dziedzicznych ma korzystać z platformy?

Odpowiedź: docelowo praktycznie wszystkie systemy dziedziczne (kilkadziesiąt).

3. Jakie sposoby/protokoły komunikacji wspierają integrowane systemy?

Odpowiedź: Klasyczne (dostępowe bazodanowe, webserwisowe, kolejkowe).

4. Jakie mają być wspierane tryby przetwarzania danych (batch, real-time, near real-time, event driven)?

Odpowiedź: Wszystkie, możliwość dostosowania trybu do rzeczywistych potrzeb.

5. Jakiego typu dane mają być przetwarzane (structured/unstructured/semi-structured)?

Odpowiedź: Wszystkie podane. Zależnie od wymagań dla systemu źródłowego.

6. Jaki jest planowany wolumen danych przetwarzanych dziennie przez platformę?

Odpowiedź: Brak danych na tym etapie. W perspektywie paru lat po wdrożeniu należy oczekiwać wolumen rządu do 10GB surowych (bez indeksów) danych dziennie.

7. Czy rozważają Państwo integrację z technologiami chmurowymi? Jeśli tak, to czy planowane jest przejście w pełni do chmury, czy wdrożenie modelu hybrydowego? Jakie są główne powody rozważania tej integracji?  
Odpowiedź: Nie.
8. Czy w kontekście wdrożenia technologii chmurowych analizowane są wyłącznie rozwiązania OCHK (Chmura Krajowa), czy również inne platformy, takie jak AWS, Azure czy Google Cloud? Jakie kryteria będą decydować o wyborze dostawcy?  
Odpowiedź: Nie dotyczy. Nie zakładamy wykorzystania chmury na tym etapie.
9. Na jakich technologiach i architekturze oparta jest obecna hurtownia danych? Jakie konkretne bazy danych, frameworki ETL i narzędzia do zarządzania danymi są obecnie wykorzystywane?  
Odpowiedź: Klasycznej architekturze hurtowni: ETL z StageArea, Warstwa danych detalicznych, Warstwa danych analitycznych, dedykowane komponenty do przetwarzania danych na potrzeby systemów, narzędzia raportowo-analityczne.
10. Jakie systemy źródłowe są już zintegrowane z hurtownią danych? Czy są to systemy operacyjne, CRM, ERP, systemy legacy czy inne źródła?  
Odpowiedź: Dla planowanego rozwiązania integracje będą realizowane od podstaw w celu wyeliminowania problemów.
11. Jakie dodatkowe systemy wymagają integracji? Z jakich źródeł pochodzą dane, gdzie są obecnie przechowywane i w jakim formacie (np. pliki płaskie, API, bazy danych relacyjne, NoSQL)?  
Odpowiedź: Proszę założyć wszelkie domyślne formaty plików płaskich, bazy relacyjne i nie relacyjne, API (REST, SOAP), zasoby webowe przefiltrowane.
12. Z jaką częstotliwością dane powinny być przesyłane do różnych systemów? Czy wymagane jest przetwarzanie w trybie rzeczywistym, czy w określonych interwałach (np. co godzinę, dziennie, tygodniowo)?  
Odpowiedź: Odpowiedzi udzielono w pkt. 17 część 2.
13. Jakie są szacunkowe wolumeny danych, które będą przetwarzane? Czy istnieją prognozy dotyczące wzrostu ilości danych w kolejnych latach?  
Odpowiedź: Odpowiedzi udzielono w pkt. 18 część 2.
14. Czy organizacja korzysta już z narzędzi do wizualizacji danych? Jeśli tak, to jakie (np. Power BI, Tableau, Qlik Sense)? Czy istnieją określone standardy raportowania i wizualizacji?  
Odpowiedź: Odpowiedzi udzielono w pkt. 19 część 2.
15. Czy w organizacji dostępne są narzędzia do kontroli wersji, takie jak Git lub SVN? Jakie są obecne standardy zarządzania kodem i wersjonowania w projektach związanych z danymi?  
Odpowiedź: Odpowiedzi udzielono w pkt. 20 część 2.
16. Jakie środowiska są obecnie wykorzystywane w organizacji? Czy istnieją dedykowane środowiska dla deweloperów (Dev), testowe (Preprod, Sandbox) oraz produkcyjne (Prod)?  
Odpowiedź: Odpowiedzi udzielono w pkt. 21 część 2.

17. Jakie środowiska powinny zostać wdrożone w ramach projektu? Czy potrzeba tylko środowisk testowych i produkcyjnych, czy także np. staging, UAT (User Acceptance Testing) lub dedykowanych środowisk dla integracji?

Odpowiedź: Odpowiedzi udzielono w pkt. 22 część 2.

18. Czy konieczne jest zaprojektowanie i wdrożenie nowego procesu CI/CD? Czy w organizacji istnieją już rozwiązania do automatyzacji wdrożeń, które należy uwzględnić, a jeśli tak, to jakie narzędzia są stosowane (np. Jenkins, GitLab CI/CD, Azure DevOps)?

Odpowiedź: Odpowiedzi udzielono w pkt. 23 część 2.

19. Część 3/ETL Prosimy o informację jaki rodzaj danych będzie źródłem:

- Bazy danych i tabele
- Pliki ekstraktów
- Json/ xml / webservises

Odpowiedź: wszystkie wskazane, dodatkowo może wystąpić przetwarzanie plików użytkownika (csv, xlsx, txt), jak i konkretnie zorganizowanych stron webowych (być może poprzedzone procesem automatyzacji pobrania i uporządkowania danych wejściowych).

20. Część 3/ETL – Jaka będzie szacunkowa liczba obiektów do przeladowania per typ.

Odpowiedź: nie rozumiemy pytania.

21. Część 3/ETL – Jakie są przewidywane wymagania wydajnościowe, jaki wolumen do załadowania w jakim czasie?

Odpowiedź: Na etapie konsultacji nie precyzujemy wymagań.

#### Część IV

Zbudowanie platformy Zarządzania Danymi (Data Governance) w całym cyklu życia danych.

1. Co rozumiemy przez Data Governance?

Odpowiedź: Zarządzanie Danymi (Data Governance) to systematyczne podejście do zarządzania danymi w organizacji. Obejmuje to tworzenie i wdrażanie polityk, procesów, procedur oraz standardów dotyczących zbierania, przechowywania, przetwarzania i udostępniania danych. Głównym celem Data Governance jest zapewnienie wysokiej jakości danych, ich bezpieczeństwa, spójności, a także efektywnego i etycznego wykorzystania danych w celach strategicznych i operacyjnych firmy.

2. Jak platforma Zarządzania Danymi ma się do pozostałych modułów Huba Danych CeZ, Repozytorium Kanonicznego Modelu Danych i zarządzanie modelem danych systemów CeZ

Odpowiedź: W ramach Data Governance zostaną opracowane m.in. polityki, procesy, procedury i standardy dotyczące identyfikacji i monitoringu jakości danych oraz zbierania, przechowywania, przetwarzania i udostępniania danych. Jednym z efektów Data Governance jest Kanoniczny Model Danych i zarządzanie tym modelem. Głównym celem Data Governance jest zapewnienie ładu danych, wysokiej jakości danych, ich bezpieczeństwa i spójności w pozostałych modułach Huba Danych CeZ, a także w modelach danych systemów CeZ.

3. Jakie systemy powinny być objęte działaniem platformy Data Governance?

Odpowiedź: Docelowo planujemy objąć wszystkie systemy CeZ. Jednak ze względu na złożoność modeli danych w systemach planujemy w pierwszym etapie Opracowanie Planu wdrażania Strategii DG w CeZ i obejmowanie jej etapami.

4. Czy w ODS będą przechowywane dane objęte regulacjami (osobowe, medyczne itp.)?

Odpowiedź: Zakładamy przechowywanie danych objętych regulacjami.

5. Jakiego rodzaju dane mają być przechowywane przez platformę Zarządzania Danymi?

Odpowiedź: Przewidujemy przechowywanie danych tj. dane referencyjne, słowniki danych, metadane, Kanoniczny model danych z uwzględnieniem wszystkich domen danych w CeZ.

6. O backupowaniu i odzyskiwaniu jakich danych jest mowa w wymaganiu 9?

Odpowiedź: Mowa tutaj o możliwych dostępnych metodach i narzędziach tworzenia kopii zapasowych, ich częstotliwości oraz procedur odzyskiwania danych w razie awarii platformy Zarządzania Danymi (Data Governance) i danych, które są przedmiotem funkcjonowania lub efektem działania Data Governance.

7. Wydajność jakich rozwiązań ma być monitorowana przez platformę? Czy mowa o pozostałych modułach Huba Danych (ODS, ETL, AI/ML?)

Odpowiedź: Wszystkie moduły mają być monitorowane (w tym będą zdefiniowane KPI dla działania usług aplikacyjnych). Ponadto monitorowaniem mają być objęte procesy zdefiniowane w ramach strategii Data Governance oraz mierniki jakości danych.

8. Bezpieczeństwo jakich danych ma być zapewnione? Czy mowa o danych przechowywanych na samej platformie Zarządzania Danymi, czy też danych przechowywanych w pozostałych modułach Huba Danych (ODS, ETL, AI/ML?)

Odpowiedź: Ze względu na strategiczne znaczenie i istotność danych przetwarzanych oraz serializowanych w systemach przez CeZ planowane rozwiązanie jest objęte maksymalnymi regulacjami dotyczącymi bezpieczeństwa. Nie zakłada się żadnych ustępstw. Inne zasady dot. bezpieczeństwa są stosowane w zakresie Danych Otwartych.

W ramach strategii Data Governance mają powstać rekomendacje dotyczące zarządzania bezpieczeństwem danych. Zakładamy, że będą one dotyczyły zarówno bezpieczeństwa danych, które są przedmiotem funkcjonowania lub efektem działania Data Governance, jak również dane przechowywane, przetwarzane w modułach Huba Danych.

9. Czy istnieją oczekiwania dotyczące konkretnych funkcjonalności platformy Data Governance, które są dla CEZ szczególnie istotne (np. katalog danych, lineage danych, polityki jakości danych, monitorowanie jakości danych)?

Odpowiedź: Platforma, jak i Strategia Data Governance powinna zapewniać skuteczne zarządzanie danymi w organizacji, kładąc nacisk na jakość, bezpieczeństwo, zgodność z regulacjami oraz dostępność danych. Poniżej przedstawiamy najważniejsze oczekiwania dotyczące szczególnie istotnych funkcjonalności:

- a. Zarządzanie i monitoring jakości danych (mechanizmy monitorowania i poprawy jakości danych, automatyczne wykrywanie i korekta błędów danych, standaryzacja danych, zarządzanie zmianami w danych, narzędzia do dokumentowania i wymiany wiedzy o danych).

- b. Zarządzanie cyklem życia danych (definiowanie i monitorowanie cyklu życia danych, mechanizmy klasyfikacji danych, automatyzacja anonimizacji i maskowania danych).
  - c. Bezpieczeństwo i kontrola dostępu, zgodność z regulacjami prawnymi (zarządzanie uprawnieniami i rolami, szyfrowanie danych, monitorowanie dostępu do danych, mechanizmy zapewnienia zgodności z m.in. RODO, HIPAA, inne).
  - d. Integracja i zarządzanie danymi (Data Catalog, Kanoniczny Model Danych, integracja z hurtowaniami danych, zarządzanie linią danych Data Lineage).
  - e. Monitorowanie, audyt i raportowanie (automatyczne alerty o naruszeniach polityk Data Governance, raporty o stanie i wykorzystaniu danych, śledzenie zmian w danych).
10. Jaka jest aktualna infrastruktura IT Centrum e-Zdrowia? (serwery, centra danych, chmura obliczeniowa - jeśli dotyczy).
- Odpowiedź: Działanie w modelu on-premise. Nie możemy udostępnić specyfikacji infrastruktury oraz jej opisu ze względów bezpieczeństwa.
11. Jak obecnie wygląda zarządzanie danymi w Centrum e-Zdrowia? (procesy, role, odpowiedzialności).
- Odpowiedź: Wynika z dokumentu System Zarządzania Bezpieczeństwem Informacji (SZBI) – na podstawie normy ISO 27001.
12. Czy istnieją udokumentowane polityki i procedury dotyczące zarządzania danymi w CEZ?
- Odpowiedź: Istnieje System Zarządzania Bezpieczeństwem Informacji (SZBI) – na podstawie normy ISO 27001.
13. Jakie systemy dziedzinowe w CEZ będą źródłem danych dla platformy Data Governance? (proszę o listę i krótki opis).
- Odpowiedź: Docelowo mają to być wszystkie systemy dziedzinowe CeZ.
14. Czy istnieją wymagania dotyczące zgodności z przepisami (np. RODO, ustawa o ochronie danych osobowych w ochronie zdrowia), które platforma Data Governance ma wspierać?
- Odpowiedź: System ma być w pełni zgodny z literą prawa. Jest to jedno z pryncypiów CeZ. Co do zasady Data Governance ma wspierać mechanizmy zapewnienia i monitorowania zgodności z RODO, HIPAA, ISO/IEC 27001, ISO IDMP, HL7 na możliwie wysokim poziomie zautomatyzowania. Oczekujemy podczas konsultacji rekomendacji jakie wymagania należy uwzględnić w tym zakresie.
15. Jakie funkcjonalności platformy Data Governance są dla CEZ absolutnie kluczowe i niezbędne do poprawnego działania?
- Odpowiedź: Odpowiedź udzielona w pkt 9.
16. Czy CEZ ma szczególne wymagania dotyczące integracji platformy Data Governance z innymi systemami używanymi w CEZ, które powinny być uwzględnione w OPZ?
- Odpowiedź: Wynikają one z odpowiedzi na pytanie 9 oraz z założenia o możliwym maksymalnym zautomatyzowaniu procesów integracji i jej monitorowaniu a także odporności rozwiązania na ewentualne zmiany danych.
17. Czy istnieją konkretne standardy bezpieczeństwa (np. ISO 27001, NIST) lub regulacje prawne (np. RODO) dotyczące bezpieczeństwa danych, które musimy uwzględnić?
- ISO 27001, ISO 22301, NIS2 w tym opracowywana ustawa o Krajowym Cyberbezpieczeństwie



Odpowiedź: ISO 27001, ISO 22301, NIS2 w tym opracowywana ustawa o Krajowym Cyberbezpieczeństwie oraz HL7, Normy ISO IDMP, RODO. Oczekujemy również w miarę możliwości uwzględnienia Europejski AI Act.

Osoba sporządzająca: Hryniewicz-Braham Agnieszka