



AGENCJA
BADAŃ
MEDYCZNYCH

REGIONALNE CENTRA MEDYCNY CYFROWEJ

STANDARD

WARSZAWA, MARZEC 2023

SPIS TREŚCI

1. WIZJA REGIONALNYCH CENTRÓW MEDYCZNY CYFROWEJ.....	4
2. UZASADNIENIE DLA TWORZENIA REGIONALNYCH CENTRÓW MEDYCZNY CYFROWEJ	5
3. CELE UTWORZENIA REGIONALNYCH CENTRÓW MEDYCZNY CYFROWEJ.....	6
3.1. Zadania RCMC	7
4. ŹRÓDŁA DANYCH	10
4.1. Poziom regionalny (RCMC przy CWBK).....	10
4.2. Zbieranie Danych medycznych.....	11
5. GŁÓWNE FUNKCJONALNOŚCI.....	13
5.1. Poziom regionalny RCMC.....	13
6. STANDARDY (WYMAGANIA) REGIONALNEGO CENTRUM MEDYCZNY CYFROWEJ	16
6.1. Infrastruktura	16
6.1.1. Sprzęt i zaplecze techniczne	16
6.1.2. Infrastruktura biurowa RCMC.....	16
6.1.3. Biobank	16
6.2. Kadry	18
6.2.1. Kadra RCMC.....	18
6.2.2. Kadra Biobanku	20
6.3. Doskonalenie kadry - szkolenia	21
6.3.1. Kadra RCMC.....	21
6.3.2. Kadra Biobanku	22
6.3.3. Krzywa uczenia	23
6.4. Zapewnienie ciągłości działania	23
6.5. Bezpieczeństwo danych	24
6.6. Minimalna jakość wyjściowa danych zapewniająca możliwość analizy zagregowanych danych (typy, jakość, format i wielkość plików)	27
6.7. Wprowadzenie standardu danych medycznych.....	28
5.7.2 Dane obrazowe.....	29
6.7.1. Dane sekwencyjne	29
6.7.2. Inne dane	30

6.8. Udostępnianie danych i ich wymiana	30
6.9. Aspekty etyczne i prawne	31
7. OCZEKIWANE REZULTATY I KORZYŚCI	33

1. WIZJA REGIONALNYCH CENTRÓW MEDYCZNY CYFROWEJ

Regionalne Centra Medycyny Cyfrowej (RCMC) będą strukturami łączącymi bazy danych uniwersytetów, szpitali klinicznych, Biobanków i innych.

Centra Medycyny Cyfrowej będą pełniły rolę jednostki organizacyjnej w strukturach Wnioskodawcy, odpowiedzialnej za procesy związane z wdrożeniem innowacyjnych instrumentów w zakresie cyfryzacji.

Celem powstania Regionalnych Centrów Medycyny Cyfrowej jest standaryzacja pozyskiwania i przetwarzania wysokiej jakości Danych medycznych do celów naukowych i analitycznych oraz zapewnienie możliwości bezpiecznej wymiany ustrukturyzowanych danych

Centra Medycyny Cyfrowej powstają w celu uporządkowania systemu zarządzania Danymi medycznymi na poziomie głównych ośrodków uniwersyteckich i klinicznych, a w dalszej perspektywie mają na celu wybudowanie połączeń pomiędzy tymi jednostkami i zbieranie danych z jednostek współpracujących przy rozwoju RCMC. Dane pozyskiwane przez RCMC muszą być, zależnie od źródła, Anonimowane, Pseudonimizowane lub jawne. Podział ten zależy od wizji jednostki tworzącej RCMC. Przy aktualnym stanie prawnym RCMC musi posiadać możliwość pełnego zanomizowania danych, tak aby móc przekazać je innym RCMC. Dane zbierane w ramach RCMC nie mogą służyć budowaniu prywatnych baz danych. Jednostki tworzące RCMC zobowiązują się dzielić danymi między sobą w celu rozwoju nauki i budowaniu współpracy. Rozwijając RCMC należy wziąć pod uwagę, zmieniające się otoczenie prawne oraz być gotowym do dzielenia się danymi z innymi jednostkami.

Planując rozwój RCMC jednostka musi przeprowadzić szereg analiz aby prawidłowo wprowadzić funkcje systemu RCMC w swoich strukturach. W pierwszym etapie konieczne jest przeprowadzanie m.in.: analizy prawnej, procesowej, wdrożeniowej, informatycznej i technicznej posiadanych zasobów oraz uwzględnienie możliwości ich integracji oraz interoperacyjności. Możliwe jest rozwijanie projektu w oparciu o stopniową adopcję technologii tzw. agilowe podejście iteracyjne.

RCMC rozwinią samodzielnie narzędzia i techniki pracy z danymi, prowadząc do rozwoju medycyny cyfrowej.

RCMC będzie stanowić wsparcie dla prowadzenia Badań klinicznych na różnych etapach (np. obliczanie mocy, analiza wyników).

Wymiana i analiza wyników między Centrami stanowić będzie przełom m.in. w chorobach rzadkich, w których badania zazwyczaj wykonywane są na małych grupach pacjentów.

RCMC będzie również przesyłać dane zgodnie z żądaniami Centrali CMC (ABM) celem analiz big data związanych z analizą genomu populacji polskiej, co jest zgodne z unijnym programem cyfrowej transformacji zdrowia¹, rozporządzenie Parlamentu Europejskiego i Rady (UE) 2021/522 z dnia 24 marca 2021 r. w sprawie ustanowienia Programu działań Unii w dziedzinie zdrowia („Program UE dla zdrowia”) na lata 2021-2027 oraz uchylenia rozporządzenia (UE) nr 282/2014.

2. UZASADNIENIE DLA TWORZENIA REGIONALNYCH CENTRÓW MEDYCYNY CYFROWEJ

ABM w ramach swojej działalności sfinansowała powstanie Centrów Wsparcia Badań Klinicznych, które należą do Polskiej Sieci Badań Klinicznych. Kolejnym etapem jest utworzenie w ramach jednostek, które posiadają w swojej strukturze CWBK, infrastruktury wspierającej rozwój cyfryzacji medycyny oraz umożliwiającej maksymalne wykorzystanie danych zdrowotnych pacjentów i osób, które chcieliby włączyć się do szeroko zakrojonego programu rozwoju medycyny cyfrowej.

Regionalne Centra Medycyny Cyfrowej (RCMC) zrzeszone w Sieć Centrów Medycyny Cyfrowej (SCMC) z nadzczną rolą ABM, docelowo w postaci Centrali Medycyny Cyfrowej ABM (CMC ABM), będą służyły zarówno analizie danych zbieranych przez dany ośrodek w czasie rzeczywistym, wsparciu Badań klinicznych i opieki szpitalnej w obszarze rozwiązań cyfrowych, jak i analizie retrospektywnej. Rodzaj gromadzonych danych będzie obejmował m.in. dane dotyczące zdrowia (rozumiane zgodnie z art. 4 pkt. 15 RODO) dane omicznne pozyskane z gromadzonych w ośrodkach naukowych próbek z projektów badawczych (w tym z niekomercyjnych Badań klinicznych finansowanych przez ABM) opisanych

¹ <https://www.gov.pl/web/zdrowie/programy-unii-w-dziedzinie-zdrowia>

w Podręczniku „*Standardy jakości dla Biobanków polskich*” v.2.0², oraz inne dane zbierane w szpitalnych systemach informacyjnych (HIS) lub innych repozytoriach (m.in. dane obrazowe, dane dotyczące wykonywanych procedur oraz dane pochodzące z notatek lekarskich).

Dane będą kolekcjonowane zgodnie z istniejącymi standardami jakości (dopasowanymi przez jednostkę i jednocześnie pozostającymi w zgodzie z obecnym kierunkiem europejskim w tym obszarze), zapewniając ich przydatność do analiz, zachowując jednocześnie gwarancję bezpieczeństwa udostępniania i przechowywania. W dalszej kolejności ABM podejmie wysiłki, aby zbierane dane były możliwe do wykorzystania przez inne podmioty zewnętrzne (rządowe, akademickie) z zachowaniem międzynarodowych standardów, bezpieczeństwa danych oraz tworzonej nowej legislatury UE. Idea Konkursu jest zgodna z działaniami UE na rzecz wspólnych europejskich przestrzeni danych zapewniających dostęp do dużej liczby wiarygodnych danych, przy jednoczesnym zachowaniu kontroli nad źródłami, które generują te dane. Docelowo europejskie przestrzenie mają być budowane z udziałem Europejskiej Chmury dla Otwartej Nauki (EOSC), co ma za zadanie zapewnienie europejskim naukowcom i przedsiębiorstwom otwartego multidyscyplinarnego środowiska, w którym będą mogli wyszukiwać i ponownie wykorzystywać dane.

Wtórna analiza danych w dalszym ciągu jest poważnym wyzwaniem logistyczno-prawnym, ale może stanowić przełom w ocenie rozwiązań diagnostycznych i terapeutycznych, zarówno na poziomie pojedynczego Regionalnego Centrum Medycyny Cyfrowej, przy współpracy poszczególnych Centrów w Sieci, jak i docelowo w przypadku analiz zbiorczych w ramach Centrali Medycyny Cyfrowej ABM. Stworzenie systemu, który pozwoli maksymalnie wykorzystać Dane medyczne jest więc zasadne i ma potencjał, żeby zmienić znacząco sposób przetwarzania Danych medycznych w Polsce.

3. CELE UTWORZENIA REGIONALNYCH CENTRÓW MEDYCYNY CYFROWEJ

Głavnym celem utworzenia Regionalnych Centrów Medyczny Cyfrowej w ramach jednostek, które posiadają w swojej strukturze CWBK i należą do Polskiej Sieci Badań Klinicznych, jest stworzenie zaplecza infrastrukturalnego, kadrowego oraz systemowego umożliwiającego docelowo powstanie Sieci Centrów Medycyny Cyfrowej. Umiejscowienie

² https://wydawnictwo.umw.edu.pl/upload/files/standardy_jakosci_dla_biobankow_polskich_2.0%281%29.pdf

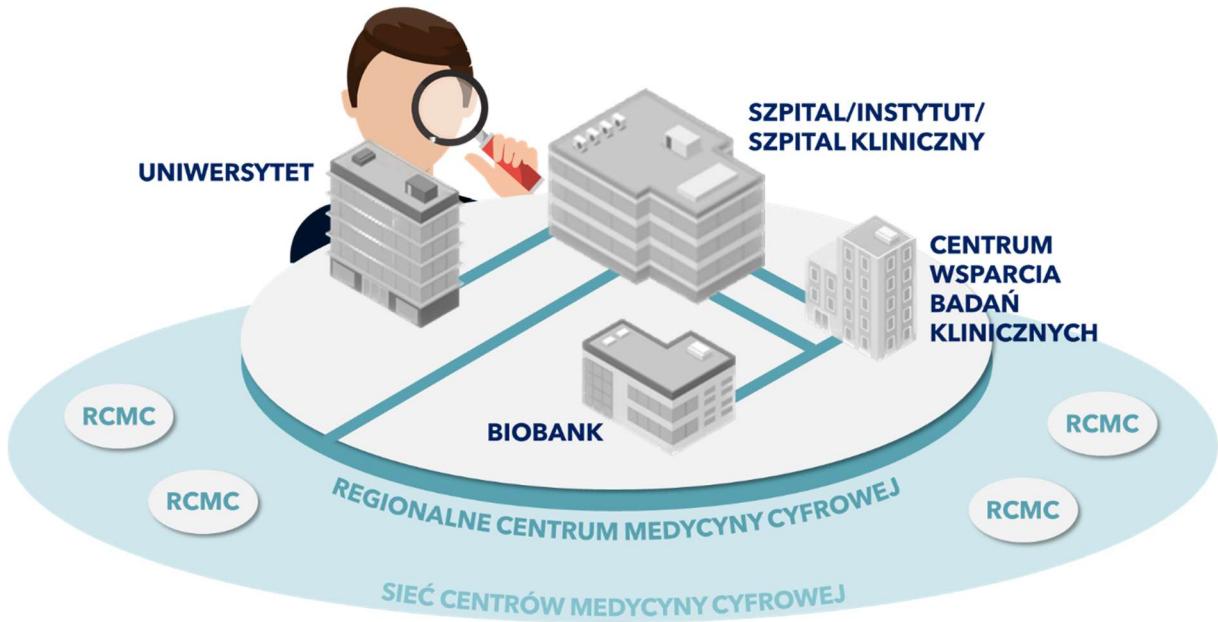
RCMC w strukturze organizacyjnej jednostki jest uzależnione od celów stawianych RCMC, jego wewnętrznych uwarunkowań organizacyjnych i jest decyją jednostki.

Cel główny będzie wspierany przez realizację celów pomocniczych:

- a) Pomoc RCMC w projektowaniu wybranych/wskazanych przez badaczy elementów Badań klinicznych (np. założeń statystycznych), magazynowaniu danych i ich analizie w czasie rzeczywistym, zgodnie z istniejącymi standardami, jak również odpowiedzialność za retrospektywną analizę danych, ich Anonimizację i Pseudonimizację (jeśli dotyczy) i docelowy transfer przy ścisłej współpracy z ABM.
- b) Generowanie przez RCMC wysokiej jakości zbiorów danych klinicznych i omicznych, tworzących bazę do tworzenia cyfrowych narzędzi o charakterze prognostycznych, predykcyjnych i terapeutycznych algorytmów sztucznej inteligencji wykorzystywanych w prewencji, diagnostyce i terapii chorób.
- c) Opracowanie przez RCMC technicznych standardów udostępniania danych i analizy danych wtórnych w porozumieniu z ABM.
- d) Powstanie w RCMC algorytmów i narzędzi pozwalających na analizy zintegrowanych danych oraz wydajnego oprogramowania do zabezpieczania i udostępniania danych.
- e) Powstanie w RCMC zaplecza pozwalającego na bezpośrednią analizę danych (m.in. danych omicznych) uzyskanych z próbek biologicznych pochodzących od uczestników Badań klinicznych lub osób, które chcieliby włączyć się do programu rozwoju medycyny cyfrowej.
- f) Świadczenie usług medycyny cyfrowej w ramach projektów badawczych ze szczególnym uwzględnieniem tych finansowanych przez ABM, m.in. z zakresie analiz wyników badań genetycznych, projektowania badań w aspekcie cyfryzacji czy analizy danych.

3.1. Zadania RCMC

Regionalne Centra Medycyny Cyfrowej będą pełniły rolę jednostki organizacyjnej w strukturach jednostki odpowiedzialnej za procesy związane z wdrożeniem innowacyjnych instrumentów w zakresie cyfryzacji. RCMC może znajdować się w strukturze organizacyjnej Uniwersytetu/ Szpitala/ Instytutu/ Szpitala Klinicznego/ CWBK/ Biobanku. RCMC jest zobowiązane świadczyć wsparcie na rzecz pozostałych jednostek.



Przykładowe zadania RCMC:

1. Tworzenie, rozwój, upowszechnianie i wdrożenie inteligentnych rozwiązań, w tym sztucznej inteligencji m.in. w zakresie:
 - tworzenia cyfrowych narzędzi o charakterze prognostycznych, predykcyjnych i terapeutycznych algorytmów sztucznej inteligencji opartych o wygenerowane w ośrodkach zbiory danych klinicznych i omiczych,
 - algorytmów dawkowania leków,
 - algorytmów generujących ostrzeżenia o niepożądanych skutkach przy przepisywaniu kombinacji wielolekowych,
 - algorytmów poprawiających monitoring pacjenta,
 - algorytmów poprawiających zarządzanie procesem leczniczym,
 - integracje i analizę danych w ramach RCMC z Systemów PACS (Picture Archiving and Communication System) oraz VNA (Vendor Neutral Archive) służące do archiwizacji i udostępniania obrazowej dokumentacji medycznej z różnych źródeł,
 - integracje i analizę danych w ramach RCMC z Systemów CAD (Computer-Assisted Diagnosis) umożliwiające diagnozę na podstawie obrazowania,
 - systemów wspomagania decyzji lekarskich,
 - rozwój systemów wewnętrznych IT jednostki umożliwiających digitalizację dokumentacji medycznej, tworzenie algorytmów AI wspomagających procesy zarządcze w jednostce,
 - wsparcie prac badawczo-rozwojowych związanych z medycyną cyfrową,

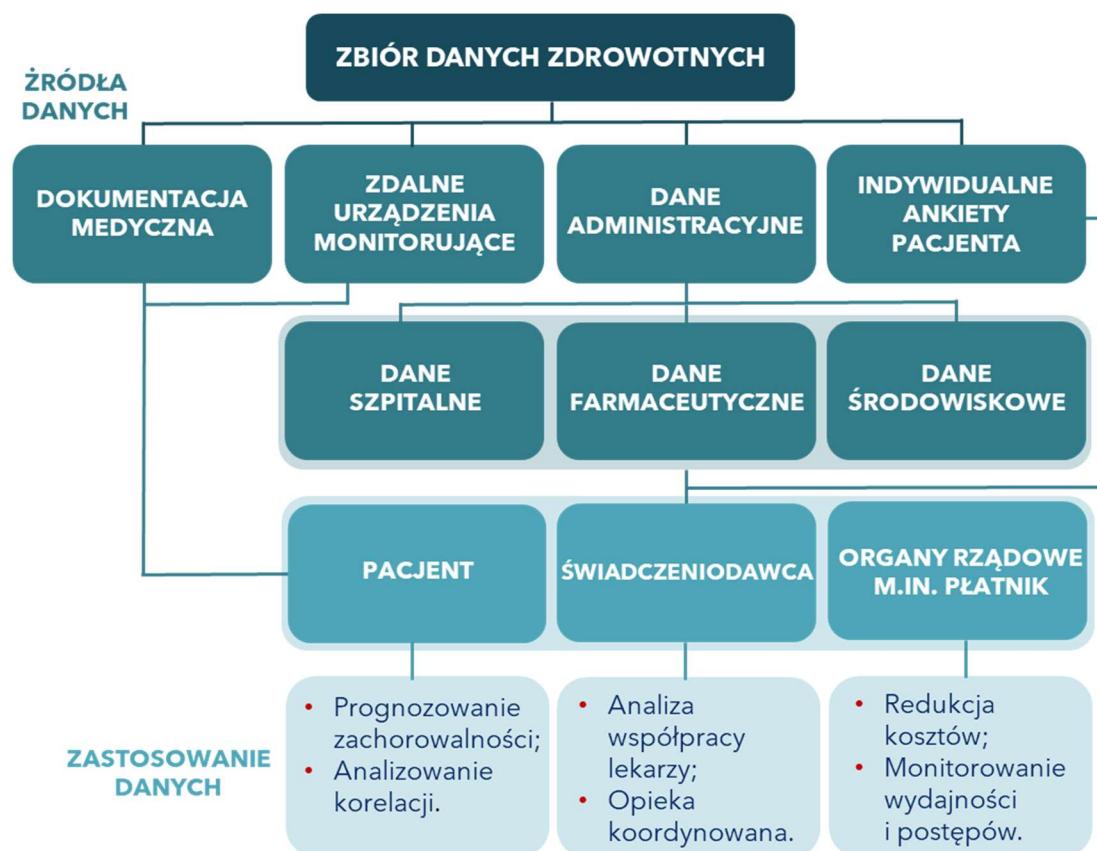
- koordynacja współpracy pomiędzy lekarzami, informatykami i biostatystykami w celu rozwijania algorytmu AI,
- tworzenie lokalnych baz danych, które z łatwością mogą zostać włączone w budowane rozwiązania systemowe,
- wsparcie prac związanych z tworzeniem i rozwojem ustandaryzowanych Biobanków. Celem poprawy jakości pozyskiwanych danych przewiduje się również utworzenie, rozwinięcie i/lub doposażanie współpracujących z Centrami Wsparcia Badań Klinicznych Biobanków, tak aby przechowywane próbki, a następnie pozyskiwane z nich dane spełniały standardy niezbędne do dalszej sprawnej wymiany informacji i analiz w ramach Sieci CMC.

4. ŹRÓDŁA DANYCH

4.1. Poziom regionalny (RCMC przy CWBK)

Jednostka będzie musiała wykazać dokładny podział obowiązków i kompetencji pomiędzy RCMC i CWBK oraz sposoby przepływu między dostarczycielami danych i ryzyka z tym związane.

RCMC będzie odpowiedzialne za zbiór Danych medycznych (ang. *data management*) zawierający m.in. dane opisowe, farmaceutyczne, obrazowe, wyniki badań pochodzące zarówno z systemów HIS, jak i eCFR oraz innych źródeł, jak choćby z urządzeń monitorujących pacjenta czy wykonujących analizy. Istotnym jest zatem, aby zespół RCMC został zaznajomiony z procesem badawczym w CWBK i wszystkimi możliwymi źródłami danych, które temu procesowi towarzyszą, celem maksymalnego usprawnienia procesu implementacji narzędzi do pracy z danymi.



ZBIERANIE RÓŻNORODNYCH DANYCH

PRZETWARZANIE



Pakiety danych, dane strumieniowe, dane ustrukturyzowane oraz nieuszkodzone.



Przetwarzanie i zarządzanie wszystkich danych.

REALIZACJA



Generowanie raportów z wykonanych analiz



Kompletny pakiet narzędzi analitycznych z możliwością śledzenia i zarządzania

UDOSTĘPNIANIE



Zakres zbieranych danych powinien być dostosowany do specyfiki i profilu badawczego danego RCMC.

4.2. Zbieranie Danych medycznych

Tworzone RCMC będzie musiało mieć gotowość i kompetencje, aby czerpać z wielu różnych źródeł danych, różnych nie tylko pod względem ich charakteru ale przede wszystkim - o różnym stopniu zdigitalizowania.

Minimalny zakres danych zbieranych z własnych źródeł:

- Szpitalne systemy informatyczne (HIS)³ ze szczególnym uwzględnieniem:
 - danych demograficznych,
 - jednostek chorobowych (ICD10) - głównych oraz współistniejących,
 - wykonanych procedur (ICD9),
 - przyjmowanych leków (nazwa substancji czynnej, kod EAN),
 - pobytu w ramach hospitalizacji i wizyt (minimum data rozpoczęcia i zakończenia),
 - czasy wystąpienia zdarzeń,

³ Dane pozyskane w trakcie Badań klinicznych muszą być oznaczone

- dane opisowe – dokumentacja medyczna niepowiązana z wykonaniem usługi (np. epikryzy, notatki lekarskie, obserwacje pielęgniarskie, formularze opisujące stan pacjenta itp.)⁴,
- Diagnostyka laboratoryjna (wyniki badań laboratoryjnych),
- Diagnostyka obrazowa (opisy badań oraz obrazy),
- Diagnostyka histopatologiczna (opisy badań oraz obrazy),
- Informacje o dostępności materiału biologicznego przechowywanego w Biobanku.
- Badania genomu (whole genome sequencing, whole exome sequencing).

Rekomendowane jest zbieranie danych z innych źródeł:

- Wyników badań omicznich (metabolomicznych, proteomicznych, transkryptomicznych),
- Pozostałych zakładów/pracowni współpracujących z RCMC,
- Telemetrii – dane z wyrobów medycznych i zdalnych urządzeń monitorujących (m.in. holter, opaski sportowe),
- Indywidualnych ankiet dotyczących samopoczucia Pacjenta i/lub stylu życia i/lub jakości życia w oparciu o wystandardyzowane, międzynarodowe wzory np. SF-36, EQ-5D, HRQoL-14.

W przypadku Dawców-wolontariuszy minimalny zakres danych zawarty został w formularzu

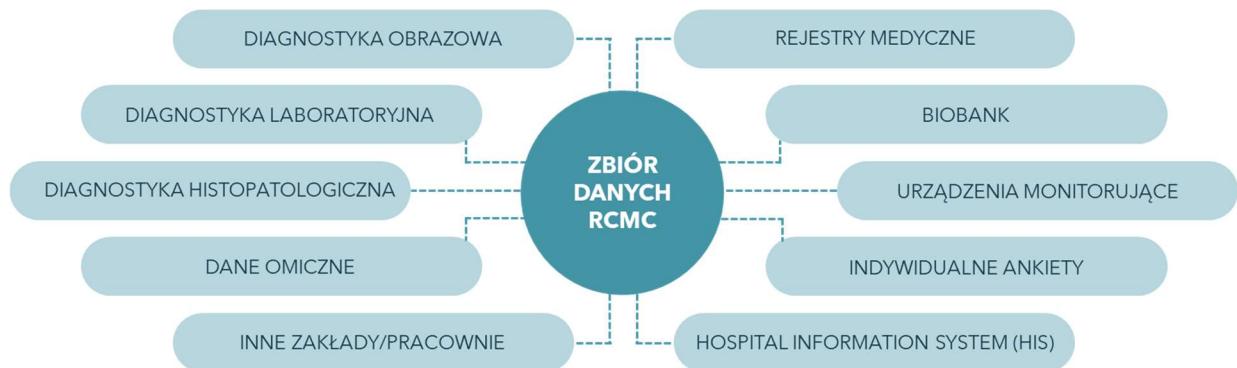
Załącznik nr 1 do Standardu RCMC - Formularz Świadomej Zgody.

Wszystkie rodzaje wyżej wymienionych danych powinny być odpowiednio Anonimizowane lub Pseudonimizowane i oznaczone tym samym unikatowym kodem pacjenta. Zakres i Anonimizacji i Pseudonimizacji, w kontekście współpracy pomiędzy interesariuszami (uniwersytet, szpital kliniczny, Biobank i inni) określa RCMC. Na poziomie RCMC dane pacjenta powinny być Pseudoanimizowane . Dane przekazywane przez RCMC innym RCMC i ABM muszą być zanonimizowane, chyba, że inne przepisy (m.in. przepisy o rejestrach medycznych) pozwalają na przekazanie danych jawnych. Zapewni to zachowanie spójności wszystkich danych w centralnej jednostce przechowującej dane.

⁴ Jeżeli dana opisowa jest w postaci pola tekstowego a nie ustrukturyzowanego formularza należy ją wprowadzić jako wartość tekstową

5. GŁÓWNE FUNKCJONALNOŚCI

5.1. Poziom regionalny RCMC



A. System informatyczny RCMC musi:

- umożliwiać przetwarzanie, skalowanie, zarządzanie, przeszukiwanie i analizę statystyczną zbieranych danych,
- generować raporty i podsumowania,
- posiadać możliwość obsługi wszystkich rodzajów danych – ustrukturyzowanych i nieustrukturyzowanych,
- selekcjonować i indeksować dane w celu przyspieszenia procesu ich przetwarzania,
- stosować zunifikowane słowniki,
- zapewniać Pseudonimizację i Anonimizację danych zależnie od potrzeb jednostki,
- docelowo posiadać możliwość integracji z systemem centralnym poprzez udostępnienie application programming interface (ang. API),
- docelowo umożliwiać wymianę/dzielenie się danymi z innymi RCMC zgodnie z obowiązującym stanem prawnym,
- mieć strukturę modularną umożliwiającą opracowywanie i dodanie nowych narzędzi i funkcjonalności,
- posiadać stosowne zabezpieczenia zapewniające ochronę gromadzonych danych.

B. Wymagane jest aby system informatyczny RCMC:

- wykorzystywał algorytmy AI,
- posiadał oprogramowanie umożliwiające uczenie maszynowe (ang. ML – machine learning),
- umożliwiał tworzenie modeli matematyczno-statystycznych.

C. Wymagane jest aby został stworzony system informatyczny (Moduł asystent lekarza) strukturyzujący dane wprowadzane do HIS.

Moduł asystenta lekarza oszczędzi czas lekarzy podczas normalizacji i strukturyzowania dokumentacji medycznej on-line. Dzięki asystentowi notatka będzie w pełni analizowana natychmiast po jej przygotowaniu. Na podstawie dowolnego tekstu wprowadzonego przez lekarza i rozpoznanego przez silnik asystenta, powinien podpowiadać, jak sprawić, by zdanie było trafne, jasne, jednoznaczne i porównywalne niezależnie od języka/słownika/słownictwa, w którym zostało utworzone i/lub użyte. Ujednolicenie specyficznego słownictwa medycznego umożliwia pełną wymienność danych gromadzonych w notatkach medycznych.

Asystent powinien posiadać/ umożliwiać:

- ustrukturyzowane rozpoznawanie mowy i tekstu⁵,
- integrację z systemami IT w placówkach medycznych,
- używanie złożonych skal np. RECIST 1.1 i PIRADS,
- własne szablony lub możliwe do dostosowania do swoich potrzeb te z ogólnodostępnej biblioteki,
- gromadzenie danych w chmurze i/lub na terenie placówki.

D. Wymagane jest aby system informatyczny RCMC miał Moduł Feasibility.

- Moduł musi umożliwiać spójne prezentowanie danych pochodzących z różnych źródeł (nałożenie abstrakcji na zimportowane dane), poprzez jednolite nazwy atrybutów, łączenie grup danych w zdefiniowane typy obiektów takie jak np.: diagnoza, hospitalizacja, badanie, lekarz oraz musi pozwalać na łatwe wyszukiwanie pacjentów według zadanej kwerendy.
- Zakres danych zimportowanych z różnych źródeł, musi być zgodny z minimalnym zakresem danych wskazanym w standardzie (moduł powinien zapewniać Pseudonimizację i Anonimizację danych zależnie od potrzeb jednostki). Moduł musi być wyposażony w mechanizm umożliwiający tworzenie zapytań bez posiadania specjalistycznej wiedzy z zakresu obsługi baz danych, programowania czy statystyki. Użytkownik z podstawową wiedzą medyczną powinien być w stanie wygenerować zapytanie.

⁵ Automatyczne ekstrahowanie kodów EAN, ICD-9, ICD-10 oraz konceptów SNOMED CT z niestrukturyzowanych dokumentów

- Moduł musi posiadać możliwość tworzenia kryteriów definiowania i wyszukiwania grup pacjentów z posiadanej bazy danych.
- Moduł musi posiadać możliwość wyszukiwania grup pacjentów na podstawie dowolnie zdefiniowanych schematów leczenia.
- Moduł musi umożliwiać wyświetlenie wszystkich innych zdarzeń występujących między dwoma zależnymi w czasie zdarzeniami. Moduł musi umożliwiać zawężenie wyników wyszukiwania do danego przedziału czasu np. ostatni miesiąc, ostatni kwartał, dowolny zakres dat itp.
- Moduł musi umożliwiać definiowanie kryteriów wyszukiwania (w tym opartych o schematy leczenia).

6. STANDARDY (WYMAGANIA) REGIONALNEGO CENTRUM MEDYCZNY CYFROWEJ

6.1. Infrastruktura

6.1.1. Sprzęt i zaplecze techniczne

Jednostka w wyniku realizacji projektu ma zapewnić infrastrukturę konieczną do gromadzenia, przetwarzania i zapewnienia bezpieczeństwa zebranych danych. W ramach utworzenia i rozwoju RCMC Jednostka musi opracować indywidualną koncepcję rozwoju RCMC w oparciu o zidentyfikowane i zdefiniowane potrzeby infrastrukturalno-techniczne. Jednostka w wyniku realizacji projektu RCMC powinna posiadać:

- Serwery umożliwiające efektywne działanie opracowanego systemu RCMC*,
- Przestrzeń dyskową dostosowaną do charakterystyki systemu RCMC*,
- Pozostały sprzęt niezbędny do efektywnego działania systemu RCMC określony w oparciu o przeprowadzoną analizę potrzeb i analizę wdrożeniową,
- Zaplecze lokalowe dla kadr zarządzających i obsługujących RCMC,
- Zaplecze lokalowe dla sprzętu teleinformatycznego RCMC,
- Inne zaplecze lokalowe niezbędne do efektywnego funkcjonowania RCMC.

* Dopuszczalne są rozwiązania chmurowe.

6.1.2. Infrastruktura biurowa RCMC

Celem Projektu jest wdrożenie standardu w nowoutworzonym Regionalnym Centrum Medycyny Cyfrowej. Planowana liczba pomieszczeń, liczba i rodzaj zakupionego wyposażenia muszą być adekwatne do potrzeb jednostki. Jednostka ma zapewnić infrastrukturę konieczną do pracy personelu zatrudnionego w RCMC. Jednostka ma zapewnić adekwatną liczbę pomieszczeń i stanowisk pracy dla personelu zatrudnionego w RCMC oraz do pracy nad danymi przechowywanymi w RCMC.

6.1.3. Biobank

W ramach struktur RCMC, jednostka może rozwinać lub doposażyć Biobank, mogący być samodzielną jednostką niezależną od RCMC lecz ściśle z nim współpracującą w celu poprawy jakości pozyskiwanych danych i ich przetwarzania w ramach RCMC. Zależności

organizacyjne i funkcjonalne między Biobankiem i RCMC są uzależnione od wewnętrznych uwarunkowań organizacyjnych i planów jednostki i pozostają w jego gestii. Dzięki temu, informacje pozyskane z próbki materiału biologicznego pobranego od danego pacjenta mogłyby być wykorzystywana wielokrotnie w różnych badaniach. Istotnym elementem deponowania próbki jest udostępnienie historii próbki tj. warunków w jakich była ona pobrana, transportowana i przechowywana, zgodnie z obowiązującymi standardami Biobankowania. Rekomendowane jest kolekcjonowanie co najmniej dwóch rodzajów materiału biologicznego od pacjenta np. krew, bioptat, mocz, plwocina, itp. w zależności od rodzaju badania i profilu naukowego Biobanku. Za materiał podstawowy uznaje się krew obwodową/surowicę. W przypadku pacjentów onkologicznych istnieje dodatkowo wymóg Biobankowania próbek tkanki nowotworowej (np. fragmentu guza) w ilości umożliwiającej wykonanie sekwencjonowania, o ile pacjent wyrazi zgodę i istnieje kliniczna możliwość pozyskania takiego materiału. Umożliwi to korelację zmian dotyczących badanej jednostki chorobowej w różnych tkankach/płynach ustrojowych oraz poszukiwanie nowych biomarkerów danej choroby.

Jednostka w oparciu o zidentyfikowane potrzeby opracuje infrastrukturę i zasady analizy danych dla próbek biologicznych pochodzących z Badań klinicznych, Biobanków i od Dawców-wolontariuszy. Opracowana infrastruktura i zasady muszą spełniać standardy Podręcznika „*Standardy Jakości dla Biobanków Polskich 2.0*”. oraz obowiązujące przepisy prawa. Zgodność z międzynarodowymi standardami umożliwi także realizację współpracy międzynarodowej między ośrodkami, co poprawi wzmacni potencjał naukowy i umożliwi publikowanie wyników w czasopismach o wysokim wskaźniku odziewania w preferowanej formule Open Access.

W przypadku tworzenia, rozwijania lub doposażenia Biobanku jednostka, musi kierować się zasadami opisanymi w standardzie Podręcznik „*Standardy Jakości dla Biobanków Polskich 2.0*”.

Końcowym rezultatem tworzenia i/lub rozwoju Biobanków powinno być zaprojektowanie oraz walidacja procesu funkcjonowania Biobanku, w sposób umożliwiający unifikację SOP-ów między różnymi Biobankami zrzeszonymi w ramach Sieci CMC. Biobank będzie musiał wyrazić zgodę na przejście certyfikacji zewnętrznej, potwierdzającej działania w zgodzie z Podręcznikiem „*Standardy Jakości dla Biobanków Polskich 2.0*”.

6.2. Kadry

6.2.1. Kadra RCMC

RCMC będzie wymagało stworzenia nowych etatów, zależnie od potrzeb i stopnia zawansowania jednostki rozwijającej RCMC.

RCMC powinno zapewnić sprawne funkcjonowanie poprzez integrację, analizę, przetwarzanie i udostępnianie zbieranych danych w zależności od profilu naukowego RCMC w rozumieniu obszarów terapeutycznych w których jednostka prowadzi Badania naukowe. Kierownik RCMC powinien posiadać wiedzę i doświadczenie w zakresie rozwijania nowych narzędzi informatycznych w tym m.in. ich zastosowania w połączeniu z metodami współczesnej biologii molekularnej, nadzorowaniu i uczestniczeniu w testach i wdrożeniach systemów IT.

Personel powinien składać się z pracowników o 4 typach kompetencji:

INŻYNIER DANYCH (Data Engineer)	ADMINISTATOR DANYCH	BIOSTATYSTYK	DATA SCIENTIST
Osoba odpowiedzialna za prawidłowe zbieranie i przetwarzanie surowych danych, ocenę przydatności nowych źródeł informacji oraz projektowanie i uruchamianie nowych relacyjnych baz danych, pozwalających na przechowywanie oraz przetwarzanie informacji napływających do systemu RCMC.	Osoba, odpowiedzialna za sposób przechowywania, przetwarzania i udostępniania danych w odniesieniu do wyznaczonych przez jednostkę celów.	Osoba wykorzystująca narzędzia statystyczne do rozwiązywania problemów z zakresu medycyny, zdrowia publicznego lub biologii. Posiadająca doświadczenie w obsłudze algorytmów służących do analizy (wstępnej i wtórnej), wyników badań ośmiorzędnych	Specjalista ds. analityki, odpowiedzialny za analizowanie i interpretowanie danych, niezbędnych w podejmowaniu decyzji. Łączy elementy matematyki, statystyki i programowania komputerowego. Wykorzystuje zaawansowane techniki analityczne, takie jak uczenie maszynowe i modelowanie predykcyjne, wraz z zastosowaniem zasad naukowych.

Kompetencje nie są równoznaczne z liczbą etatów. Jeden pracownik może łączyć kilka kompetencji lub może być konieczne zatrudnienie większej liczby osób o określonych kompetencjach – Wnioskodawca powinien uzasadnić planowaną liczbę personelu, wraz z uwzględnieniem czasu potrzebnego na uzyskanie pełnej sprawności funkcjonowania

RCMC. Jednostka decyduje o liczbie oraz stanowiskach jakie utworzy w RCMC, a liczba i rodzaj stanowisk powinny zapewniać prawidłowe funkcjonowanie RCMC, jednak z uwzględnieniem minimum 4 kompetencji opisanych w tabeli powyżej. Zatrudnienie powinno odbywać się w kolejności analityk, następnie architekt. Data scientist powinien dołączyć dopiero na etapie funkcjonującej architektury.

Ponadto kadra RCMC powinna składać się ze specjalistów, których kompetencje powinny obejmować m.in.:

- umiejętność tworzenia modeli analitycznych opisujących systemy informatyczne (modele logiczne bazy danych, modele obiektów i logiki usług - API),
- ogólną znajomość relacyjnych baz danych i SQL, tworzenie i utrzymanie aplikacji internetowych, znajomości języków skryptowych (np. R, Python),
- umiejętność wykonywania analiz biostatystycznych i informatycznych,
- tworzenia dokumentacji analitycznej na etapie definiowania koncepcji, projektowania produkcji i wdrażania systemów IT, współpracy z programistami, testerami i architektami systemów IT,
- zapewnienie działania infrastruktury informatycznej,
- znajomość zasad ochrony i przetwarzania danych.

Dla realizacji celów RCMC niezbędna jest odpowiednia liczba wykwalifikowanych osób w ramach efektywnej i uporządkowanej struktury organizacyjnej. Strategia zatrudnienia pracowników powinna być oparta na bieżącej analizie sytuacji RCMC, planach rozwoju oraz uwzględniająca indywidualne potrzeby jednostki w kontekście zadań realizowanych w RCMC.

6.2.2. Kadra Biobanku

Biobank powinien opracować i wdrożyć procedurę zarządzania zasobami ludzkimi. Struktura organizacyjna Biobanku może się różnić w zależności od charakteru jednostki oraz zasobów, jakie posiada. Biobank powinien posiadać zatwierdzony schemat organizacyjny, w którym określone są zależności pomiędzy komórkami organizacyjnymi/funkcjonalnymi. W strukturze organizacyjnej Biobanku powinna być uwzględniona komórka/osoba odpowiedzialna za Systemem Zapewnienia i Zarządzania Jakością (SZiZJ).

Biobank powinien podlegać pod Komitet Naukowy/Radę Naukową lub inne ciało doradcze w zakresie prowadzonych projektów naukowych, technicznych i/lub administracyjnych.

Jednostka, w strukturze której znajduje się Biobank, powinna zapewnić odpowiednią liczbę wykwalifikowanych pracowników do płynnego przeprowadzenia wszystkich procesów realizowanych w Biobanku w sposób zapewniający trwałość po zakończeniu okresu realizacji projektu.

Należy określić wymagane kompetencje dla interdyscyplinarnego zespołu, zdefiniować zakres obowiązków oraz warunki pracy obejmujące także zagrożenia stanowiskowe. Kwalifikacje pracownika powinny obejmować wykształcenie, odbyte szkolenia lub doświadczenie zawodowe adekwatne do zakresu obowiązków wykonywanych w Biobanku.

Biobank powinien wskazywać odpowiedzialności i uprawnienia pracowników, określić zastępstwo za danego pracownika, aby zapewnić ciągłość wszystkich procesów zachodzących w Biobanku.

W celu zapewnienia sprawnego funkcjonowania Biobanku rekomenduje się zatrudnienie interdyscyplinarnego zespołu zapewniającego sprawne działanie Biobanku i jednocześnie umożliwiającego certyfikację zgodnie ze „Standardem Jakości dla Biobanków Polskich 2.0”. Kierownik Biobanku powinien posiadać kompetencje zapewniające kompleksowy nadzór nad działalnością Biobanku w tym: inicjowanie, prowadzenie i koordynowanie działalności naukowej w obszarze Biobankowania, definiowanie potencjału naukowo-badawczego Biobanku, wyznaczanie kierunków rozwoju Biobanku, tworzenie oferty usługowej Biobanku dla potencjalnych kontrahentów, budowanie partnerstwa naukowego w zakresie badań i rozwoju. W zakresie kadry Biobanku, Jednostka musi kierować się zasadami opisanymi w standardzie Podręcznik „Standardy Jakości dla Biobanków Polskich 2.0”.

6.3. Doskonalenie kadry - szkolenia

6.3.1. Kadra RCMC

Kompetentna kadra jest warunkiem sukcesu RCMC, dlatego szkolenie musi zostać zaplanowane przed uruchomieniem RCMC. Przeszkolony personel RCMC musi rozumieć swój zakres zadań (patrz: tabelki powyżej) i potrafić ze sobą współpracować w tych obszarach. Szkolenie musi mieć charakter praktyczny i odbywać się w obrębie jednostek docelowych (CWBK, szpital, uczelnia) tak aby szkolić się na istniejących, rzeczywistych danych a nie na symulacjach. Szkolenie teoretyczne powinno stanowić minimum. Należy również zadbać o procedury przeciwdziałające rotacji pracowników, ponieważ wyszkolenie personelu będzie czasochłonne.

Rekomenduje się opracowanie i wdrożenie polityki szkoleń wstępnych i okresowych oraz innych działań podnoszących kompetencje pracowników zatrudnionych w RCMC. Szkolenia mogą dotyczyć:

- wprowadzenia nowej metody w procesie,
- wprowadzenia zmiany w procesie technologicznym,
- odnowienia (przypomnienia) wiedzy z zakresu stosowanych procesów technologicznych,
- pozyskania nowej wiedzy w obszarach związanych oraz łączących się z Biobankowaniem,
- zarządzanie jakością i ryzykiem.

RCMC powinno przeszkolić każdego pracownika w zakresie szkoleń stanowiskowych w zakresie SOP-ów i procedur obowiązujących w RCMC a także umożliwić realizację szkoleń podnoszących kwalifikacje. Udział w szkoleniu zewnętrznym zawsze powinien być potwierdzony odpowiednim dokumentem, takim jak certyfikat, dyplom, zaświadczenie, o ile jest to przewidziane przez organizatora. W przypadku szkoleń wewnętrznych wystarczająca jest lista obecności.

6.3.2. Kadra Biobanku

Biobank powinien utworzyć oraz stosować politykę dotyczącą szkoleń okresowych pracowników. Biobank powinien przeszkolić każdego pracownika (wewnętrznie, tj. przez pracownika Biobanku, kierownika Biobanku, osoby uprawnione przez kierownika Biobanku, lub zewnętrznie, tj. przez podmioty niezwiązane z Biobankiem, ale posiadające odpowiednie kwalifikacje). Biobank powinien utworzyć plan szkoleń oraz aktualizować go. Plan szkoleń powinien być przygotowywany przez osobę odpowiedzialną za zarządzanie jakością a akceptować go powinien kierownik Biobanku. Akceptacja planu szkoleń musi zostać potwierdzona np. podpisem oraz datą akceptacji.

Biobank powinien okresowo weryfikować wiedzę pracowników wykonujących proces technologiczny, aby mieć pewność, iż proces wykonywany jest zgodnie z wewnętrznymi procedurami.

Uczestnictwo w szkoleniu zewnętrznym zawsze powinno być potwierdzone odpowiednim dokumentem, takim jak certyfikat, dyplom, zaświadczenie, o ile jest to przewidziane przez organizatora. W przypadku szkoleń wewnętrznych wystarczająca jest lista obecności. Kopie

dokumentów potwierdzających udział pracownika Biobanku w szkoleniu powinny być przechowywane przez kierownika Biobanku bądź inną uprawnioną osobę. Biobank powinien zweryfikować wiedzę uczestników szkolenia wewnętrznego wykorzystując narzędzia takie jak: test pisemny, pytania ustne czy test sprawdzający umiejętności praktyczne.

6.3.3. Krzywa uczenia

Jednostka powołująca RCMC powinna zaproponować sposoby weryfikacji kompetencji personelu związane z tym konkretnym zakresem kompetencji CWBK (zakres zagadnień, rodzaj danych), poprzez opracowanie modelu weryfikacji i dokumentacji zdobycia kompetencji. Co ważne, planując dalszy rozwój CWBK i RCMC konieczne będzie zaplanowanie pełnego przeszkolenia i wdrożenia każdej nowozatrudnionej osoby, co oznacza, że konieczne będzie przewidzenie buforu potrzebnego na adaptację, zanim zatrudniony personel przejdzie do wykonywania obowiązków. Zatem jednostka musi przewidzieć opóźnienia związane z krzywą uczenia i przedstawić sposoby radzenia sobie z ryzykami stanowiącymi następstwo takiego stanu rzeczy.

6.4. Zapewnienie ciągłości działania

RCMC w celu zapewnienia ciągłości działania powinno zapewnić warunki umożliwiające odtworzenie działalności po katastrofie bądź innym incydencie zakłócającym działanie (np. cyberataku). W celu zapewnienia ciągłości działania jednostka zobowiązana jest do opracowania Planu ciągłości działania (ang. *Business Continuity Plan*) w tym Plany odtworzeniowe po awarii (ang. *Disaster Recovery Plan*), uwzględniające:

- A. Zapewnienie możliwości odtworzenia się z bezpiecznej kopii.
- B. Spełnienie zakładanego poziomu świadczenia RCMC po odtworzeniu.

RCMC powinno zapewnić:

- utrzymanie i bezpieczną eksploatację systemu informacyjnego,
- bezpieczeństwo fizyczne i środowiskowe, uwzględniające kontrolę dostępu,
- bezpieczeństwo i ciągłość dostaw usług, od których zależy działanie RCMC,
- Kopie zapasowe jako gwarancja odtworzenia po awarii:

- Dla zachowania ciągłości działania, np. na wypadek awarii powinny być realizowane kopie zapasowe ze wszystkich systemów, składających się na świadczoną usługę, zgodnie z ustalonym uprzednio harmonogramem i nadanymi priorytetami ich realizacji.
- Kopie te powinny być testowane okresowo poprzez ich odtwarzanie na dedykowanym środowisku odtworzeniowym i powinny podlegać obowiązkowej i każdorazowej weryfikacji poprawności oraz powinny zawierać raporty z odtworzenia.
- Kopie zapasowe dla zapewnienia im bezpieczeństwa, powinny być regularnie tworzone niezależnie od środowiska on-premise, czy chmurowego (np. na trwałe taśmy lub inne systemy back-up) i podlegać odmiejscowieniu na wypadek np. pożaru jednego z ośrodków przetwarzania (w przypadku środowisk on-premise), przy kopiach zapasowych rekomenduje się stosować zasadę 3-2-1. Zasada polega na tym, żeby przechowywać przynajmniej trzy kopie swoich danych (3 x back-up) i przechowywać dwie kopie back-upu na różnych urządzeniach (2 x back-up), z których jedno (1 x back-up) znajduje się w innej lokalizacji niż siedziba firmy, np. chmurze. Kopie powinny być zaszyfrowane.

6.5. Bezpieczeństwo danych

Jedną z nadrzędnych idei Konkursu jest bezpieczeństwo Danych medycznych pacjentów. RCMC będą musiały w kontekście bezpieczeństwa i przetwarzania danych z badań, w tym klinicznych funkcjonować zgodnie z podstawami prawnymi wypracowanymi przez Unię Europejską. Dokumenty te wskazują na konieczność ich szczególnej ochrony, gdyż kontekst ich przetwarzania może powodować poważne ryzyko dla podstawowych praw i wolności. System RCMC musi zapewniać ściśle kontrolowany dostęp do zasobów systemów, na których dane są przechowywane i przetwarzane, jak również zapewnić maksymalne zabezpieczenie danych poprzez wybrane rozwiązanie. W kontekście cyberbezpieczeństwa dopuszczalne jest przechowywanie danych na serwerach dostawców lub usługodawców spełniających wymogi regulatora, niezależnie od wyboru rozwiązania technicznego (chmura lub serwer tradycyjny).

System RCMC po uruchomieniu produkcyjnym lub po przystąpieniu do Sieci CMC musi spełniać wymagania OWASP ASVS co najmniej na poziomie L1, co powinno zostać potwierdzone raportem z testów penetracyjnych, przeprowadzonych przez niezależną

firmę, specjalizującą się w bezpieczeństwie aplikacyjnym. Rekomendowane jest, aby taki test penetracyjny wykonywać przynajmniej raz w roku.

Ponadto, system RCMC powinien spełniać poniższe wymagania:

- Dane muszą być chronione przed niepowołanym dostępem przy pomocy mechanizmu reglamentacji uprawnień użytkowników.
- Każdy użytkownik (pracownik Uniwersytetu, szpitala klinicznego, Biobanku czy innej jednostki współpracującej w ramach rozwoju RCMC) systemu musi mieć odrębny login i hasło.
- Jakakolwiek funkcjonalność systemu (niezależnie od ilości modułów) będzie dostępna dla użytkownika dopiero po jego zalogowaniu (z wykorzystaniem podwójnego uwierzytelnienia).
- Dostęp do systemu powinien być realizowany tylko z autoryzowanych przez Administratora urządzeń służbowych.
- Umożliwić dodawanie, zablokowanie, usuwanie oraz edycję kont użytkowników.
- Zgodność z rozporządzeniem Rady Ministrów z dnia 12 kwietnia 2012 r. w sprawie Krajowych Ram Interoperacyjności, minimalnych wymagań dla rejestrów publicznych i wymiany informacji w postaci elektronicznej oraz minimalnych wymagań dla systemów teleinformatycznych.
- Dla danych krytycznych wprowadzanych ręcznie, wymagane jest dodatkowe sprawdzenie przy użyciu narzędzi do walidacji zgodnie z zasadami opartymi na analizie ryzyka.
- Anonimizację i Pseudoanimizację gromadzonych danych w zależności od potrzeb jednostki.
- Odpowiedni poziom cyberbezpieczeństwa - zapewnienie bezpieczeństwa przepływu danych zgodnie z obowiązującymi przepisami prawa oraz w zgodzie z aktualnymi rekomendacjami Centrum e-Zdrowia w zakresie budowy systemów cyberbezpieczeństwa oraz z zaleceniami Agencji Unii Europejskiej ds. Cyberbezpieczeństwa ENISA dla sektora medycznego.
- Być zgodny z obowiązującym na dzień uruchomienia systemu stanem prawnym oraz niezwłocznie dostosowującym się do zmian.
- Być zgodny z Narodowymi Standardami Cyberbezpieczeństwa (NSC).
- Musi wylogowywać lub blokować sesję użytkownika po określonym czasie braku aktywności.
- System musi być zgodny z wymaganiami rozporządzenia 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 (ogólne rozporządzenie o ochronie danych- RODO) oraz musi być przygotowany zgodnie z zasadami ochrony danych w fazie projektowania oraz domyślnej ochrony danych (ang. *Privacy by design i Privacy by default*). System musi gwarantować stosowanie zasady minimalizmu danych. System musi przewidywać możliwość korzystania osób ze swoich praw, np. prawo do sprostowania, prawo dostępu do danych, prawa do przenoszenia danych, umieszczenia obowiązków informacyjnych wynikających z art. 13 lub 14 RODO.

- Musi gwarantować rozliczalność prowadzonych działań użytkowników w systemie, odnotowywać wszelką działalność użytkowników - każde wprowadzenie/ zmianę danych oraz pierwotne dane wraz z informacją, kto i kiedy je wprowadził. Żadne dane raz wprowadzone nie mogą zostać nadpisane w sposób trwały powodujący ich nieodwracalne utracenie.
- Musi tworzyć i utrzymywać log systemu przez co najmniej 6 miesięcy, rejestrujący wszystkich użytkowników systemu i wykonane przez nich czynności z możliwością analizy historii zmienianych wartości danych.
- Administrator musi mieć dostęp do Dziennika Zdarzeń, który obejmować będzie rejestr wszystkich logowań, zmian, cofnięć i skreśleń/ usunięcia danych.
- Musi być wyposażony w zabezpieczenia przed nieautoryzowanym dostępem. Zabezpieczenia muszą funkcjonować na poziomie klienta (aplikacja) i serwera (serwer baz danych).
- Musi umożliwić eksport danych lub migracji na nowe środowisko serwerowe w przypadku zaistnienia potrzeby unowocześnienia lub wymiany sprzętu.

Istotnym aspektem bezpieczeństwa udostępniania danych jest dostęp jednostek współtworzących projekt do danych. Dostęp ten powinien być odpowiednio kontrolowany:

- Użytkownicy i jednostki powinni mieć dostęp wyłącznie do tych danych, do których dostępu udzielił im administrator tych danych.
- Każdy zbiór danych przechowywany na serwerach RCMC musi mieć wyznaczonego pracownika, odpowiedzialnego za przechowywane dane.
- Dane mogą być przesyłane pomiędzy jednostkami, pod warunkiem że uzyskana zostanie odpowiednia zgoda komisji bioetycznej.

- Użytkownicy systemu muszą zostać przeszkoleni o zasad bezpieczeństwa obowiązujących w RCMC i zadeklarować, że nie będą przechowywali szczególnych kategorii danych na własnych komputerach.

Zatem konstrukcja systemu musi zakładać możliwość bezpiecznego udostępnienia danych innym podmiotom, ze szczególnym uwzględnieniem podmiotów publicznych. W przypadku jednostek komercyjnych po przesłaniu odpowiedniej aplikacji i uzyskaniu zgody (przyznanej przez kierownika Projektu i odpowiednią komisję bioetyczną), będzie możliwe nadanie dostępu do danych wrażliwych (wyłącznie wewnątrz systemu RCMC).

6.6. Minimalna jakość wyjściowa danych zapewniająca możliwość analizy zagregowanych danych (typy, jakość, format i wielkość plików)

Jednostka musi dbać o wysoką jakość danych w hurtowni danych zarówno na poziomie definicji i architektury danych, jak i ich zawartości. Jakość definicji danych należy rozumieć jako stopień zgodności opisu danych z rzeczywistym pojęciem lub obiektem, tak aby dane były tak samo, jednoznacznie rozumiane przez wszystkich, którzy mają do nich dostęp. Może to być zgodność opisu danych z ogólnie obowiązującymi standardami lub normami. Jakość architektury danych określa jakość modelu danych i baz danych w sposób umożliwiający współdzielony dostęp do nich przez wszystkich uprawnionych odbiorców.

Wysokiej jakości architektura danych zapewnia:

- istnienie właściwych związków (relacji) pomiędzy danymi (spójność danych),
- stabilność modeli danych z punktu widzenia dostępu do nich przez różnych odbiorców danych, niezależnie od zastosowanych aplikacji dostępowych,
- elastyczność modeli danych z punktu widzenia wprowadzanych zmian lub rozbudowy modeli danych.

Jakość zawartości danych określa stopień porównywalności i dokładności, z jakim gromadzone dane odzwierciedlają rzeczywistość i spełniają oczekiwania odbiorców. Miarami jakości zawartości danych są między innymi:

- zgodność zawartości danych z definicją (czy dane opisują rzeczywistość, którą mają opisywać, np. dane nie spełniają tego kryterium, jeśli są niejednoznaczne),
- kompletność danych (czy występują puste wartości w danych),
- zgodność zawartości danych z obowiązującymi regułami (poprawność danych, np. czy są zgodne z dostępnymi danymi wzorcowymi),

- dokładność zawartości danych,
- brak duplikatów danych.

W celu integracji powyższych danych w pierwszej kolejności należy zgromadzić dane w jednym, bezpiecznym i przygotowanym do analizy miejscu (w ramach jednego RCMC). Konieczne jest zapewnienie takiego zarządzania danymi z projektów (Badań naukowych, prac rozwojowych, eksperymentów), które umożliwi transfer danych pomiędzy projektami w sytuacji pozyskania odpowiedniej zgody pacjenta i komisji bioetycznej. Proponowanym, ale nie obowiązkowym rozwiązaniem tego problemu jest przechowywanie danych w oddzielnych maszynach wirtualnych lub instancjach dyskowych, które w razie potrzeby (uzyskania zgody komisji bioetycznej) mogą być zamontowane do innych projektów i połączone ze sobą.

6.7. Wprowadzenie standardu Danych medycznych

Standardy zbieranych danych nie mogą ograniczać wymiany danych. Wytypowanym przez ABM zbiorem tych standardów są zasady FAIR (*Findability, Accessibility, Interoperability, and Reuse*)⁶. W RCMC powinny być składowane zdarzenia medyczne i wytworzone EDM, przynajmniej w zakresie określonym w rozporządzeniu Ministra Zdrowia z dnia 26 czerwca 2020 r. w sprawie szczegółowego zakresu danych zdarzenia medycznego przetwarzanego w systemie informacji oraz sposobu i terminów przekazywania tych danych do Systemu Informacji Medycznej z późniejszymi zmianami.

RCMC powinno stosować dokładnie taką samą indeksację zasobów zdarzenia medycznego, jak w źródłowej bazie P1 zgodnie z ustalonymi standardami w P1. Rozszerzenie zakresu informacyjnego względem zakresu określonego w przepisach prawa powinno zostać ustandaryzowane RCMC w ramach współpracy pomiędzy jednostkami RCMC. Kolejne atrybuty dla danych zdarzeń pacjentów, powinny być przechowywane w logice zgodnej ze standardem HL7 CDA. Dla wszystkich centrów powinien obowiązywać jeden standard. Zapewni to możliwość interoperacyjności systemów RCMC, a w dalszej kolejności stworzy potencjał interoperacyjności z całym system e-zdrowia w Polsce i Europie. Docelowa standaryzacja danych zostanie przeprowadzona w ramach Centrali CMC ABM.

⁶ <https://www.go-fair.org/fair-principles> (data dostępu 12.01.2023r.)

5.7.2 Dane obrazowe

Dane obrazowe należy zapisywać w formacie DICOM i przechowywane na serwerach lub w chmurze, połączonych z infrastrukturą szpitalną służącą badaniom obrazowym: diagnostyce, radiologii, histopatologii i innym. System RCMC powinien posiadać systemy Anonimizujące i Psedudonimizujące obrazy, zależnie od konkretnych potrzeb i wymagań prawnych.

Dane powinny koniecznie zawierać następujące informacje:

- pozyskany obraz w formacie DICOM (wymóg zgodny ze Standardami e-Zdrowia)⁷,
- odniesienie do jednoznacznego identyfikatora pacjenta, dla którego wykonano badanie,
- rodzaj lub typ wykonanego badania,
- informację o wykorzystanym standardzie (krajowym lub europejskim), według którego wykonano badanie,
- informację czy dane są surowe czy przetworzone (jeżeli przetworzone, to w jaki sposób),
- diagnozę.

6.7.1. Dane sekwencyjne

Dane sekwencyjne zapisane w formatach tekstowych (skompresowanych lub nie) często zawierają warianty genetyczne, które są specyficzne dla pacjenta i wrażliwe. Są to dane o dużej objętości i wymagają wydajnych serwerów obliczeniowych na etapach kontroli jakości i przetwarzania. Wnioskodawca będzie musiał wykazać zapewnienie zdolności do zarówno zabezpieczenia, jak i przetwarzania danych.

Dane powinny koniecznie zawierać następujące informacje:

- pozyskane surowe sekwencje w formacie fastq,
- odniesienie do jednoznacznego identyfikatora pacjenta, dla którego wykonano badanie,
- rodzaj lub typ wykonanego badania,
- informacje o sekwenatorze którym wykonano badanie,
- informacje kontaktowe osoby odpowiedzialnej za badanie.

⁷ <https://www.gov.pl/web/ia/standardy-e-zdrowia> (dostęp z dnia 10.02.2023)

Dodatkowo, powinna istnieć możliwość przechowywania danych przetworzonych, pod warunkiem, że protokół, według którego zostały one utworzone będzie odpowiednio opisany i będzie istniała możliwość ponownego wykorzystania tego protokołu.

VCF jest formatem pliku tekstowego (zapisanego w sposób skompresowany). Zawiera linie meta-informacji, linię nagłówka, a następnie linie danych, z których każda zawiera informacje o pozycji w genomie.

Jakość plików *vcf (*Variant Call Format*) powinna być określana zgodnie z aktualną specyfikacją VCF/BCF. Grupą prowadzącą zarządzanie i rozszerzanie formatu jest zespół formatu plików grupy roboczej Global Alliance for Genomics and Health Data. Główną wersję specyfikacji (oraz jej aktualizacje) można znaleźć na stronie <https://github.com/samtools/hts-specs>

Ostatnia wersja specyfikacji to v4.3. Głównym miernikiem jakości sekwencji w formacie vcf jest wartość QUAL

Jako zasób danych referencyjnych można traktować bazę The International Genome Sample Resource <https://www.internationalgenome.org/>

6.7.2. Inne dane

Inne dane laboratoryjne, metabolomiczne, proteomiczne, i inne omiczne, zazwyczaj w formatach tekstowych tabel, ale w niektórych przypadkach w innych formatach specyficznych dla producentów sprzętu medycznego.

Dane powinny zawierać koniecznie następujące informacje:

- informacje o typie pozyskanych danych,
- standardzie krajowym lub europejskim, według którego wykonano badanie,
- odniesienie do jednoznacznego identyfikatora pacjenta.

6.8. Udostępnianie danych i ich wymiana

Ze względu na rosnącą liczbę zagrożeń dotyczących szczególnych kategorii danych, zdecydowano się na wprowadzanie rozwiązania federacyjnego z dopuszczalnym rozwiązaniem centralnym w przyszłości. Rozwiązanie takie polega na przechowywaniu danych w jednostce która dane pozyskała i posiada prawne umocowanie do ich przechowywania i przetwarzania, zgodne z RODO oraz krajowymi ustawami dotyczącymi szczególnych kategorii danych.

Dodatkowym elementem systemu federacyjnego jest wprowadzenie relacji i standardów wymiany danych pomiędzy jednostkami zbierającymi dane i jednostką centralną. W przypadku Polski jednostką Centralną jest Centrum e-Zdrowie. Dodatkowo jako jednostka Centralna będzie rozpatrywana Centrala Medycyny Cyfrowej ABM (CMC ABM) po jej uruchomieniu i prawnym umocowaniu. W strukturze federacyjnej jednostki zbierające dane informują centralę o posiadaniu konkretnych typów danych (na przykład z badań klinicznych, obrazowych czy sekwencyjnych), ale nie przesyłają całości tych danych na do centrali.

Należy podkreślić, że technicznie możliwość przesyłania danych istnieje, lecz jest ona wykorzystywana wyłącznie w uzasadnionych przypadkach i przy uzyskaniu odpowiednich zgód na poziomie prawnym i bioetycznym. Możliwe jest stworzenie sieci komunikacji opartej o odpowiednio zabezpieczony i szyfrowany protokół SSH-2, która pozwoli na połączenie i przesyłanie danych. Agencja rekomenduje zastosowanie standardu FHIR - Fast Healthcare Interoperability Resources. FHIR może być zastosowany do przekazywania danych gromadzonych w systemach medycznych w zakresie objętym standardem.

6.9. Aspekty etyczne i prawne

RCMC powinno funkcjonować zgodnie z zasadami etycznymi oraz przepisami prawa krajowego i międzynarodowego dotyczącymi ludzkiego materiału biologicznego i danych obowiązującymi w Polsce w tym poszanowania godności oraz podstawowych praw wolności uczestników badań, zachowania bezstronności w zakresie organizacji, zarządzania oraz analizy ryzyka.

RCMC powinno dokumentować swoją działalność poprzez archiwizację dokumentacji takiej jak pozwolenia, decyzje, opinie i zgłoszenia. Rekomenduje się także określenie czasu przechowywania dokumentacji i związanych z nią danych. Należy określić relacje pomiędzy RCMC a podmiotem, którego jest częścią. Informacje na temat przyjętych zasad postępowania w ważnych z punktu widzenia uczestnika badań kwestiach etycznych i prawnych, np. zasady transferu próbek i danych do innych podmiotów, umieszcza się na stronach internetowych jednostek zaangażowanych w rozwój RCMC z odwołaniem do konkretnych dokumentów. Informacje przeznaczone dla uczestników badań powinny być napisane językiem prostym i zrozumiałym dla przeciętnego odbiorcy.

Zgodnie z art. 35 RODO jeżeli dla przetwarzanych danych osobowych istnieje wysokie prawdopodobieństwo wystąpienia ryzyka naruszenia praw lub wolności osób fizycznych, przed rozpoczęciem przetwarzaniem należy przeprowadzić ocenę skutków dla ochrony danych (*DPiA - Data Protection Impact Assessment*), czyli zidentyfikować zagrożenia, które mogą naruszyć dane osobowe oraz zastosować odpowiednie środki zapobiegawcze i mechanizmy bezpieczeństwa, które nie dopuszczają lub zminimalizują ryzyko naruszenia tych danych.

7. OCZEKIWANE REZULTATY I KORZYŚCI

1. Wytworzenie API (ang. Application Programming Interface) w celu dalszego rozwoju systemu - integracja z systemem centralnym poprzez udostępnienie API.
2. Otwartość systemu. System musi mieć możliwość rozbudowy o dodatkowe moduły.
3. System musi umożliwić eksport danych lub migracji na nowe środowisko serwerowe w przypadku zaistnienia potrzeby unowocześnienia lub wymiany sprzętu.
4. Utworzony, rozwinięty lub doposażony Biobank posiada certyfikaty zgodne ze standardem Podręcznika „*Standardy jakości dla Biobanków polskich*” v.2.0.
5. Pomoc w projektowaniu badania (m.in. obliczaniu mocy badania, implementacji narzędzi cyfrowych) – polepszenie jakości badania na etapie projektowym (ang. *quality by design*).
6. Stworzenie bazy danych klinicznych i omicznych stanowiącej podstawę do generowania cyfrowych narzędzi o charakterze prognostycznych, predykcyjnych i terapeutycznych algorytmów sztucznej inteligencji.
7. Analiza danych w czasie rzeczywistym – poprawa bezpieczeństwa pacjentów oraz jakości uzyskanych wyników.
8. Analiza danych po ukończeniu Badania naukowego – nadzór nad jakością danych od początku do końca badania i uzyskanie wiarygodnych wyników badań o dużym potencjale publikacyjnym.
9. Magazynowanie danych z Badań klinicznych realizowanych w ramach umów z ABM i nadzór nad ich integralnością i bezpieczeństwem we współpracy z ABM.
10. Praca z danymi w oparciu o międzynarodowe standardy oraz wypracowywanie dobrych praktyk we współpracy z ABM.
11. Ujednolicenie standardów analiz i jakości uzyskiwanych danych.
12. Standaryzacja danych z różnych HIS.
13. Standaryzację danych z różnych Biobanków.
14. Wypracowanie rozwiązań dotyczących Anonimizacji / Pseudanonimizacji, integracji, zabezpieczania i transferu danych zgodnie z obowiązującymi standardami, ustawodawstwem i we współpracy z ABM.
15. Docelowe transferowanie danych w ramach Sieci, w tym do Centrali Medycyny Cyfrowej ABM, celem realizacji analiz wtórnego.
16. Opracowanie standardów udostępniania danych zewnętrznym podmiotom celem poprawy dostępności do danych w ścisłej współpracy i pod przewodnictwem ABM. Standard zostanie wdrożony po zmianach legislacyjnych.
17. Opracowanie infrastruktury i zasad analizy danych - próbek biologicznych pochodzących z Badań klinicznych, Biobanków i od Dawców-wolontariuszy.

18. Promowanie realizowanego programu medycyny cyfrowej za granicą i nawiązanie współpracy międzynarodowej w zakresie standaryzacji i analiz danych.

Załączniki:

1. Formularz Świadomej Zgody