



PLAN ZARZĄDZANIA DANYMI W PROJEKCIE BADAWCZYM

WERSJA ROZSZERZONA

Plan Zarządzania Danymi w projekcie badawczym. Wersja rozszerzona

Konkurs: Sonata 16

Obszar badawczy: HS Nauki humanistyczne, Społeczne i o Sztuce

Panel dyscyplin: HS2 Kultura i twórczość kulturowa

Pomocnicze określenia: HS2_004 Bibliologia i informatologia;

HS3_012 Dziedzictwo kulturowe

(w tym: inwentaryzacja pamiątek i zabytków kultury, aktywność regionalistyczna)

Opracowanie: dr Anna Bujko, mgr Jolanta Gątecka

Biblioteka Uniwersytecka

Uniwersytetu Warmińsko-Mazurskiego w Olsztynie

ul. Oczapowskiego 12B, 10-719 Olsztyn

tel. 89 524 51 17, 89 523 49 97

e-mail: anna.bujko@uwm.edu.pl, joga@uwm.edu.pl



Licencja Creativ Commons ZERO (CC0) 1.0 Universal

1. Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych

1.1. Sposób pozyskiwania i opracowania nowych danych i/lub ponownego wykorzystania dostępnych danych

• Jakie standardy, metody i oprogramowanie posłużą do pozyskiwania lub wytwarzania nowych danych?

W drodze **kwerendy bibliotecznej i archiwalnej** zostaną pozyskane następujące dane badawcze:

1. przeznaczone do udostępnienia (dane opracowane):

- a) baza danych z plikami graficznymi po konwersji, zdeponowanymi w systemie plików serwera systemu bibliotecznego ALEPH oraz na serwerze dedykowanym Wykonawcy projektu;
- b) zbiór danych obejmujący zawartość bazy danych przekonwertowany z formatu .xls do formatu .pdf udostępniony w repozytorium otwartych danych badawczych w formule Open Access.

2. nie przeznaczone do udostępnienia:

- a) pliki graficzne:
 - macierzyste;
 - po konwersji;
- b) pliki (kopie) cyfrowe dokumentów bibliotecznych i archiwalnych oraz opracowań naukowych, materiałów prasowych, wspomnień itp.;
- c) kopie w formie papierowej dokumentów bibliotecznych i archiwalnych oraz opracowań naukowych, materiałów prasowych, wspomnień itp.

Dane badawcze w postaci **plików graficznych po konwersji** zostaną zarejestrowane w **bazie danych** i będą stanowiły element składowy **rekordu bibliograficznego** jako pole/podpola rekordu opisujące zawartość i lokalizację pliku graficznego. Do budowy bazy danych zostanie wykorzystany zintegrowany system biblioteczny ALEPH: a) moduł katalogowania w formacie MARC 21 oraz b) moduł ADAM, służący do zarządzania zasobami cyfrowymi. Rekordy zostaną stworzone zgodnie z obowiązującymi zasadami katalogowania. System biblioteczny ALEPH gwarantuje wysoką jakość i spójność dzięki możliwości weryfikacji i walidacji danych na poziomie rekordów i pól/podpól, rozbudowany system raportowania (ujawnienie rekordów zawierających błędy) oraz możliwość globalnych zmian pól/podpól/słów obejmujących całą bazę.

Dane badawcze w postaci **plików (kopii) cyfrowych** dokumentów bibliotecznych i archiwalnych oraz opracowań naukowych, materiałów prasowych, wspomnień itp. zostaną zebrane lub/i wytworzone zgodnie z obowiązującymi standardami digitalizacji, opisu bibliograficznego oraz opisu metadanych.

• Jakie dane już istniejące (własne lub będące w posiadaniu strony trzeciej) zostaną ponownie wykorzystane?

Pliki cyfrowe dokumentów bibliotecznych i archiwalnych oraz opracowań naukowych, materiałów prasowych, wspomnień itp. dostępne w bibliotekach i archiwach cyfrowych.

• W jaki sposób zostanie udokumentowane pochodzenie danych?

W rekordach bazy danych opisujących pliki graficzne (skany) zostanie dodane pole o aktualnym właścicielu egzemplarza. Pochodzenie plików (kopii) cyfrowych dokumentów bibliotecznych i archiwalnych oraz opracowań naukowych, materiałów prasowych, wspomnień itp. zostanie udokumentowane **zgodnie z zasadami opisu bibliograficznego**. Przy plikach cyfrowych zgromadzonych w drodze kwerendy w archiwach i bibliotekach cyfrowych zostanie dodany **link** przekierowujący do **miejsca zdeponowania (pochodzenia) pliku**.

• Jak będzie wyglądać organizacja plików i zarządzanie różnymi wersjami?

Dane badawcze w formie plików graficznych macierzystych i po konwersji oraz plików (kopii) cyfrowych dokumentów bibliotecznych i archiwalnych oraz opracowań naukowych, materiałów prasowych, wspomnień itp. będą usystematyzowane i uporządkowane w poszczególnych folderach i opatrzone odpowiednim identyfikatorem (nazwą). Na nazwę pliku będą składały się następujące elementy:

- 1) **akronim** stanowiący nazwę obecnego właściciela egzemplarza książki, czasopisma, dokumentu, opracowania itd.;
- 2) aktualna **sygnatura**;
- 3) **numer karty** (w przypadku plików cyfrowych materiałów źródłowych).

Przykład:

- dla dokumentu archiwalnego: AAN, MO, 37/207, k. 20; APO, 53/892/0/2/399, k.30
- dla egzemplarza książki: AP, 3067; BUUWM, 6430 itd.

Dane badawcze w formie rekordów bazy danych będą oznaczone przypisanym na stałe unikalnym **numerem systemowym** w ramach bazy logicznej i uporządkowane według zdefiniowanych kryteriów wyszukiwawczych, w zależności od potrzeb użytkownika. Dane badawcze w formie kopii w tradycyjnej formie papierowej zostaną zorganizowane **w porządku alfabetycznym zgodnie z zasadami opisu bibliograficznego**.

Mechanizm **wersjonowania** będzie zastosowany do **zbioru danych**, obejmującego zawartość bazy danych, który będzie udostępniony w otwartym repozytorium danych badawczych. Wersje zbioru danych – o ile powstaną – będą zdeponowane w tym samym rekordzie, w którym znajduje się wersja podstawowa (numer DOI będzie posiadał zbiór danych niezależnie od zdeponowanych wersji). Zachowany zostanie porządek numeryczny wersji (1, 2, 3, ... n).

• Czy konieczna będzie digitalizacja danych analogowych lub wydanych w formie papierowej?

Projekt zakłada digitalizację około 31000 ... znajdujących się w 26000 **egzemplarzach wydawnictw zwartych i ciągłych**. Digitalizacji będą również wymagały zdobyte w wyniku kwerendy **dokumenty biblioteczne i archiwalne oraz opracowania naukowe, materiały prasowe, wspomnienia itp.**

1.2. Pozyskiwane lub opracowane dane (rodzaj, format, ilość)

• Jaki rodzaj, format i objętość danych planują Państwo pozyskać, wytworzyć lub ponownie wykorzystać?

Dane badawcze będzie stanowiło:

- 1) około **31000 plików graficznych macierzystych w formacie .tif** w rozdzielczości **300 DPI**, skany kolorowe 1:1;
- 2) około **31000 plików graficznych po konwersji w formacie .pdf**;
- 3) **26000 kompletnych rekordów bazy danych** zawierających: a) dane o egzemplarzu, rodzaju występujących ...; b) pola i podpola tworzone w oparciu o listę podpowiedzi nazw ... (hasła osobowe w odniesieniu do właścicieli indywidualnych oraz korporatywne w odniesieniu do właścicieli instytucjonalnych); c) około 31000 plików graficznych prezentujących opisywane ... w formacie .pdf.;
- 3) **pliki (kopie) cyfrowe** dokumentów bibliotecznych i archiwalnych oraz opracowań naukowych, materiałów prasowych, wspomnień itp. (dokładny format cyfrowy i objętość będzie możliwa do określenia dopiero po odbytej kwerendzie);
- 4) **kopie w formie papierowej** dokumentów bibliotecznych i archiwalnych oraz opracowań naukowych, materiałów prasowych, wspomnień itp. (ilość będzie możliwa do określenia dopiero po odbytej kwerendzie).

2. Dokumentacja i jakość danych

2.1. Metadane i dokumenty

(np. metodologia lub pozyskanie danych oraz sposób porządkowania danych)

Podstawą pozyskania danych będzie kwerenda archiwalna i biblioteczna. Elementy graficzne przedstawiające ... zostaną zeskanowane. Dane badawcze w postaci plików przetworzonych do formatu .pdf na podstawie macierzystych plików graficznych w formacie .tif (oba rodzaje plików zostaną zarchiwizowane na dedykowanym serwerze Biblioteki Uniwersyteckiej UWM w uporządkowanych folderach według typu plików i numeru systemowego utworzonego rekordu), zostaną opracowane i uporządkowane dzięki systemowi ALEPH moduł zarządzania zasobami cyfrowymi ADAM, umożliwiającemu ich załączenie do tworzonych rekordów w module katalogowania w formacie MARC 21. Dodatkowo zostaną opisane metadanymi technicznymi (nazwa, wielkość, typ, format, lokalizacja pliku).

• Jakie informacje należy uzupełnić, aby potencjalni użytkownicy (komputery lub ludzie) byli w stanie w przyszłości odczytać i zinterpretować zebrane dane?

Odczytanie i interpretację danych udostępnionych w formacie Open Access w formie:

1) **bazy danych** stworzonej w oparciu o system informatyczny ALEPH zapewnią:

a) **indeks** właścicieli ...;

b) **metadane** opisujące pliki cyfrowe po konwersji (np. 1. Nazwa obecnego właściciela egzemplarza, 2. Opis bibliograficzny: autor, tytuł egzemplarza, miejsce rok wydania; 3. Sygnatura; 4. Informacje o instytucji finansującej badania, 5. Typ pliku, 6. format pliku, 7. Licencja itd.);

2) **zbioru danych** zdeponowanego w otwartym repozytorium zapewni **format .csv** lub/i **.xls** lub/i **.pdf**, w jakim zostanie zapisany plik;

3) **metadanych** zapewni zastosowany format zgodnie ze standardami **OpenAIRE**.

CSV

• Czy możliwe jest ich komputerowe odczytanie?

Będzie możliwość komputerowego **odczytania metadanych** opisujących zbiór danych zamieszczony w repozytorium. Maszynowe odczytanie zdeponowanego zbioru danych będzie zależało od formatu, w jakim plik zostanie zapisany (jeżeli obecnie plik będzie zapisany w formacie uniemożliwiającym maszynowe odczytanie danych, przewiduje się w przyszłości konwersję pliku do formatu umożliwiającego komputerowy odczyt i interpretację danych).

• W jaki sposób powstanie odpowiednia dokumentacja?

Podstawę dokumentacji danych będzie stanowiła **registratura danych badawczych** obejmująca wszystkie dane badawcze, stworzona w formacie .xls w porządku według schematów podanych poniżej. Zapisy w registraturze będą miały swoje odzwierciedlenie w folderach, w których będą znajdowały się wszystkie cyfrowe dane badawcze (dla danych w tradycyjnej formie papierowej odpowiednikiem folderów będą tecki):

Schemat:

a) dla opracowań i materiałów źródłowych opublikowanych:

XXX, YYY, ZZZ, gdzie: XXX to akronim instytucji, YYY to sygnatura egzemplarza lub dokumentu archiwalnego i bibliotecznego;

b) dla materiałów źródłowych nieopublikowanych:

XXX, YYY, ZZZ, gdzie: XXX to akronim instytucji, YYY to sygnatura egzemplarza lub dokumentu archiwalnego i bibliotecznego, ZZZ to numer karty.

1. Pliki macierzyste i po konwersji:

1.1 pliki macierzyste (ZWM – Znaki Własnościowe Macierzyste):

- 1.1.a) instytucja A: ZWM XXX, YYY; ZWM XXX, YYY; ZWM XXX, YYY...;
- 1.1.b) instytucja B: ZWM XXX, YYY; ZWM XXX, YYY; ZWM XXX, YYY...;
- 1.1.c) instytucja C: ZWM XXX, YYY; ZWM XXX, YYY; ZWM XXX, YYY...;
- 1.1.d) instytucja D: ZWM XXX, YYY; ZWM XXX, YYY; ZWM XXX, YYY...;
- 1.1.e) instytucja E: ZWM XXX, YYY; ZWM XXX, YYY; ZWM XXX, YYY...;

1.2. pliki po konwersji (ZWK – Znaki Własnościowe po Konwersji):

- 1.2.a) instytucja A: ZWK XXX, YYY; ZWK XXX, YYY; ZWK XXX, YYY...;
- 1.2.b) instytucja B: ZWK XXX, YYY; ZWK XXX, YYY; ZWK XXX, YYY...;
- 1.2.c) instytucja C: ZWK XXX, YYY; ZWK XXX, YYY; ZWK XXX, YYY...;
- 1.2.d) instytucja D: ZWK XXX, YYY; ZWK XXX, YYY; ZWK XXX, YYY...;
- 1.2.e) instytucja E: ZWK XXX, YYY; ZWK XXX, YYY; ZWK XXX, YYY...;

Przykład: BUUWM [Biblioteka Uniwersytecka UWM w Olsztynie], 6430 [sygnatura]

2. Dokumenty archiwalne i biblioteczne (MZN – Materiały Źródłowe Niepublikowane):

- 2.1. instytucja A: MZN XXX, YYY, ZZZ; MZN XXX, YYY, ZZZ; MZN XXX, YYY, ZZZ;
- 2.2. instytucja B: MZN XXX, YYY, ZZZ; MZN XXX, YYY, ZZZ; MZN XXX, YYY, ZZZ;
- 2.3. instytucja C: MZN XXX, YYY, ZZZ; MZN XXX, YYY, ZZZ; MZN XXX, YYY, ZZZ;
- 2.4. instytucja D: MZN XXX, YYY, ZZZ; MZN XXX, YYY, ZZZ; MZN XXX, YYY, ZZZ;
- 2.5. instytucja E: MZN XXX, YYY, ZZZ; MZN XXX, YYY, ZZZ; MZN XXX, YYY, ZZZ;

Przykład: AAN [Archiwum Akt Nowych], MO 283/3072 [sygnatura], k. 11 [karta]

3. Opracowania naukowe, materiały prasowe, wspomnienia itp. (MZP – Materiały Źródłowe Publikowane):

- 3.1. instytucja A: MZP XXX, YYY, ZZZ; MZP XXX, YYY, ZZZ; MZP XXX, YYY, ZZZ;
- 3.2. instytucja B: MZP XXX, YYY, ZZZ; MZP XXX, YYY, ZZZ; MZP XXX, YYY, ZZZ;
- 3.3. instytucja C: MZP XXX, YYY, ZZZ; MZP XXX, YYY, ZZZ; MZP XXX, YYY, ZZZ;
- 3.4. instytucja D: MZP XXX, YYY, ZZZ; MZP XXX, YYY, ZZZ; MZP XXX, YYY, ZZZ;
- 3.5. instytucja E: MZP XXX, YYY, ZZZ; MZP XXX, YYY, ZZZ; MZP XXX, YYY, ZZZ;

Przykład: BUUWM [Biblioteka Uniwersytecka UWM w Olsztynie], 6430 [sygnatura]

4. Rekordy w bazie danych uporządkowane według numeru systemowego;

5. Zbiór danych ze wszystkimi wersjami:

- 5.1. v 1 [wersja 1];
- 5.2. v 2;
- 5.3. v 3;
- 5....
- 5.n. v n.

W działach:

- a) o numerze 1 w registraturze i folderach zostanie zachowany ten sam **porządek numeryczny** według numerów sygnatur;
- b) o numerze 2 i 3 w registraturze zostanie zachowany **porządek alfabetyczny** zgodnie z zasadami opisu bibliograficznego dla opracowań, materiałów źródłowych opublikowanych i materiałów źródłowych niepublikowanych, a w folderach 2 i 3 – **porządek numeryczny** według numerów sygnatur.

• Jakie standardy środowiskowe (jeżeli w ogóle) posłużą do objaśnienia (meta)danych?

Do budowy bazy danych z plikami graficznymi po konwersji zostanie wykorzystany **system biblioteczny ALEPH**.

Dane badawcze:

- 1) w postaci rekordów będą stworzone zgodnie ze standardem opisu bibliograficznego w formacie MARC 21, który jest powszechnie obowiązującym formatem katalogowania danych bibliograficznych w formie elektronicznej na całym świecie;
- 2) w postaci plików graficznych po konwersji przechowywanych na serwerze dedykowanym lub w systemie plików serwera ALEPH będą dostępne poprzez link zamieszczony w jednym z pól rekordu bibliograficznego, stanowiący element metadanych opisanych w **formacie MARC 21**;
- 3) mogą być przekonwertowane **z formatu MARC 21 do formatu Dublin Core**.

• Jakie międzynarodowe standardy lub schematy (tj. Dublin Core, DDI) posłużą do organizacji metadanych?

Dane badawcze będą opatrzone metadanymi stworzonymi zgodnie z obowiązującymi standardami, w zależności od miejsca ich przechowywania i udostępniania:

- a) **pliki graficzne po konwersji** będą opisane (scharakteryzowane) w systemie bibliotecznym ALEPH w standardzie MARC 21, umożliwiającym dostęp do metadanych zgodnych z elementami wymaganymi w formacie Dublin Core;
- b) **zbiór danych badawczych** w formie pliku .csv lub/i .xls lub/i .pdf zdeponowany w otwartym repozytorium cyfrowym będzie opisany metadanymi **zgodnie ze standardami przyjętymi dla danego repozytorium** umożliwiającymi eksport metadanych według różnych standardów (OpenAIRE, DDI, JSON, DataCite, Dublin Core, JSON-LD Schema.org, OAI_ORE).

2.2. Stosowane środki kontroli jakości danych

• W jaki sposób metody pozyskiwania, analizy i przetwarzania danych mogą wpływać na ich jakość?

nie dotyczy

• W jaki sposób można wyeliminować błędy pomiarowe i problem stronniczości?

nie dotyczy

• W jaki sposób zminimalizować ryzyko dotyczące poprawności danych?

Dane zostaną wytworzone przez wykwalifikowanych wykonawców na urządzeniach posiadających odpowiednie oprogramowanie. Ich wysoką jakość zapewni:

- 1) kontrola i korekta utworzonych rekordów oraz wprowadzanie ewentualnych uzupełnień do rekordów, kontrola merytoryczna wprowadzanych nazw właścicieli ... w oparciu o indeks utworzony przez zespół badawczy;
- 2) **kontrola jakości danych** poprzez sprawdzenie wytworzonych rekordów pod względem poprawności formalnej, tzn. zgodności zawartości pól, podpól i wskaźników z przyjętym formatem, bieżące monitorowanie zgodności wszystkich operacji w systemie komputerowym (poprawność pól opisu bibliograficznego, poprawność załączonych plików graficznych), korektę w zakresie

poprawności opisu w systemie bibliotecznym, kontrolowanie prac związanych z modulem opracowania zbiorów systemu komputerowego;
3) **kontrola poprawności zapisu plików graficznych** (poprawność formatu, właściwa rozdzielczość, poprawność wyświetlania grafiki itp.);
4) **nadzór merytoryczny** zespołu badawczego.
Przed nieuprawnioną modyfikacją w rekordach będzie chronił **system uwierzytelniania** (logowanie do bazy danych, posiadanie adekwatnych uprawnień).

3. Przechowywanie i tworzenie kopii zapasowych

3.1. Przechowywanie i tworzenie kopii zapasowych danych i metadanych podczas badań

• Jaką dysponują Państwo pojemnością pamięci i gdzie będą przechowywane dane?

Dane badawcze: kompletne rekordy w bazie danych zarchiwizowane na **serwerze Zintegrowanego Systemu Bibliotecznego ALEPH**, pliki graficzne macierzyste w formacie .tif oraz pliki po konwersji w formacie .pdf w formie elektronicznej **na urządzeniach pamięci masowej** sprzętu wykorzystywanego w fazie realizacji projektu (5 komputerów) i jednocześnie na **dedykowanym serwerze Biblioteki Uniwersyteckiej**. Bezpieczeństwo danych, ciągłość funkcjonowania serwerów i stacji roboczych, ciągłość funkcjonowania systemu bibliotecznego ALEPH zapewni zespół Oddziału ds. Komputeryzacji jednostki, w której będzie realizowany projekt.

• Jak wyglądają procedury tworzenia kopii zapasowych?

Kopie zapasowe zostaną utworzone **automatycznie po utworzeniu danych**. Dodatkowo kopie zapasowe danych i metadanych będą wykonywane **regularnie pod kontrolą Oddziału ds. Komputeryzacji** jednostki, w której będzie realizowany projekt, kopie przyrostowe codziennie, kopia pełna nie rzadziej niż raz na tydzień. Zgodnie z obowiązującą umową kopie zapasowe wytworzonych danych w systemie bibliotecznym ALEPH będą kontrolowane przez **firmę realizującą ciągłą obsługę serwisową** (ALEPH Polska).

• Czy potrzebne będą specjalne metody przenoszenia danych z urządzeń mobilnych, stanowisk terenowych lub sprzętu domowego na główny serwer w miejscu pracy?

Migracji ze stanowisk terenowych i mobilnych będą wymagały dane badawcze w postaci plików graficznych macierzystych i plików graficznych po konwersji. Do tego celu zostanie wykorzystany **system chmurowy** który umożliwi przesył, pobranie i zapisanie plików na serwerze głównym.

3.2. Sposób zapewnienia bezpieczeństwa danych oraz ochrony danych wrażliwych podczas badań

• W jaki sposób można będzie odzyskać dane utracone w wyniku incydentu?

Dane zebrane w wyniku realizacji projektu zostaną zapisane na serwerze dedykowanym jednostki realizującej projekt. Dla wszystkich danych zostaną stworzone **kopie zapasowe**, które będą przechowywane na oddzielnej lokalizacji. Mogą wystąpić p o j e d y n c z e przypadki utraty danych podczas tworzenia rekordu w bazie danych lub skanowania grafiki z powodu błędu ludzkiego, co będzie wymagało **powtórzenia czynności stworzenia rekordu lub wykonania skanu**. Te incydentalne przypadki – o ile w ogóle wystąpią – nie wpłyną na realizację i zakończenie projektu.

• Kto uzyska dostęp do danych w czasie trwania projektu i jak wyglądać będzie kontrola dostępu do danych, zwłaszcza w przypadku współpracy między kilkoma partnerami?

Dostęp do urządzeń, systemów służących do wytworzenia danych oraz do danych będzie kontrolowany na poziomie wykonawcy przez **systemy uwierzytelniania** (ALEPH, Active Directory). Adekwatny dostęp do danych będą mieli **wszyscy uprawnieni wykonawcy projektu** po uprzednim **zalogowaniu** na urządzeniu, a następnie do systemu bibliotecznego ALEPH (w przypadku bazy

elektronicznej) lub po zalogowaniu na urządzeniu (do zdigitalizowanych plików graficznych znajdujących się w uporządkowanych folderach umożliwiających identyfikację pliku z egzemplarzem książki lub czasopisma w zasobach współdzielonych, dostępnych dla wszystkich wykonawców projektu). **Dane wrażliwe (login i hasło)** będą chronione hasłem w systemach uwierzytelniania i systemie ALEPH, które będą zarządzane przez uprawnionych administratorów zgodnie z zasadami polityki bezpieczeństwa. Hasła będą dystrybuowane przy użyciu bezpiecznych **kanałów komunikacji**.

4. Wymogi prawne, kodeks postępowania

4.1. Sposób zapewnienia zgodności z przepisami dotyczącymi danych osobowych i bezpieczeństwa danych w przypadku przetwarzania danych osobowych

Na Uniwersytecie Warmińsko-Mazurskim w Olsztynie **dane osobowe są chronione na wszystkich poziomach**. Istniejące procedury zostaną zastosowane w projekcie w odniesieniu do wykonawców zatrudnionych w ramach projektu. W razie wątpliwości kierownik projektu zwróci się do Inspektora Ochrony Danych, zatrudnionego w jednostce, który posiada wszelkie kompetencje do udzielenia porad i informacji (<http://www.uwm.edu.pl/daneosobowe/inspektor-ochrony-danych>).

• Czy w procesie pozyskiwania danych niezbędna będzie ich anonimizacja?

nie dotyczy

• Czy należy usunąć informacje umożliwiające identyfikację osób lub zataić tożsamość uczestników (np. za pomocą pseudoanimacji) przed udostępnieniem danych?

nie dotyczy

4.2. Sposób zarządzania innymi kwestiami prawnymi, np. prawami własności intelektualnej lub własnością. Obowiązujące przepisy

• Kto będzie właścicielem danych?

Prawa autorskie do bazy danych i do plików graficznych będą należeć do **Wnioskodawcy**.

• Jakie zostaną zastosowane licencje?

Dane badawcze zostaną udostępnione na następujących licencjach:

- 1) baza danych z plikami graficznymi po konwersji na licencji **Creativ Commons ZERO (CC0) 1.0 Universal (Public Domain)**;
- 2) zbiór danych zdeponowany w repozytorium danych badawczych zostanie udostępniony na licencji **Creativ Commons Uznanie autorstwa 4.0**.

• Czy przed udostępnieniem danych konieczne będzie uzyskanie odpowiedniego zezwolenia w zakresie praw autorskich?

nie dotyczy

5. Udostępnienie i długotrwałe przechowywanie danych

5.1. Sposób i termin udostępnienie danych. Ewentualne ograniczenia w udostępnianiu danych lub przyczyny embarga

• W jaki sposób o Państwa danych dowiedzą się potencjalni użytkownicy?

Dwa rodzaje danych badawczych zostaną udostępnione w formule Open Access dla wszystkich zainteresowanych użytkowników bez ograniczeń:

- 1) w postaci **rekordów w bazie danych** będą dostępne pierwszego dnia po terminie zakończenia w **serwisie internetowym Biblioteki Uniwersyteckiej poprzez WWW OPAC** w formie odrębnej bazy logicznej (https://bart.uwm.edu.pl/F?func=file&file_name=base-list).
- 2) w postaci **zbioru danych badawczych** zostaną udostępnione pierwszego dnia po zakończeniu projektu w **repozytorium danych badawczych znajdującym się w bazie <https://www.re3data.org/>, stanowiącej globalny rejestr repozytoriów danych badawczych;**
- 3) w postaci **metadanych opisujących zbiór danych** zostaną udostępnione pierwszego dnia po zakończeniu projektu w repozytorium;
- 4) w postaci **metadanych opisujących pliki graficzne po konwersji** zamieszczone w rekordach w bazie danych pierwszego dnia po zakończeniu projektu.

Dodatkowo formą rozpowszechnienia informacji o danych badawczych będą powstałe w wyniku realizacji projektu artykuły naukowe dostępne w formule Open Access oraz referaty wygłoszone na międzynarodowych konferencjach naukowych.

• Jak długo będą one przechowywane?

Wszystkie dane badawcze (udostępnione i nieudostępnione) wytworzone i zgromadzone w ramach projektu będą przechowywane przez **minimum 10 lat**.

• Czy istnieją jakieś ograniczenia i przeszkody uniemożliwiające ich pełne lub częściowe udostępnienie?

Ograniczenia i przeszkody uniemożliwiające pełne udostępnienie dotyczą:

- 1) **danych surowych ze względu na ich ilość** (zbyt duża liczba plików uniemożliwia udostępnienie danych w sposób uporządkowany i umożliwiający skorzystanie dla zainteresowanych użytkowników), a dokładnie:
 - a) około 31000 plików graficznych macierzystych zgromadzonych na serwerze dedykowanym (przewiduje się możliwość udostępnienia plików .tif na potrzeby publikacji wyników naukowych badań zespołowi badawczemu);
 - b) około 31000 plików graficznych po konwersji zgromadzonych na serwerze dedykowanym;
- 2) **danych stanowiących kopie materiałów należących do instytucji trzecich**, a dokładnie:
 - a) plików (kopii) cyfrowych dokumentów bibliotecznych i archiwalnych oraz opracowań naukowych, materiałów prasowych, wspomnień itp.;
 - b) kopii w formie papierowej dokumentów bibliotecznych i archiwalnych oraz opracowań naukowych, materiałów prasowych, wspomnień itp.

• Czy wydawcy czasopism będą wymagać składowania danych na poparcie ustaleń publikacji?

Jeżeli będzie to wymagane przez wydawcę lub redakcję czasopisma wraz z artykułem zostanie opublikowany zbiór danych badawczych.

• Czy udostępnianie danych wymaga zgody uczestników badania?

nie dotyczy

5.2. Sposób wyboru danych przeznaczonych do przechowywania oraz miejsce długotrwałego przechowywania danych (np. repozytorium lub archiwum danych)

- **Które dane trzeba zachować, a które zniszczyć z przyczyn wynikających z umów, przepisów prawnych lub regulacji?**

nie dotyczy

- **W jaki sposób zostanie podjęta decyzja o tym, które dane zachować?**

Przechowywane będą **wszystkie** dane badawcze surowe i przetworzone zgromadzone, wytworzone i opracowane podczas projektu.

- **Jaka będzie procedura selekcji przeznaczonych do utrwalenia danych?**

nie dotyczy

- **Z jakiego repozytorium będą Państwo korzystać?**

Zbiór danych badawczych zostanie zdeponowany w repozytorium zarejestrowanym w agregatorze **Registry of Reserche Data Repositories** (<https://www.re3data.org/>), np. repozytorium RepOD (<https://repor.icm.edu.pl/>).

- **Czy przestrzega ono zasad FAIR Data?**

Repozytorium RepOD lub inne, które zostanie wybrane jako miejsce zdeponowania zbioru danych **przestrzega zasad FAIR Data** (zbiór danych i metadane będą: a) możliwe do znalezienia; b) dostępne; c) interoperacyjne; d) możliwe do ponownego wykorzystania).

- **Czy Państwa instytucja tworzy regularnie kopie zapasowe?**

Zob. punkt 3.1.

5.3. Metody lub narzędzia programowe umożliwiające dostęp do danych i korzystanie z danych

- **Czy aby dane mogły być przechowywane przez dłuższy okres i zachowały długi okres ważności należy je przekształcić do formatu standardowego lub otwartego?**

Dane w postaci bazy danych udostępnionej w sieci Internet oraz plików graficznych macierzystych i przetworzonych zostaną umieszczone na dedykowanym serwerze Biblioteki Uniwersyteckiej UWM, gdzie będą przechowywane przez minimum 10 lat. W przyszłości przewiduje się **konwersję danych do utworzonego repozytorium instytucjonalnego**. Konwersję bazy danych i danych surowych zakłada się również w ewentualnym przypadku **implementacji oprogramowania** (system chmurowy) bądź zmian w funkcjonalnościach systemu bibliotecznego. Dane bibliograficzne i metadane przeznaczone do ciągłego udostępniania i długotrwałego przechowywania będą przechowywane na **serwerze Zintegrowanego Systemu Bibliotecznego ALEPH**. Zbiór danych wraz z metadanymi zdeponowany w repozytorium danych badawczych będzie **dostępny tak długo, jak długo będzie istniało repozytorium**.

- **Czy do skanowania lub konwersji niezbędny będzie dodatkowy sprzęt lub oprogramowanie?**

Realizacja projektu będzie wymagała stworzenia pięciu stanowisk pracy umożliwiających budowę bazy danych oraz skanowanie grafik. Każde stanowisko pracy będzie składało się z komputera mobilnego lub stacjonarnego, skanera, oprogramowania komputerowego oraz licencji na program umożliwiający obróbkę plików graficznych (np. Adobe Creative Cloud). W ramach projektu zostaną zakupione **dwa komputery stacjonarne wraz z oprogramowaniem, jeden komputer mobilny (laptop) wraz z oprogramowaniem, 5 licencji na korzystanie z programu do obróbki plików graficznych.**

- **Jaki mechanizm posłuży do udostępniania danych (np. odpowiedzi na żądanie, repozytorium)?**

Aby uzyskać dostęp do danych przeznaczonych do udostępnienia (baza danych i zbiór danych z metadanymi) nie jest konieczne specjalistyczne oprogramowanie.

Dostęp i korzystanie z bazy danych zapewni katalog dostępny przez przeglądarkę internetową z wykorzystaniem modułu **WWW OPAC Zintegrowanego Systemu Bibliotecznego ALEPH**.

Ten katalog online to proste w obsłudze ogólnodostępne narzędzie wyszukiwawcze oferujące użytkownikom szybki dostęp do bazy danych i plików graficznych poprzez wiele opcji wyszukiwawczych. Możliwe jest **wyszukiwanie proste i zaawansowane** przez słowa występujące we wszystkich polach opisu oraz przeglądanie indeksów i krzyżowanie/zawężanie otrzymanych wyników przy pomocy operatorów logicznych.

Repozytorium danych badawczych umożliwi odczytywanie zdeponowanego zbioru danych przez **oprogramowanie typu open source**. Istnieje możliwość **eksportu metadanych** opisujących zdeponowany w repozytorium zbiór danych (np. eksport do OpenAIRE, DDI, JSON, DataCite, Dublin Core, JSON-LS Schema.org, OAI_ORE).

5.4. **Sposób zapewniający stosowanie unikalnego i długotrwałego identyfikatora (np. cyfrowego identyfikatora obiektu DOI) dla każdego zestawu danych**

- **Czy trwale przypisany identyfikator zostanie uzyskany?**

Z dwóch typów danych udostępnianych identyfikator będzie posiadał **zbiór danych badawczych** zdeponowany w otwartym repozytorium.

- **Jakie istniejące trwale identyfikatory pozostaną w użyciu (np. cyfrowe identyfikatory dokumentu elektronicznego, numery dostępne)?**

Wymieniony wyżej zbiór danych badawczych wraz z metadanymi będzie posiadał **numer DOI**. Ten sam numer będą posiadały wszystkie udostępnione wersje.

6. **Zadania związane z zarządzaniem danymi oraz zasoby**

6.1. **Kto będzie odpowiadał za zarządzanie danymi (tj. kto będzie ich opiekunem)?**

- **Czy Państwa instytucja posiada personel odpowiedzialny za zarządzanie danymi w trakcie trwania projektu, tj. od ich utworzenia, początkowego przechowywania, aż do archiwizacji i długoterminowego przechowywania danych?**

Za zarządzanie danymi badawczymi będzie odpowiedzialna osoba zatrudniona w projekcie w zespole badawczym (Wykonawca 4) z zakresem obowiązków odpowiednim dla **opiekuna systemu informatycznego**. Wynagrodzenie za zarządzanie danymi zostanie ujęte w kosztorysie projektu.

- **Jeśli NIE to kto jest odpowiedzialny za archiwizację i długoterminowe zarządzanie danymi w Państwa Jednostce?**

nie dotyczy

- **Jeśli TAK to jakie zajmuje stanowisko w obrębie Jednostki?**

Osoba zarządzająca danymi badawczymi w projekcie pełni obecnie funkcję zastępcy **Dyrektora Biblioteki Uniwersyteckiej, bibliotekarza systemowego oraz nadzoruje prace Oddziału ds. Komputeryzacji**. Po zakończeniu projektu za prace związane zarządzaniem danymi w celu utrzymania ciągłości i trwałości efektów projektu będzie odpowiadał wyznaczony pracownik Oddziału ds. Komputeryzacji jednostki, w której będzie realizowany projekt.

- **Jaką rolę pełni Kierownik projektu w zarządzaniu danymi?**

Kierownik projektu **przygotuje koncepcję organizacji plików**, będzie nadzorował **aktualizację rejestry danych badawczych** oraz **prace wykonawców w zakresie poprawności zapisu plików**.

6.2. Środki (np. finansowe i czasowe) przeznaczone do zarządzania danymi i zapewnienia możliwości odnalezienia, dostępu, interoperacyjności i ponownego wykorzystania danych

- **Czy potrzebne są dodatkowe zasoby do zarządzania danymi, takie jak osoby, czas, sprzęt lub oprogramowanie?**

Zarządzanie danymi podczas trwania projektu stanowi część obowiązków **Wykonawcy 4**. Koszty przechowania danych zostaną poniesione przez jednostkę realizującą projekt, która **zapewni miejsce na dysku serwera, system informatyczny z wersją umożliwiającą dostęp online (WWW OPAC)**. Po zakończeniu projektu do zarządzania danymi zostanie wyznaczona osoba z Oddziału ds. Komputeryzacji.

- **Jakie koszty związane będą z zapewnieniem standardów FAIR w projekcie?**

nie dotyczy

- **W jaki sposób zostaną opłacone?**

nie dotyczy