

Polityka długoterminowego przechowywania danych repozytorium Dane Badawcze UW

1. Organizacja repozytorium

1.1. Podmiot prowadzący

Repozytorium Dane Badawcze UW („Repozytorium”) prowadzone jest przez Uniwersytet Warszawski, z siedzibą w Warszawie (00-927) przy ul. Krakowskie Przedmieście 26/28 („UW”) za pośrednictwem serwisu internetowego Repozytorium (<https://danebadawcze.uw.edu.pl/>).

1.2. Cel działania Repozytorium

Repozytorium służy udostępnianiu za pośrednictwem sieci Internet danych badawczych, takich jak dane tabelaryczne, zdjęcia, materiały audiowizualne i wszelkie inne rodzaje danych wytworzonych, zebranych lub opisanych na potrzeby prowadzenia badań naukowych na Uniwersytecie Warszawskim lub we współpracy z Uniwersytetem.

1.3. Zakres danych i grupa docelowa

Repozytorium Dane Badawcze UW jest instytucjonalnym repozytorium danych badawczych Uniwersytetu Warszawskiego, prowadzonym z myślą o wszystkich członkach społeczności naukowej UW.

W Repozytorium mogą być gromadzone, przechowywane i udostępniane dane badawcze ze wszystkich dziedzin nauki.

1.4. Podstawa działania Repozytorium

Repozytorium działa na podstawie regulaminu dostępnego na stronie <https://danebadawcze.uw.edu.pl/terms-of-use-page.xhtml>. Dodatkowe informacje dotyczące deponowania danych badawczych są zawarte w Poradniku użytkownika dostępnym pod adresem <https://repod.icm.edu.pl/guides/pl/1.0.0-SNAPSHOT/user/index.html> oraz w witrynie informacyjnej Repozytorium dostępnej pod adresem <http://info.danebadawcze.uw.edu.pl>.

Repozytorium powstało w ramach programu „Inicjatywa Doskonałości – Uczelnia Badawcza”, realizowanego przez Uniwersytet Warszawski w latach 2020–2026.

1.5. Personel Repozytorium

Zadania związane z bieżącym prowadzeniem Repozytorium i zarządzaniem danymi w Repozytorium ujęte są w dokumentach określających zakres obowiązków odpowiednich pracowników Interdyscyplinarnego Centrum Modelowania Matematycznego i Komputerowego UW (ICM UW).

Obowiązki związane z bieżącym prowadzeniem Repozytorium są uwzględniane przez kierownictwo ICM UW podczas określania struktury organizacyjnej ICM UW (zgodnie z §5 Regulaminu ICM UW) oraz w procesie planowania zatrudnienia.

1.6. Finansowanie działalności Repozytorium

Utworzenie Repozytorium i jego prowadzenie do roku 2025 zostało sfinansowane ze środków programu „Inicjatywa Doskonałości – Uczelnia Badawcza”.

Polityka otwartego dostępu Uniwersytetu Warszawskiego oraz Strategia Uniwersytetu Warszawskiego na lata 2023–2032 wskazują Repozytorium jako element infrastruktury badawczej Uniwersytetu, do której należy zapewnić komfortowy, zintegrowany i ustandaryzowany dostęp, oraz wskazują bezpośrednią zależność między realizacją tego zobowiązania a potrzebą uwzględnienia rosnących kosztów i poziomu skomplikowania zarządzania danymi oraz ich zbiorami.

1.7. Informacje dotyczące wykorzystania danych

W obrębie systemu gromadzone są informacje dotyczące pobrań poszczególnych plików i zbiorów danych. Informacje o liczbie pobrań poszczególnych plików są dostępne publicznie.

W odniesieniu do całości udostępnianych zasobów Repozytorium publicznie dostępne są ponadto informacje o łącznej liczbie pobrań plików oraz całych zbiorów danych.

1.8. Audyt i certyfikacja

W czerwcu 2025 r. rozpoczął się proces aplikowania o certyfikat Core Trust Seal dla Repozytorium.

1.9. Postępowanie na wypadek zakończenia funkcjonowania Repozytorium

W wypadku zaistnienia okoliczności powodujących konieczność zakończenia działalności Repozytorium, Uniwersytet Warszawski podejmie starania mające na celu przeniesienie wszystkich zgromadzonych w nim danych do innej lokalizacji, z zachowaniem ciągłości poprawnego funkcjonowania identyfikatorów DOI nadanych zbiorom danych.

Wybór nowej lokalizacji dla zbiorów danych zależeć będzie od bieżącej dostępności infrastruktury umożliwiającej przeniesienie danych.

1.10. Zmiany polityki długoterminowego przechowywania danych

Kolejne wersje niniejszego dokumentu znajdują się w witrynie informacyjnej Repozytorium, dostępnej pod adresem <http://info.danebadawcze.uw.edu.pl>.

2. Funkcjonowanie Repozytorium i zarządzanie danymi

2.1. Kolekcje

W ramach repozytorium funkcjonują kolekcje wszystkich wydziałów UW oraz kolekcje innych jednostek UW, takich jak centra i instytuty. Zasady funkcjonowania kolekcji są takie same, jak zasady funkcjonowania kolekcji głównej.

2.2. Pozyskiwanie danych

Repozytorium nie określa szczegółowych wymogów dotyczących właściwości zbiorów danych, z wyjątkiem konieczności zdeponowania co najmniej jednego pliku w ramach zbioru. Preferowane są otwarte formaty plików.

Repozytorium każdorazowo zachowuje oryginalny format zdeponowanego pliku.

Repozytorium akceptuje zbiory danych o różnej wielkości, z uwzględnieniem kryterium technicznej dostępności zbiorów o określonej wielkości dla użytkowników. Limit wielkości pojedynczego pliku w zbiorze wynosi 5 GB.

Dla każdego przesłanego do Repozytorium pliku wyliczana jest suma kontrolna MD5. Umożliwia to porównanie jej z sumą kontrolną pliku wyliczoną po stronie użytkownika i zweryfikowanie poprawności przesyłu danych.

Repozytorium umożliwia przysyłanie plików z danymi zarówno przez graficzny interfejs użytkownika dostępny przez przeglądarkę www, jak i za pomocą interfejsu API.

Tworząc zbiór danych, użytkownik tworzy początkowo jego wersję roboczą, która może podlegać zmianom i uzupełnieniom w zakresie metadanych i plików. Po jej opublikowaniu i nadaniu numeru wersji użytkownik nie ma możliwości zmodyfikowania składających się na nią metadanych i plików.

Użytkownik ma jednak możliwość utworzenia kolejnych, numerowanych wersji zbioru danych na bazie istniejącej, opublikowanej wersji.

2.3. Wymagania dotyczące deponowania zbiorów danych

Repozytorium nie pobiera opłat od użytkowników deponujących oraz pobierających dane.

Korzystanie z funkcji przechowywania i udostępniania danych w Repozytorium jest możliwe dla osób posiadających konto w Centralnym Systemie Uwierzytelniania UW (CAS UW) lub – w wypadku braku takiego konta – posiadających konto, do którego przypisane są adres e-mail i hasło uwierzytelniające, z odpowiednimi uprawnieniami nadanymi przez Administratorów.

Zakres informacji wymaganych podczas tworzenia i edycji zbioru danych wskazany jest bezpośrednio w formularzu metadanych. Zbiór danych może zostać przesłany do weryfikacji jeśli zawiera co najmniej jeden plik, a wszystkie wymagane pola metadanych zostały wypełnione.

Konfiguracja zestawów metadanych oraz pól obowiązkowych jest możliwa na poziomie repozytorium i na poziomie kolekcji. Konfigurację przeprowadzają Administratorzy repozytorium.

2.4. Minimalny zestaw metadanych w Repozytorium

Repozytorium, jako repozytorium ogólnego przeznaczenia, wymaga wprowadzenia jedynie podstawowych informacji o zbiorach danych. Są to pola:

- tytuł;
- autor (nazwisko i imię autora);
- osoba do kontaktu (e-mail);
- opis (treść);
- słowa kluczowe (treść);
- obszar tematyczny.

Dodatkowo we wszystkich kolekcjach Repozytorium dostępne są następujące pola, których wypełnienie jest opcjonalne:

- tłumaczenie tytułu (treść; język tłumaczenia);
- autor (system identyfikacji; identyfikator; afiliacja; ROR instytucji afiliującej);
- osoba do kontaktu (nazwisko i imię; afiliacja);
- opis (data);
- dyscyplina MEiN 2022;
- dziedzina B+R UNESCO, OECD;
- słowa kluczowe (słownik; adres URL słownika);
- powiązana publikacja (cytowanie; rodzaj powiązania; typ identyfikatora; identyfikator; adres URL);
- uwagi;
- język;
- język metadanych;
- język danych;
- język dokumentacji;
- informacje o grantie (instytucja finansująca; skrócona nazwa instytucji; ROR instytucji finansującej; program grantowy; numer grantu);
- powiązany zbiór danych (cytowanie; typ powiązania; typ identyfikatora; identyfikator; URL).

Więcej opcjonalnych pól metadanych można wypełnić w trybie edycji po zapisaniu wersji roboczej zbioru.

Odrębne wymogi dotyczące obowiązkowych pól metadanych mogą obowiązywać w obrębie poszczególnych kolekcji.

Deponując dane, użytkownik musi ponadto wskazać licencje lub warunki, na których udostępniane będą poszczególne pliki.

2.5. Walidacja metadanych

Do utworzenia wersji roboczej zbioru danych, a następnie jego opublikowania, niezbędne jest wprowadzenie minimalnego wymaganego zakresu metadanych. Wypełnienie odpowiednich pól jest weryfikowane automatycznie przez oprogramowanie Repozytorium w momencie zapisu wersji roboczej zbioru. Dodatkowo poprawność opisu jest sprawdzana w momencie weryfikacji zbioru przed jego opublikowaniem.

2.6. Zbiory objęte embargiem

W wypadku zbiorów danych objętych embargiem użytkownik określa datę, od której pliki w zbiorze danych staną się dostępne. Od momentu publikacji takiego zbioru publicznie dostępne są jego metadane. Embargo ustawić można jedynie w odniesieniu do tych zbiorów, które nie posiadają jeszcze żadnej opublikowanej wersji.

2.7. Maksymalny okres embargo

Maksymalny okres embargo w Repozytorium wynosi 36 miesięcy.

2.8. Pliki w zbiorach udostępnione w sposób ograniczony

W wypadku plików w zbiorze danych udostępnianych w sposób ograniczony, Repozytorium pozwala na wystosowanie prośby o uzyskanie dostępu do określonego pliku. Prośba kierowana jest do użytkownika Repozytorium, który zdeponował zbiór danych zawierający ten plik. Obsługa Repozytorium otrzymuje kopię e-maila zawierającego prośbę o dostęp do zasobu tego rodzaju, ale nie ingeruje w dalszą korespondencję pomiędzy użytkownikiem chcącym uzyskać dostęp do plików w zbiorze a użytkownikiem, który go zdeponował.

2.9. Kontrola antywirusowa

W momencie przesyłania do Repozytorium pliki podlegają kontroli antywirusowej.

W razie wykrycia nieprawidłowości zapis pliku zostaje uniemożliwiony, a użytkownik otrzymuje informację o niepowodzeniu operacji przesyłania pliku.

2.10. Weryfikacja zdeponowanych zbiorów danych

Po zamieszczeniu i opisanu danych badawczych w Repozytorium użytkownik przekazuje je do weryfikacji. Weryfikacji dokonuje Administrator repozytorium. Składają się na nią:

- weryfikacja poprawności wprowadzonych metadanych;
- weryfikacja poprawności plików.

Proces ten (od momentu przesłania do weryfikacji do momentu decyzji o publikacji lub zwróceniu zbioru do poprawek) zajmuje zwykle do 3 dni roboczych.

Formalnym potwierdzeniem publikacji wersji zbioru danych jest powiadomienie wysyłane na adres e-mail deponującego.

2.11. Wersje zbiorów danych

Każdy ze zdeponowanych przez użytkownika zbiorów danych składa się z co najmniej jednej wersji. Na wersję zbioru danych składają się metadane oraz pliki zawierające dane oraz ich dodatkową dokumentację.

Tworząc lub edytując zbiór danych, użytkownik tworzy początkowo wersję roboczą zbioru. Wersję tę można edytować zarówno w zakresie metadanych, jak i plików. Możliwość edycji zostaje zablokowana w momencie publikacji konkretnej wersji zbioru danych. Wyjątek od tej zasady stanowią drobne korekty oczywistych omyłek w zakresie metadanych, dokonywane przez administratora Repozytorium. Zmiany tego rodzaju nie wymagają opublikowania nowej wersji zbioru danych.

Publikując zbiór danych, Administrator repozytorium określa, czy wprowadzone zmiany wymagają opublikowania wersji roboczej jako nowej wersji głównej (major) czy jako wersji dodatkowej (minor).

Wersję roboczą, w której zmodyfikowane zostały pliki, opublikować można jedynie jako nową wersję główną.

Numer wersji zbioru danych stanowi element sugerowanego cytowania zbioru danych, które widoczne jest na stronie zbioru danych.

2.12. Usuwanie i wycofywanie zbiorów i ich wersji

Po opublikowaniu danej wersji zbioru danych nie ma możliwości jej usunięcia. W szczególnych przypadkach istnieje jednak możliwość wycofania wersji zbioru. Wycofanie całego zbioru danych następuje poprzez wycofanie wszystkich składających się na niego wersji.

W razie wycofania wersji zbioru, publicznie dostępne pozostają jedynie podstawowe informacje dotyczące zbioru (tzw. tombstone):

- cytowanie;
- powód wycofania.

Pełne metadane i dane wycofanej wersji pozostają do wglądu osób posiadających role systemowe umożliwiające publikowanie i wycofanie zbioru, tj. Administratorów repozytorium.

2.13. Forma archiwizacji danych

Dane podlegają archiwizacji i udostępnianiu w postaci dostarczonej do Repozytorium przez użytkownika. Administrator może przed opublikowaniem zbioru danych poprosić o wprowadzenie do niego korekt lub uzupełnień, a w wypadku drobnych i oczywistych pomyłek (np. literówek) wprowadzić niezbędne poprawki samodzielnie.

W wypadku wybranych formatów danych tabelarycznych tworzone są dodatkowo kopie tych danych w innych formatach, w celu zwiększenia dostępności danych dla użytkowników korzystających z oprogramowania różnego typu. Konwersja ta odbywa się automatycznie. Zdeponowany plik każdorazowo zachowywany jest w oryginalnym formacie.

W wypadku plików tabelarycznych dodatkowo generowana jest wartość UNF (universal numerical fingerprint), co pozwala użytkownikom na weryfikację poprawności dokonanej konwersji.

2.14. Identyfikator DOI

Każdy ze zbiorów danych zdeponowanych w Repozytorium opatrzony zostaje identyfikatorem DOI w obrębie przypisanego do Repozytorium prefiksu 10.58132. Identyfikator ten w momencie utworzenia pierwszej wersji roboczej zbioru danych jest rezerwowany w obrębie infrastruktury

ekosystemu Datacite. Aktywacja identyfikatora DOI następuje w momencie opublikowania pierwszej wersji zbioru danych.

Identyfikator DOI zbioru danych stanowi element sugerowanego cytowania, które widoczne jest na stronie zbioru danych.

2.15. Identyfikator DOI a wersje zbioru danych

Wszystkie wersje zbioru danych mają ten sam identyfikator DOI, ich wyróżnikiem jest zaś numer wersji. Informacja o wszystkich wersjach zbiorów prezentowana jest na stronie zbioru danych w zakładce „Wersje”.

W obrębie metadanych możliwe jest ponadto wskazanie innych obiektów (publikacji, zbiorów danych) powiązanych ze zbiorem danych, identyfikatorów DOI tych obiektów, jak również określenie typu powiązania.

2.16. Logi systemowe

W wypadku wszystkich zasobów Repozytorium logowaniu podlega każde pobranie pliku.

W wypadku każdego działania tego rodzaju w logach zapisywane są informacje o:

- identyfikatorze pliku;
- rodzaju pobrania;
- e-mailu użytkownika pobierającego plik;
- imieniu i nazwisku użytkownika pobierającego plik (jeśli dotyczy);
- stanowisku użytkownika pobierającego plik (jeśli dotyczy);
- dacie i godzinie pobrania pliku;
- identyfikatorze zalogowanego użytkownika (jeśli dotyczy);
- identyfikatorze pliku danych;
- identyfikatorze zbioru danych;
- identyfikatorze wersji zbioru danych;
- identyfikatorze księgi gości (jeśli dotyczy).

2.17. Wsparcie użytkowników Repozytorium

Repozytorium umożliwia użytkownikom kontakt poprzez e-mail (danebadawczeuw@icm.edu.pl) oraz formularz kontaktowy znajdujący się na stronie Repozytorium. Zgłoszone problemy rozwiązywane są przez odpowiednich pracowników ICM UW.

W wypadku pytań dotyczących poszczególnych zbiorów danych Repozytorium umożliwia kontakt z osobami wskazanymi jako osoby do kontaktu w sprawach dotyczących konkretnego zbioru danych.

Ponadto użytkownicy posiadający odpowiednie role systemowe otrzymują automatyczne powiadomienia e-mailowe dotyczące:

- utworzenia zbioru danych;
- przesłania zbioru danych do weryfikacji;
- zwrócenia zbioru danych do poprawy;
- opublikowania zbioru danych;
- przesłania prośby o dostęp do pliku w ograniczonym dostępie;
- przyznania roli systemowej;
- odebrania roli systemowej.

3. Prawne aspekty funkcjonowania Repozytorium

3.1. Prawa do zbiorów danych

Wyjaśnienie prawnego statusu deponowanych danych na mocy Regulaminu spoczywa na użytkowniku Repozytorium. Dostarczając dane do Repozytorium użytkownik oświadcza, iż posiada wystarczające uprawnienia do ich zdeponowania i udostępnienia. Wyjaśnienie ewentualnych wątpliwości dotyczących prawnego statusu deponowanych danych na mocy Regulaminu spoczywa na użytkowniku dostarczającym dane do Repozytorium.

3.2. Licencja dla Uniwersytetu Warszawskiego

Przesyłając dane do Repozytorium użytkownik udziela Uniwersytetowi Warszawskiemu niewyłącznej licencji obejmującej utrwalanie, przechowywanie i zwielokrotnianie danych badawczych techniką cyfrową, w celu prowadzenia Repozytorium. Licencja ta obejmuje całość dostarczonych danych badawczych, a w szczególności metadane, prawa związane z danymi badawczymi jako zestawieniem plików oraz pliki stanowiące elementy składowe danych badawczych. Licencja obejmuje także udostępnianie danych badawczych osobom trzecim, z wyjątkiem sytuacji, w której dostęp do konkretnego pliku został ograniczony.

3.3. Oświadczenie dotyczące metadanych i zestawienia plików

Dostarczenie danych do Repozytorium jest równoznaczne ze złożeniem przez użytkownika oświadczenia o treści zgodnej ze wzorcem Creative Commons Zero („CC0”), którego pełne postanowienia dostępne są pod adresem: <https://creativecommons.org/publicdomain/zero/1.0/legalcode>, którego przedmiotem są metadane opisujące dane badawcze oraz ewentualne prawa związane ze zbiorem jako zestawieniem plików.

Objęcie powyższym oświadczeniem metadanych opisujących dane badawcze umożliwia swobodną wymianę metadanych z innymi serwisami agregującymi informacje o dostępnych zbiorach danych.

Objęcie powyższym oświadczeniem praw związanych ze zbiorem jako zestawieniem plików (np. praw do utworu zbiorowego, praw sui generis do baz danych) ułatwia ocenę zakresu swobody użytkownika, która w takiej sytuacji wynika tylko z zasad określonych dla poszczególnych plików wchodzących w skład zestawu (np. wskazanej licencji CC).

3.4. Warunki udostępniania dla plików

Do każdego z plików wchodzących w skład deponowanego zbioru danych użytkownik może przypisać licencję z listy licencji dostępnej w repozytorium, udostępnić go na zasadach dozwolonego użytku lub w sposób ograniczony.

W wypadku plików udostępnionych w sposób ograniczony, decyzja o udostępnieniu pliku konkretnemu użytkownikowi jest każdorazowo podejmowana przez deponującego i może być uzależniona od zaakceptowania przez użytkownika dodatkowych warunków (np. możliwość użycia pliku jedynie do celów naukowych).

3.5. Wielokrotne licencjonowanie

Dopuszczalna jest sytuacja, w której te same pliki dostępne są na różnych licencjach w różnych wersjach zbioru danych (wielokrotne licencjonowanie). Jeżeli każdy z tych plików jest publicznie dostępny, użytkownik pliku może wybrać, do której z tych licencji będzie się stosował.

3.6. Forma licencji

W wypadku, w którym użytkownik udostępnia zasoby, w tym pliki oraz metadane, na podstawie licencji niewyłącznej, jej pisemna forma nie jest wymagana (polskie prawo pozwala na udzielanie licencji niewyłącznych w dowolnej formie).

W wypadku, w którym użytkownik udostępnia pliki zbioru danych na zasadach dozwolonego użytku, udostępnienie i wykorzystanie danych możliwe jest na mocy obowiązujących przepisów prawa. W tym wypadku nie istnieją więc żadne postanowienia, w odniesieniu do których konieczne byłoby rozważanie kwestii ich formy.

3.7. Udostępnianie plików w sposób ograniczony

W wypadku, w którym użytkownik udostępnia pliki zbioru danych w sposób ograniczony, dodatkowe warunki dostępu i wykorzystania plików opisane są przy każdym udostępnionym na tych zasadach pliku.

Udostępnienie plików konkretnemu użytkownikowi każdorazowo zależy od decyzji użytkownika, po stronie którego znajdują się również ewentualne dodatkowe działania mające na celu weryfikację osoby, która poprosiła o udostępnienie danych.

3.8. Informowanie o sposobie wykorzystania danych

Informacja dotycząca zasad i warunków, na jakich użytkownik może udostępnić dane oraz metadane, zawarta jest w Regulaminie Repozytorium. Każdy użytkownik akceptuje Regulamin w momencie rejestracji konta.

Informacja o tym, na jakiej licencji udostępnione zostały poszczególne pliki w zbiorze danych, prezentowana jest przy każdym z nich w zakładce „Pliki” na stronie zbioru danych oraz dodatkowo na stronie każdego pliku.

3.9. Monitorowanie sposobu wykorzystania danych

Repozytorium nie monitoruje zgodności wykorzystania danych z licencjami, na jakich zostały udostępnione.

Repozytorium nie monitoruje, czy dane pobrane z Repozytorium są wykorzystywane zgodnie z ich licencją. W wypadku naruszenia postanowień licencji lub warunków dostępu, podjęcie ewentualnych kroków prawnych należy do podmiotu posiadającego prawa do zasobu, którego takie naruszenie dotyczy.

Repozytorium nie ingeruje również w proces dodatkowej weryfikacji użytkowników w wypadku plików udostępnionych na warunkach ograniczonego dostępu (np. wyłącznie do celów naukowych). Ta ewentualność również pozostaje po stronie użytkowników Repozytorium deponujących pliki na zasadach ograniczonego dostępu.

W wypadku zbiorów udostępnionych na innych zasadach uzyskanie dostępu do nich nie wymaga zalogowania się ani posiadania konta w Repozytorium.

4. Przechowywanie danych

4.1. Integralność danych i kopie robocze

Za zachowanie spójności i poprawności plików z danymi odpowiada oprogramowanie macierzy S3, na której są one składowane.

Dodatkowo, na wypadek awarii lub przypadkowego usunięcia zasobów z macierzy S3 (np. w wyniku błędu oprogramowania) istnieje również możliwość przywrócenia utraconych zasobów z dodatkowej kopii zapasowej składowanej w osobnej lokalizacji.

Kopie zapasowe bazy danych oraz plików przechowywane są osobno.

4.2. Weryfikacja integralności danych

W momencie przesyłania pliku do Repozytorium generowana jest jego suma kontrolna. Umożliwia to użytkownikowi porównanie sumy kontrolnej wygenerowanej lokalnie w celu potwierdzenia zgodności kopii pliku przesłanej do Repozytorium z plikiem znajdującym się po stronie użytkownika.

Repozytorium okresowo sprawdza, czy zasoby, które zgodnie z informacjami znajdującymi się w bazie danych Repozytorium powinny znajdować się na macierzy S3, rzeczywiście się na niej znajdują. W wypadku stwierdzenia rozbieżności, odzyskiwana jest kopia zapasowa plików znajdująca się na osobnej macierzy w innej lokalizacji.

Okresowo porównywana jest również zgodność sum kontrolnych znajdujących się w bazie danych Repozytorium z sumami kontrolnymi zapisanymi w metadanych plików na macierzy S3.

Informacja o sumie kontrolnej jest również wyświetlana przy każdym pliku na stronie zbioru danych oraz na stronie pliku. Umożliwia to porównanie zgodności pobranej kopii pliku po stronie użytkownika z jego kopią zdeponowaną w repozytorium.

4.3. Relacja między kopiami archiwalnymi (AIP) a kopiami podlegającymi upowszechnieniu (DIP)
Repozytorium nie tworzy kopii archiwalnych (AIP) odrębnych od kopii zbiorów danych podlegających upowszechnieniu (DIP).

5. Infrastruktura techniczna repozytorium

5.1. Oprogramowanie repozytorium

Repozytorium działa na bazie wolnego oprogramowania Dataverse w wersji 4.11, zmodyfikowanego przez zespół programistów ICM UW. Zmodyfikowana wersja oprogramowania nosi nazwę Fairchive, jest otwarta, a jej kod źródłowy dostępny jest na platformie GitHub pod adresem <https://github.com/CeON/fairchive>. Platforma ta umożliwia zgłaszanie i monitorowanie zmian i poprawek wprowadzanych do kodu oprogramowania.

5.2. Przechowywanie danych

Dane plikowe przechowywane są na macierzy S3. Metadane zbiorów i plików przechowywane są w postaci relacyjnej bazy danych w zasobie macierzowym zamontowanym bezpośrednio do serwera.

Bezpieczeństwo danych deponowanych w repozytorium zapewnia ponadto natywny mechanizm zapisu nadmiarowego macierzy dyskowej S3, na której umieszczane są dane. System wspierany jest przez firmę zewnętrzną.

Dodatkowo baza danych metadanych oraz dane plikowe przesyłane są do zasobu dyskowego znajdującego się w innej lokalizacji geograficznej. Umożliwia to odzyskanie plików z danymi oraz bazy danych repozytorium w razie ich uszkodzenia lub awarii.

Kopie zapasowe danych plikowych oraz bazy danych metadanych tworzone są raz na dobę. W odrębnej lokalizacji przechowywane są kopie bazy danych z ostatnich 30 dni oraz kopie danych z ostatnich 2 dni.

5.3. Postępowanie na wypadek dezintegracji danych

W wypadku różnicy w sumach kontrolnych deponowanych plików zapisanych w bazie danych oraz metadanych systemowych na macierzy S3 system wysyła powiadomienie do administratorów o zaistniałym problemie. Administratorzy podejmują działania w celu przywrócenia spójności systemu.

5.4. Logi operacji

W repozytorium tworzone i przechowywane są logi systemowe dla następujących operacji:



- utworzenie konta użytkownika;
- wycofanie wersji zbioru danych;
- ustawienie daty embarga dla zbioru danych;
- włączenie blokady (lock) zbioru danych;
- przypisanie roli systemowej;
- utworzenie kolekcji;
- utworzenie wpisu w księdze gości;
- utworzenie nowego zbioru danych;
- utworzenie prywatnego URL zbioru danych;
- utworzenie szablonu zbioru danych;
- usunięcie wersji roboczej zbioru danych;
- usunięcie wersji roboczej będącej jedyną wersją zbioru danych;
- usunięcie szablonu zbioru danych;
- opublikowanie zbioru danych;
- pobranie JSON Id;
- pobranie prywatnego URL zbioru danych;
- zapisanie pliku proveniencji;
- opublikowanie zbioru danych;
- opublikowanie kolekcji;
- zdjęcie blokady (lock) zbioru danych;
- wystosowanie prośby o dostęp do pliku w ograniczonym dostępie;
- zwrócenie zbioru danych do poprawy;
- odebranie roli systemowej;
- przesłanie zbioru danych do weryfikacji;
- edycja miniatury zbioru danych;
- edycja wersji roboczej zbioru danych;
- utworzenie wersji roboczej zbioru danych;
- edycja kolekcji;
- zmiana domyślnej roli dla użytkownika tworzącego zbiór danych w kolekcji;
- edycja szablonu zbioru danych Edycja wyglądu kolekcji;
- zalogowanie użytkownika;
- wylogowanie użytkownika;
- wystosowanie prośby o zmianę hasła;
- wysłanie prośby o zmianę hasła;
- pobranie z bazy danych dostępnych metod logowania;
- zmiana maksymalnej długości embarga;
- aktywacja lub dezaktywacja innych warunków udostępniania plików (ograniczony dostęp oraz dozwolony użytek);
- zmiana parametrów konta użytkownika.

5.5. Okresowy przegląd funkcjonalności

Raz do roku dokonywany jest przegląd funkcjonalności repozytorium pod kątem zidentyfikowanych nowych potrzeb użytkowników, technologii oraz funkcjonalności. Nowe funkcjonalności będą sukcesywnie opracowywane i implementowane w miarę posiadanych zasobów kadrowych i finansowych.

Wersja dokumentu: 1.0.0