

09. Tworzenie repozytoriów instytucjonalnych w oparciu o system dLibra

Celem niniejszej strony jest przedstawienie wybranych aspektów wykorzystania systemu dLibra do tworzenia repozytoriów instytucjonalnych. Strona ta ma charakter uzupełniający w stosunku do reszty dokumentacji systemu i samodzielnie nie stanowi wyczerpującego opracowania. W zamian w stosownych miejscach niniejszej strony zawarto odnośniki do wybranych fragmentów dokumentacji. Strona skonstruowana jest w formie pytań i odpowiedzi - zachęcamy do zgłaszania kolejnych pytań w komentarzach do tej strony lub w [serwisie pytań i odpowiedzi Federacji Bibliotek Cyfrowych](#).

Spis pytań

- [Czym różnią się od siebie biblioteki cyfrowe i repozytoria instytucjonalne?](#)
- [Czy w ramach jednej instytucji biblioteka cyfrowa i repozytorium to powinny być odrębne inicjatywy czy też lepiej je połączyć?](#)
- [dLibra to system do budowy bibliotek cyfrowych, czy można więc na jego bazie tworzyć repozytoria instytucjonalne?](#)
- [Istotną cechą repozytoriów instytucjonalnych jest możliwość samodzielnego deponowania prac przez autorów \(tzw. self-archiving, samoarchiwizacja\). Czy dLibra to wspiera?](#)
- [Chcę udostępniać materiały na otwartych licencjach oraz materiały, do których dostęp mogą mieć tylko pracownicy mojej instytucji - wszystko to są publikacje pracowników mojej instytucji. Czy powinienem zbudować dwa odrębne repozytoria, czy też warto i da się to połączyć w jeden system?](#)
- [Jak w systemie dLibra mogę połączyć bibliotekę cyfrową i repozytorium instytucjonalne?](#)
- [Zależy mi na widoczności materiałów z mojego repozytorium w Google Scholar. Czy dLibra wspiera to w jakiś sposób?](#)
- [Jak szybko materiały opublikowane w repozytorium na dLibrze znajdą się w Google Scholar?](#)
- [Zależy mi na widoczności materiałów z mojego repozytorium w Google Books. Czy dLibra wspiera to w jakiś sposób?](#)
- [W mojej instytucji funkcjonuje wydawnictwo, które wydaje i sprzedaje przez internet książki, skrypty i czasopisma powstające u nas. Czy mogę w jakiś sposób uwzględnić te zasoby w repozytorium instytucjonalnym?](#)
- [Autorzy z mojej instytucji publikują swoje prace w międzynarodowych wydawnictwach, które udostępniają artykuły odpłatnie. Czy mogę w jakiś sposób uwzględnić te zasoby w repozytorium instytucjonalnym?](#)
- [Czy mogę w repozytorium gromadzić również informacje na temat publikacji, których treść nie jest w żaden sposób dostępna on-line?](#)
- [W mojej instytucji pracuje kilka tysięcy pracowników naukowych. Planujemy gromadzić wszystkie publikacje pracowników w repozytorium i być może wprowadzić również mandat Open Access. Czy kilka tysięcy użytkowników równocześnie wprowadzających swoje prace nie spowoduje przeciążenia systemu?](#)

Czym różnią się od siebie biblioteki cyfrowe i repozytoria instytucjonalne?

W podstawowym zakresie zarówno biblioteki cyfrowe, jak i repozytoria instytucjonalne służą temu samemu celowi: udostępnianiu obiektów cyfrowych wraz z metadanymi, zazwyczaj z podziałem na kolekcje i dodatkowymi narzędziami takimi jak wyszukiwanie, przeglądanie, indeksy itd. Stąd też terminy te często używane są wymiennie, zwłaszcza w kontekście technicznym. Rozróżnienie następuje zazwyczaj na poziomie rodzaju udostępnionych zbiorów. Termin "biblioteka cyfrowa" zwyczajowo stosuje się w kontekście serwisów udostępniających zbiory biblioteczne w postaci cyfrowej (zarówno zbiory będące efektem digitalizacji, jak i zbiory natywnie cyfrowe / born-digital). Z kolei termin "repozytorium instytucjonalne" opisuje serwisy, w których udostępniane są efekty pracy (najczęściej naukowej) osób zatrudnionych w danej instytucji, najczęściej powstające współcześnie. Nie muszą to być materiały oficjalnie wydane - mogą to być np. również prezentacje czy powstające wewnątrz instytucji raporty (np. techniczne). Istnieją różne podejścia do określenia wzajemnej relacji tych systemów - w zależności od kontekstu czasem terminy te używane są wymiennie, czasem mówi się że biblioteka cyfrowa jest terminem szerszym, a repozytorium jest tylko pewnym podzbiorem tego co może pojawić się w bibliotece cyfrowej (np. kolekcją), a czasem podaje się że repozytorium jest czymś ogólniejszym, niż biblioteka cyfrowa gdyż np. obejmuje prace nieobjęte recenzją, niewydane w sposób formalny, a jedynie udostępnione on-line. Do tego można jeszcze dodać terminy takie jak "muzeum cyfrowe" czy "archiwum cyfrowe", które znów są pewnym wariantem opisanych wcześniej systemów różniącym się charakterem udostępnianych zbiorów, sposobem ich opisu (schemat metadanych) czy prezentacji (interfejs użytkownika, usługi dodane).

Poza potencjalnymi różnicami w zakresie charakteru gromadzonych i udostępnianych zbiorów, często podawanym wyróżnikiem repozytoriów cyfrowych jest sposób gromadzenia tych zbiorów. O ile w przypadku typowych bibliotek cyfrowych gromadzenie materiałów i wprowadzanie ich do systemu jest zadaniem pracowników biblioteki, o tyle w przypadku repozytoriów instytucjonalnych często przyjmuje się, że materiały w dużej mierze gromadzone są metodą samoarchiwizacji / self-archivingu, czyli procesu w którym autor pracy przesyła ją np. przez specjalny formularz WWW do systemu repozytorium. Następnie administratorzy/moderatorzy repozytorium weryfikują materiał i decydują o jego udostępnieniu bądź zwróceniu do autora w celu wprowadzenia niezbędnych poprawek. Zależnie od zasad działania, część zmian może być wprowadzana przez administratorów repozytorium (np. uzupełnienie opisu obiektu, przypisanie do stosownych kolekcji) bądź też realizowana automatycznie np. konwersja plików zapisanych w formatach edytowalnych (np. doc) na formaty archiwizacyjne/dystrybucyjne (np. PDF/A). W tego typu sytuacji często realizowana jest również integracja systemu uwierzytelniania pracowników instytucji z systemem uwierzytelniania i autoryzacji repozytorium instytucjonalnego.

Zobacz też:

- [Pytanie "Biblioteki cyfrowe i repozytoria instytucjonalne - jaka różnica?" w serwisie FBC "Pytania i odpowiedzi"](#)

Czy w ramach jednej instytucji biblioteka cyfrowa i repozytorium to powinny być odrębne inicjatywy czy też lepiej je połączyć?

To zawsze decyzja indywidualna. Poniżej kilka podstawowych zalet jednego i drugiego podejścia:

- Połączenie biblioteki cyfrowej i repozytorium
 - **Dla użytkowników końcowych: łatwiejszy dostęp do większej liczby obiektów** - jeden punkt dostępu do zbiorów biblioteki (cyfrowej) i zbiorów twórczości pracowników danej instytucji.
 - **Dla administratorów/redaktorów biblioteki cyfrowej/repozytorium: prostsze i wygodniejsze zarządzanie gromadzeniem i udostępnianiem** - jeden system gromadzenia i udostępniania zbiorów do utrzymania i zarządzania.
 - **Dla informatyków: prostsze i tańsze utrzymanie systemu** - jeden system informatyczny do utrzymania, administrowania, aktualizowania, monitorowania, zabezpieczania, jedna infrastruktura sprzętowa
- Biblioteka cyfrowa i repozytorium jako odrębne inicjatywy
 - **Dla użytkowników końcowych: łatwiejszy dostęp do specjalizowanego zakresu materiałów** - punkt dostępu wyłącznie zbiorów współczesnej twórczości pracowników danej instytucji.
 - **Dla pracowników instytucji: łatwiejsza możliwość zaprezentowania profilu działalności/dorobku/portfolio własnego i instytucji** - dedykowany serwis prezentujący wyłącznie publikacje i inne materiały powstające współcześnie w danej instytucji.
 - **Dla administratorów/redaktorów repozytorium: nowe możliwości promocji repozytorium** - możliwość rejestracji w wyszukiwarkach i agregatorach danych i treści naukowych, akceptujących tylko serwisy o profilu pasującym do definicji (otwartego) repozytorium instytucjonalnego.

Analiza zalet tych dwóch teoretycznie alternatywnych podejść może prowadzić do wniosku, że rozwiązaniem oferującym wszystkie zalety jest system, który pozwala w ramach jednej bazy gromadzić różnego rodzaju dane i dzielić je na odrębne interfejsy/portale dopiero na etapie prezentacji. W systemie dLibra podejście takie jest możliwe w oparciu o odpowiednio zaprojektowaną strukturę kolekcji.

Dla przykładu - założmy, że chcemy zbudować jeden system gromadzący trzy rodzaje materiałów:

- historyczne zbiory biblioteczne,
- współczesne publikacje pracowników,
- współczesne publikacje objęte prawem autorskim, dostępne tylko na terenie instytucji.

W tym celu można w systemie dLibra zaprojektować następującą strukturę kolekcji:

- (M) Główna kolekcja systemu dLibra
 - (A) Biblioteka Cyfrowa Instytucji X
 - Kolekcje
 - Podkolekcje
 - (B) Repozytorium Instytucjonalne Instytucji X
 - Kolekcje
 - Podkolekcje
 - (C) Repozytorium Wewnętrzne Instytucji X
 - Kolekcje
 - Podkolekcje

Przykład ten jest złożony, ale jego celem jest oddanie szerokich możliwości zastosowania mechanizmów oferowanych przez system dLibra. Zazwyczaj w systemie dLibra podstawą dla Aplikacji Czytelnika jest główna kolekcja (M), w myśl "prostej zasady jedno wdrożenie = jeden interfejs WWW z kompletem zbiorów". Zamiast przy strukturze kolekcji zaprezentowanej powyżej jest taki, aby głównej kolekcji (M) nie towarzyszył dedykowany portal WWW, a zamiast tego funkcjonować powinny cztery portale bazujące odpowiednio na kolekcjach (A-C). Każdy z tych trzech portali może mieć swoją własną strukturę kolekcji i podkolekcji, a do tego powinna powstać spójna polityka udostępniania zbiorów obejmująca wszystkie portale funkcjonujące w ramach takiego wdrożenia systemu dLibra. Dzięki temu, że w systemie dLibra obiekt może należeć do więcej niż jednej kolekcji możliwe jest przypisanie jednej publikacji do kilku kolekcji widocznych w kilku różnych portalach.

W powyższym przykładzie można by przyjąć np. następujące zasady:

- zdigitalizowane zbiory biblioteczne przypisuję do kolekcji portali (A) i (C),
- wszystkie publikacje pracowników przypisuję do kolekcji portali (A), (B) i (C),
- współczesne publikacje objęte prawem autorskim, które mają być udostępniane wyłącznie na terenie instytucji przypisuję wyłącznie do kolekcji portalu (C).

Przy takich założeniach uzyskujemy:

- portal (A), zawierający informacje o wszystkich (bibliotecznych i repozytoryjnych) zbiorach, które Instytucja X posiada i może udostępniać publicznie bądź takich, które są utworami pracowników Instytucji X;
- portal (B), zawierający informacje o wszystkich utworach pracowników Instytucji X;
- portal (C), zawierający wszystko to co portale (A) i (B) oraz informacje o publikacjach, które udostępniamy tylko na terenie Instytucji X.

Dostęp do publikacji w portalach (A) i (B) może być autoryzowany w razie potrzeby. Dostęp do portalu (C) jest w całości ograniczony do zakresu adresów IP sieci lokalnej Instytucji X. Na terenie Instytucji X najlepiej korzystać z portalu (C). Poza instytucją, w zależności od preferencji użytkownik może korzystać z portali (A) lub (B). Ponadto portal (B) można promować i rejestrować jako repozytorium instytucjonalne.

Zobacz też:

- Pytanie ["Czy można łączyć biblioteki cyfrowe i repozytoria?"](#) w serwisie FBC "Pytania i odpowiedzi"

dLibra to system do budowy bibliotek cyfrowych, czy można więc na jego bazie tworzyć repozytoria instytucjonalne?

Tak, dLibra może być z powodzeniem wykorzystywana do budowy repozytoriów instytucjonalnych. Posiada wszystkie niezbędne funkcje związane z gromadzeniem i udostępnieniem obiektów cyfrowych i metadanych, wspiera również samoarchiwizację / self-archiving.

Istotną cechą repozytoriów instytucjonalnych jest możliwość samodzielnego deponowania prac przez autorów (tzw. self-archiving, samoarchiwizacja). Czy dLibra to wspiera?

Tak, jest to szczegółowo opisane w dokumentacji w rozdziale [04. Self-archiving \(przekazywanie obiektów cyfrowych przez autorów\)](#). Możliwe jest również integrowanie systemu dLibra z zewnętrznymi systemami uwierzytelniania użytkowników :[03. Integracja z systemami Single Sign-On](#).

Chcę udostępniać materiały na otwartych licencjach oraz materiały, do których dostęp mogą mieć tylko pracownicy mojej instytucji - wszystko to są publikacje pracowników mojej instytucji. Czy powinienem zbudować dwa odrębne repozytoria, czy też warto i da się to połączyć w jeden system?

Da się takie dwa repozytoria uruchomić jako jeden system informatyczny z dwoma portalami WWW. Można to zrealizować w sposób analogiczny do łączenia biblioteki cyfrowej i repozytorium, co opisano w odpowiedzi na pytanie: [Czy w ramach jednej instytucji biblioteka cyfrowa i repozytorium to powinny być odrębne inicjatywy czy też lepiej je połączyć?](#)

Zalety rozdzielenia takich repozytoriów na dwa oddzielne interfejsy WWW to przede wszystkim możliwość promowania przynajmniej części publikacji - tych które są otwarte - jako w 100% otwartego repozytorium, tzn. zawierającego tylko otwarte publikacje. Jednak należy się zastanowić, w jakim celu takie 100% otwarcie jest wymagane. Dla przykładu serwis [OpenDOAR](#), jeden z najpopularniejszych na świecie agregatorów metadanych z otwartych repozytoriów, jako [najczęstsze kryteria odrzucenia zgłoszeń rejestracyjnych](#) podaje:

- Serwis jest regularnie niedostępny
- Serwis jest czasopismem, a nie repozytorium
- Serwis nie zawiera żadnych materiałów dostępnych na zasadach Open Access
- Serwis zawiera tylko informacje bibliograficzne i ewentualnie linki do zewnętrznych serwisów, nie zawiera pełnych treści
- Serwis to katalog biblioteczny albo serwis z e-bookami, które dostępne są tylko w sieci lokalnej
- Serwis wymaga logowania (nawet jeżeli jest ono bezpłatne), żeby uzyskać dostęp do teoretycznie otwartych materiałów
- Serwis udostępnia treści na zasadach komercyjnych - płatny dostęp

Jak widać OpenDOAR nie eliminuje repozytoriów, które zawierają równocześnie materiały dostępne w sposób otwarty i dostępne w sposób ograniczony (np. tylko dla pracowników, tylko na terenie instytucji).

Jak w systemie dLibra mogę połączyć bibliotekę cyfrową i repozytorium instytucjonalne?

Opisano to w odpowiedzi na pytanie [Czy w ramach jednej instytucji biblioteka cyfrowa i repozytorium to powinny być odrębne inicjatywy czy też lepiej je połączyć?](#)

Zależy mi na widoczności materiałów z mojego repozytorium w Google Scholar. Czy dLibra wspiera to w jakiś sposób?

Tak, dLibra spełnia [wymagania serwisu Google Scholar](#) dotyczące indeksowania, m.in. na stronie prezentującej metadane poszczególnych obiektów w kodzie HTML strony w sekcji <HEAD> zawarte są stosowne znaczniki z metadanymi publikacji. dLibra umożliwia też pobranie metadanych w formatach RIS i BibTeX, co ułatwia ich wykorzystanie w pracy naukowej. Na stronach Google Scholar system dLibra nie jest wymieniony jako zalecany (wymienione są tylko dwa popularne systemy repozytoryjne open-source i jeden usługodawca oferujący repozytoria hostowane), nie oznacza to jednak że dLibra nie jest zgodna z wymaganiami Google Scholar.

Poza wsparciem ze strony systemu dLibra niezwykle ważne jest, żeby pliki z pracami naukowymi publikowane w systemie dLibra były przygotowane pod kątem wymagań Google Scholar. Absolutne minimum to publikowanie plików w formacie PDF, z warstwą tekstową (nie samych skanów). Więcej informacji można znaleźć w [serwisie informacyjnym Google Scholar](#).

Jak szybko materiały opublikowane w repozytorium na dLibrze znajdą się w Google Scholar?

Trudno podać jednoznaczny i stały termin - wszystko zależy od zasad działania Google i tego jak często aktualizują oni dane z konkretnego repozytorium. Ale należy się spodziewać, że odpowiednio przygotowane publikacje powinny być widoczne w przeciągu kilku-kilkunastu dni. Na przełomie kwietnia i maja 2013 przeprowadziliśmy mały eksperyment w tym zakresie. Więcej informacji można znaleźć tutaj:

<http://dlab.psnc.pl/2013/05/06/repozytorium-instytucjonalne-na-systemie-dlibra-i-google-scholar-maly-eksperyment/>

Zależy mi na widoczności materiałów z mojego repozytorium w Google Books. Czy dLibra wspiera to w jakiś sposób?

Serwis Google Books jest przede wszystkim serwisem dla wydawców, zainteresowanych sprzedażą książek. Ponadto dostępne są tam kopie skanów, których powstanie finansowała firma Google. Instytucje, które chcą udostępnić swoje zbiory w Google Books powinny skontaktować się z firmą Google. Po ustaleniu zasad współpracy i określeniu wymagań technicznych (nie są one dostępne publicznie), PCSS może zaimplementować niezbędne mechanizmy w systemie dLibra.

W mojej instytucji funkcjonuje wydawnictwo, które wydaje i sprzedaje przez internet książki, skrypty i czasopisma powstające u nas. Czy mogę w jakiś sposób uwzględnić te zasoby w repozytorium instytucjonalnym?

Najlepszym rozwiązaniem jest podpisanie umowy z wydawnictwem, która umożliwi przekazanie publikacji do biblioteki cyfrowej (ewentualnie po pewnym okresie embargo od momentu opublikowania tej publikacji).

Jeżeli to nie jest możliwe, to najprostszym sposobem technicznym jest wykorzystanie mechanizmu [publikacji linkujących dostępnego od wersji 5.5 systemu dLibra](#). Dzięki temu do systemu dLibra można wprowadzić:

- opis publikacji dostępnej na stronach wydawnictwa,
- link do stron wydawnictwa na który ma zostać przekierowany użytkownik po próbie otwarcia treści publikacji w systemie dLibra,
- linki do plików publikacji, które system dLibra powinien zaindeksować na potrzeby wyszukiwania pełnotekstowego.

Jeżeli serwis WWW wydawnictwa zostanie skonfigurowany tak, aby umożliwiać serwerowi systemu dLibra (np. na podstawie stałego adresu IP) dostęp przy pomocy protokołu HTTP (czyli analogicznie do czytelników korzystających z przeglądarki WWW) np. do plików PDF czy HTML z tekstem artykułów, uzyska się efekt gdzie w bibliotece cyfrowej/repozytorium można wyszukiwać publikację na podstawie jej opisu i treści, a czytelnik przy próbie otwarcia treści jest przekierowywany na właściwą stronę WWW wydawnictwa, gdzie np. może dokonać zakupu artykułu.

Autorzy z mojej instytucji publikują swoje prace w międzynarodowych wydawnictwach, które udostępniają artykuły odpłatnie. Czy mogę w jakiś sposób uwzględnić te zasoby w repozytorium instytucjonalnym?

Publikacje z wydawnictw zewnętrznych (w tym zagranicznych) można włączyć w sposób analogiczny do tego opisanego w odpowiedzi na pytanie [W mojej instytucji funkcjonuje wydawnictwo, które wydaje i sprzedaje przez internet książki, skrypty i czasopisma powstające u nas. Czy mogę w jakiś sposób uwzględnić te zasoby w repozytorium instytucjonalnym?](#).

Czy mogę w repozytorium gromadzić również informacje na temat publikacji, których treść nie jest w żaden sposób dostępna on-line?

W systemie dLibra jedyną formą publikacji bez treści (i bez linku do treści) są na chwilę obecną publikacje planowane. Informacje o publikacjach, których treść nie jest w żaden sposób dostępna on-line powinny być gromadzone w osobnej bazie bibliograficznej. Gdyby była konieczność uwzględnienia takich danych w bibliotece cyfrowej, proponowanym sposobem jest import metadanych z bazy bibliograficznej do biblioteki cyfrowej i umieszczenie wraz z tymi danymi linku zwrotnego do właściwego rekordu w bazie. Jednak taka praktyka może się spotkać z rozczarowaniem ze strony użytkowników oczekujących po repozytorium / bibliotece cyfrowej dostępu do pełnego tekstu, a nie jedynie informacji bibliograficznej. Dlatego też zalecanym rozwiązaniem jest linkowanie z bazy bibliograficznej do repozytorium (oczywiście dla tych publikacji, które w repozytorium są dostępne), a nie odwrotnie.

W mojej instytucji pracuje kilka tysięcy pracowników naukowych. Planujemy gromadzić wszystkie publikacje pracowników w repozytorium i być może wprowadzić również mandat Open Access. Czy kilka tysięcy użytkowników równocześnie wprowadzających swoje prace nie spowoduje przeciążenia systemu?

Liczba nowych publikacji w przypadku dużych wdrożeń systemu dLibra sięga od kilku (Wielkopolska Biblioteka Cyfrowa) do nawet dziesięciu tysięcy miesięcznie (Jagiellońska Biblioteka Cyfrowa). Są to w większości skany materiałów bibliotecznych, czyli publikacje bardziej obciążające serwery biblioteki cyfrowej, niż pliki typowe artykuły naukowe - pliki PDF z warstwą tekstową i ewentualnie kilkoma ilustracjami. Przygotowanie repozytorium instytucjonalnego na dużą liczbę użytkowników wprowadzających swoje publikacje może obejmować następujące aspekty:

- Zapewnienie niezbędnej infrastruktury serwerowej, wykorzystującej możliwości skalowania systemu dLibra (patrz [13. Skalowanie systemu dLibra](#)). Szczególnie istotne może być tutaj wydzielenie usługi indeksującej treść i metadane dokumentów na osobny serwer oraz zapewnienie szybkiego funkcjonowania bazy danych, z której korzysta system dLibra.
- Przygotowanie kopii portalu WWW biblioteki cyfrowej dedykowanej na potrzeby wprowadzania publikacji przez autorów metodą samoarchiwizacji. Taka kopia portalu WWW może być uruchomiona na osobnym serwerze, a jej funkcjonalność może zostać ograniczona tylko do dostępu do konta użytkownika, dodawania nowych i przeglądania już dodanych publikacji. Dzięki temu nawet w przypadku niezwykle dużego ruchu generowanego przez autorów, czytelnicy będą obsługiwani przez niezależny portal.
- Zapewnienie odpowiedniej liczby moderatorów publikacji, którzy będą nadzorować poszczególne grupy autorów i wprowadzane przez nich publikacje np. opiekunowie poszczególnych instytutów. Tego typu organizacja jest możliwa poprzez przypisywanie poszczególnym autorom ich katalogów domowych oraz opiekunów tych katalogów domowych. Możliwe jest przypisywanie wielu autorów do jednego moderatora, jak i wielu moderatorów do jednej grupy autorów.

- Zapewnienie zupełnie awaryjnego rozwiązania na wypadek przeciążenia systemu. Jeżeli np. ze względu na rozporządzenie władz uczelni wszyscy pracownicy naukowcy będą próbować na ostatnią chwilę wprowadzić swoje publikacje, system może ulec przeciążeniu niezależnie od tego jakie kroki techniczne wcześniej podejmiemy. Na taki wypadek warto mieć przygotowany plan awaryjny polegający np. na posiadaniu odpowiednio pojemnego adresu e-mail, który może pełnić rolę bufora w czasie gdy system autoarchiwizacji jest niedostępny.