


WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

# Projekt techniczny

---

## System archiwizacji

(wersja z wybranymi fragmentami z dokumentacji istotnymi z punktu  
widzenia systemu Repozytorium Cyfrowe BN)

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

## Spis treści

1	Cel dokumentu.....	5
2	Słownik .....	5
3	Informacje ogólne .....	7
3.1	Diagram poglądowy .....	7
3.2	Przypadki użycia .....	9
3.3	Obiekt jako podstawowa encja biznesowa .....	10
3.3.1	Obiekt .....	10
3.3.2	Kontent.....	11
3.3.3	Segmentacja .....	12
3.4	Wolumetria danych .....	12
4	Procesy biznesowe .....	13
4.1	Archiwizacja obiektu (proces P1) .....	13
4.2	Ponowna archiwizacja obiektu (P2).....	13
4.3	Archiwizacja metadanych (P3).....	14
4.4	Weryfikacja stanu kasety (P4) .....	15
4.5	Migracja kasety (P5).....	16
4.6	Przywrócenie zawartości obiektu (P6) .....	17
4.7	Pozostałe procesy biznesowe (P7) .....	18
4.8	Proces przywrócenia bazy danych (P8).....	19
5	Architektura .....	21
6	Wymagania funkcjonalne Systemu archiwizacji .....	21
6.1	Elementy wspólne .....	21
6.2	Archiwum .....	21
6.3	Usługi pomocnicze .....	21
6.3.1	Monitorowanie obiektów do zarchiwizowania i inicjowanie archiwizacji .....	21
6.3.2	Uruchamianie procesu archiwizacji metadanych/bazy danych .....	22
6.3.3	Monitorowanie obiektów do przywrócenia i inicjowanie przywrócenia .....	22
6.3.4	Usuwanie plików z przywróconego lub niezredukowanego obiektu .....	23
6.4	Archiwizator .....	25
6.4.1	Przyjęcie żądania zarchiwizowania obiektu .....	26
6.4.2	Przyjęcie żądania zarchiwizowania metadanych/bazy danych .....	26

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

6.4.3	Przygotowywanie paczki archiwalnej z obiektem.....	27
6.4.4	Przygotowywanie paczki archiwalnej ze zrzutem metadanych/bazy danych ....	34
6.4.5	Przekazanie żądania archiwizacji paczki do instancji Archiwum.....	36
6.4.6	Odebranie informacji o statusie zakończenia zadania archiwizacji paczki w instancji Archiwum wraz z podjęciem kolejnych kroków.....	37
6.4.7	Obsługa zakończenia procesu archiwizacji paczki z obiektem .....	38
6.4.8	Obsługa zakończenia procesu archiwizacji paczki z metadanymi.....	38
6.4.9	Przyjęcie żądania przywrócenia obiektu / udostępnienia paczki archiwalnej....	38
6.4.10	Przyjęcie od instancji Archiwum żądania odczytu paczki archiwalnej .....	40
6.4.11	Odebranie informacji o statusie zakończenia zadania odczytu paczki archiwalnej wraz z podjęciem kolejnych kroków .....	41
6.4.12	Monitorowanie statusu przywrócenia obiektu .....	42
6.5	Interfejs API Systemu archiwizacji .....	42
6.5.1	Specyfikacja interfejsu API .....	42
6.5.2	Uwierzytelnianie .....	43
6.5.3	Zapisywanie zdarzeń wywołania interfejsu .....	43
6.6	Przywracacz.....	44
6.6.1	Przywrócenie plików do obiektu w Repozytorium .....	44
6.7	Weryfikator .....	47
7	Wymagania funkcjonalne zmian w Repozytorium .....	47
7.1	Zmiany encji obiekt .....	47
7.1.1	Status obiektu w kontekście archiwizacji .....	47
7.1.2	Blokowanie obiektu (tzw. LOCK) .....	50
7.1.3	Status streama w kontekście archiwizacji .....	50
7.1.4	Modyfikacja obiektu w kontekście archiwizacji .....	51
7.1.5	Konfiguracja okresów karencji .....	52
7.2	Pobieranie całej publikacji.....	52
7.3	Zamówienia na pliki.....	52
7.3.1	Budowanie zamówienia z poziomu szczegółów obiektu .....	53
7.3.2	Budowanie zamówienia z poziomu listy obiektów.....	54
7.3.3	Budowanie zamówienia a uprawnienia.....	55
7.3.4	Zmiany w uprawnieniach .....	57
7.3.5	Zarządzanie zamówieniami .....	57

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

7.3.6	Limit miejsca na zamówienia użytkownika.....	59
7.3.7	Struktura folderów i nazewnictwo plików .....	60
7.3.8	Realizacja zamówienia.....	60
7.3.9	Archiwizacja zamówienia .....	61
7.3.10	Pobieranie dużych plików .....	62
7.4	Pobieranie plików z obiektu .....	62
7.5	Ponowna regeneracja pochodnych.....	64
7.6	Obracanie skanów.....	65
7.7	Wysyłanie obiektu do procesu OCR.....	66
7.8	Zmiana interfejsu API .....	68
7.9	Przycisk do przywracania obiektu.....	68
7.10	Funkcjonalność współdzielenia i wydzielenia contentów .....	68
8	Monitorowanie usług Systemu archiwizacji .....	69
9	Procedura wyłączenia i włączenia Systemu archiwizacji .....	69
10	Schemat i opis struktury bazy danych .....	69
11	Założenia.....	70
12	Produkty końcowe projektu.....	70

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

## 1 Cel dokumentu

Celem niniejszego dokumentu jest opisanie projektu technicznego projektowanego Systemu archiwizacji wraz ze specyfikacją zmian dot. rozbudowy i modyfikacji systemu Repozytorium Cyfrowe Biblioteki Narodowej związanych z wykorzystaniem Systemu archiwizacji.

## 2 Słownik

**Aplikacja webowa do konfiguracji biblioteki taśmowej** – istniejąca aplikacja do konfiguracji biblioteki taśmowej.

**Baza danych** – (występuje na diagramie poglądowym w rozdziale 3.1) zasób służący do przechowywania informacji o obiektach, w tym metadanych.

**Biblioteka taśmowa** – urządzenie pozwalające na zapis danych na taśmach magnetycznych.

**HCP** – (występuje na diagramie poglądowym w rozdziale 3.1) sprzętowy klaster komputerów zapewniający funkcjonalność obiektowego przechowywania danych.

**Klucz stream'a** – oznaczenie/klasyfikacja pliku występująca w Repozytorium przy każdym stream'ie znajdującym się w ramach obiektu w kontencie. Jest nazywany w niniejszym opracowaniu również typem lub rodzajem stream'a/pliku w kontencie.

**Macierz** – (występuje na diagramie poglądowym w rozdziale 3.1) zasób służący do przechowywania zawartości wymagającej szybkiego dostępu w celach prezentacyjnych (kafelki, miniatury).

**Macierz pomocnicza** – (występuje na diagramie poglądowym w rozdziale 3.1) zasób służący do przechowywania plików wykorzystywanych przez System archiwizacji, np. w Archiwum do przechowywania odczytywanych/zapisywanych paczek archiwalnych.

**Migracja kasety** – operacja mająca na celu zapisanie każdej paczki archiwalnej znajdującej się na „migrowanej” kasecie, w dokładnie tej samej wersji, na inną kasetę. Do zapisu na nową kasetę może zostać wykorzystany plik z paczką archiwalną pochodzący z „migrowanej” kasety (o ile weryfikacja jego sumy kontrolnej jest pozytywna), lub też plik może pochodzić z innej kasety.

**Napęd** – fizyczne urządzenie służące do odczytu danych z kasety magnetycznej oraz zapisu danych na kasetę magnetyczną.

**Obiekt** – zasób w systemie będący cyfrową wersją **Pozycji bibliotecznej**, pochodzący z postaci analogowej (poprzez digitalizację) lub cyfrowej (tzw. BornDigital, jeżeli został przekazany od razu w formie cyfrowej). Składa się z metadanych i zawartości (kontentów).

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

**Pozycja biblioteczna (dokument)** – obiekt analogowy typu książka, czasopismo, druk lub inna pozycja z księgozbioru Biblioteki Narodowej.

**Przywrócenie plików obiektu** – proces wywoływany przez interfejs Systemu archiwizacji mający na celu przywrócenie do obiektu w Repozytorium stream'ów (plików) z kontentów.

**Ramię robota** – fizyczne urządzenie służące do przemieszczania kaset ze slotów do napędu oraz z napędu do slotu.

**Redukcja obiektu** – usunięcie plików z obiektu po jego zarchiwizowaniu. Wybór plików do usunięcia powinien nastąpić na podstawie konfiguracji

**Repozytorium Cyfrowe BN** – (dalej: Repozytorium) system informatyczny wykorzystywany w Bibliotece Narodowej do wsparcia procesu digitalizacji, przechowywania oraz udostępniania pozycji bibliotecznych.

**Slot** – półka znajdująca się w bibliotece taśmowej, służąca do przechowywania kaset.

**System archiwizacji** – (występuje na diagramie poglądowym w rozdziale 3.1) produkt niniejszego projektu, narzędzie zapewniające przede wszystkim funkcjonalność archiwizacji oraz przywracania obiektów do Repozytorium.

**Taśma magnetyczna** – (dalej: taśma lub kaseeta) nośnik danych wykorzystywany przez Bibliotekę taśmową. Przykładowymi rozmiarami taśmy magnetycznych: 1,5TB (klasa LTO-5) lub 6TB (klasa LTO-7).

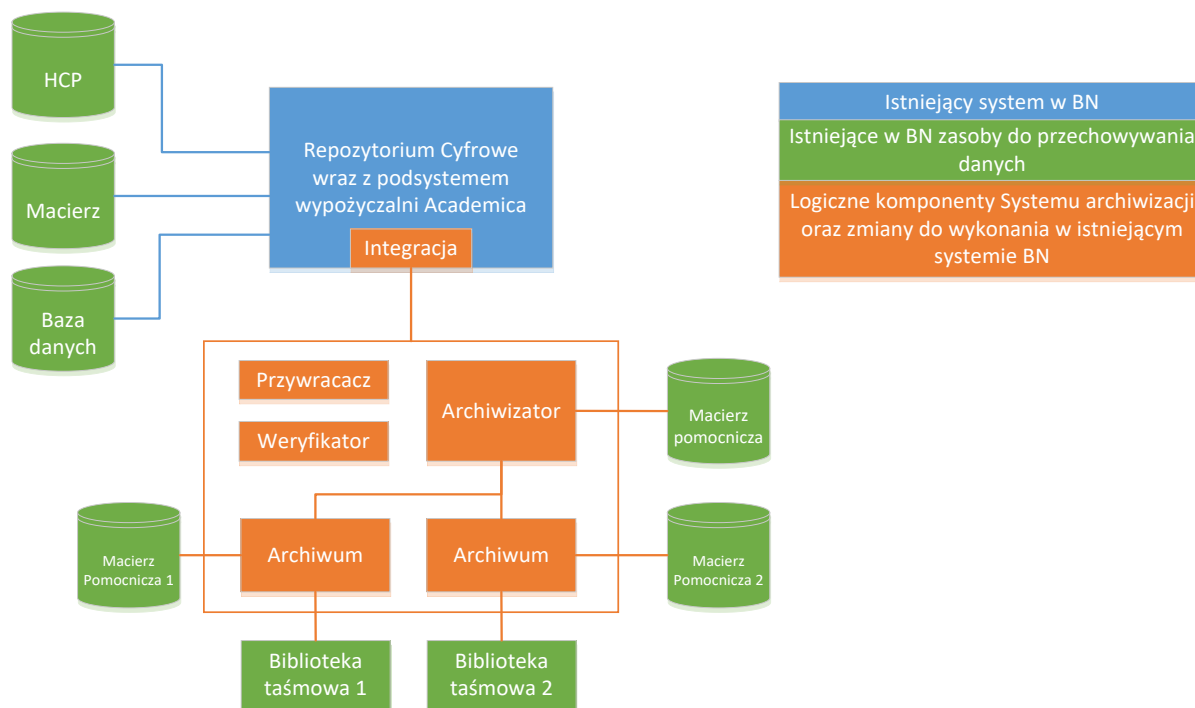
**Udostępnienie paczki archiwalnej** – proces wywoływany przez interfejs Systemu archiwizacji mający na celu udostępnienie na macierz dyskową pliku paczki archiwalnej z obiektem lub ze zrzutem metadanych bez podejmowania jakichkolwiek dalszych czynności w szczególności przywrócenia danych do Repozytorium.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

### 3 Informacje ogólne

#### 3.1 Diagram poglądowy

Poniżej zamieszczono poglądowy diagram przedstawiający system Repozytorium Cyfrowe wraz z projektowanym Systemem archiwizacji.



W ramach Systemu archiwizacji można wyszczególnić następujące komponenty logiczne:

##### 1. Archiwizator

Jeden z głównych komponentów Systemu archiwizacji, którego głównymi zadaniami są:


- Przygotowywanie danych do procesu archiwizacji oraz zlecenie Archiwum zapisania archiwizowanych danych na kasetach magnetycznych.
- Zlecenie odczytu danych z Archiwum oraz przygotowanie ich do procesu przywracania do Repozytorium.

Korzysta z macierzy pomocniczej jako zasobu do tymczasowego przechowywania plików podczas wykonywania powierzonych mu zadań.

System archiwizacji będzie zawierał jedną instancję Archiwizatora.

##### 2. Archiwum

Jeden z głównych komponentów Systemu archiwizacji, którego głównym zadaniem jest zapis oraz odczyt danych z kaset magnetycznych z wykorzystaniem biblioteki

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

taśmowej. Korzysta z macierzy pomocniczej jako zasobu do tymczasowego przechowywania plików podczas wykonywania powierzonych mu zadań. System archiwizacji będzie zawierał od jednej do wielu instancji Archiwów.

### 3. Przywracacz

Funkcjonalność, której głównym zadaniem jest przywrócenie plików zarchiwizowanych obiektów do Repozytorium. Logicznie ta funkcjonalność będzie częścią Archiwizatora.


### 4. Weryfikator

Funkcjonalność, której głównym zadaniem jest weryfikacja stanu kaset z danymi przechowywanych w ramach Archiwów. Logicznie ta funkcjonalność będzie częścią Archiwum.

### 5. Integracja

Komponent obejmujący niezbędne zmiany w Repozytorium oraz dodatkowe usługi udostępniające Repozytorium funkcjonalności zapewnione przez System archiwizacji.



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

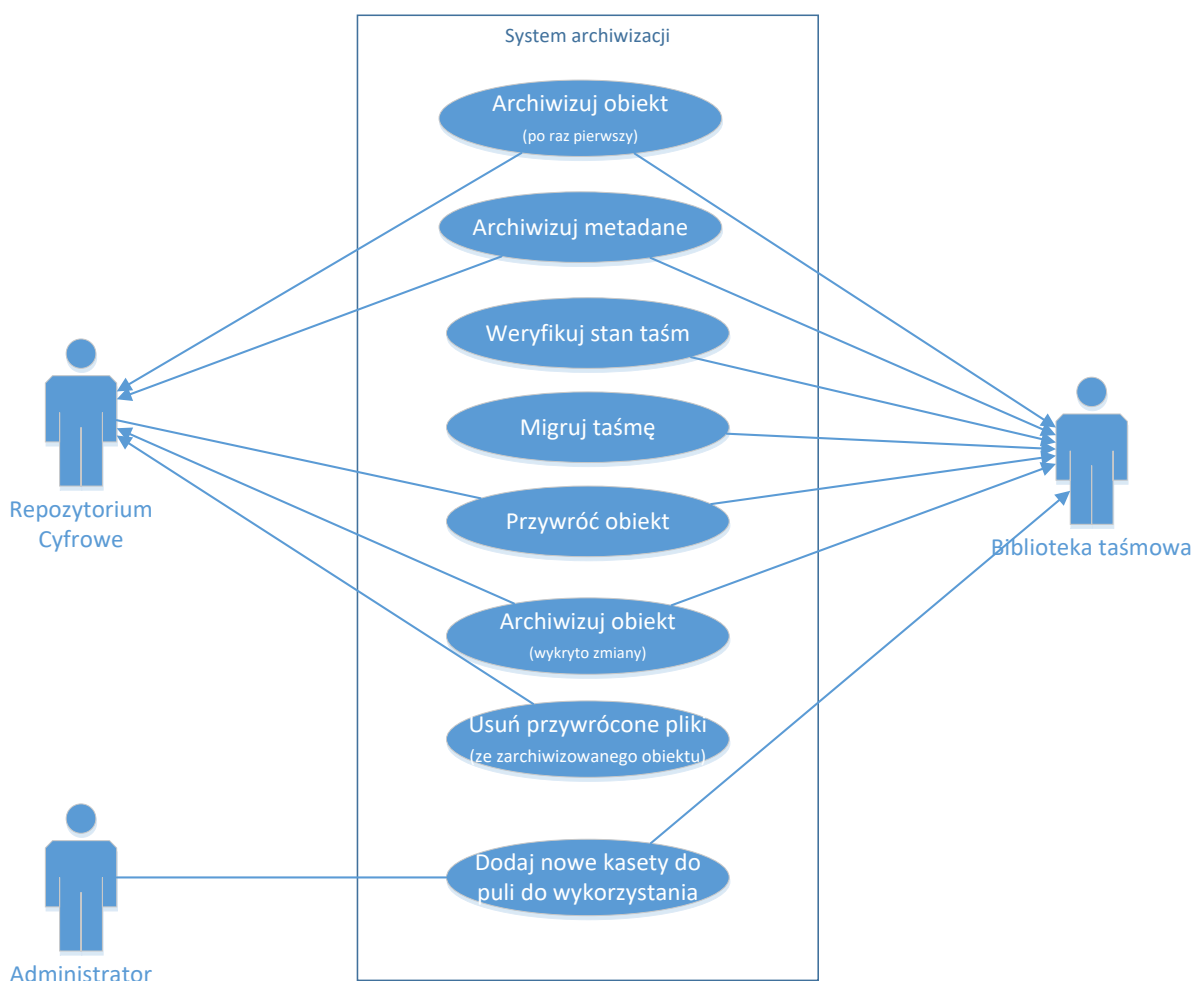
### 3.2 Przypadki użycia

Poniżej zamieszczono następujące diagramy przypadków użycia Systemu archiwizacji:

1. Diagram przypadków użycia Systemu archiwizacji.
2. Diagram przypadków użycia zmian w Repozytorium w zakresie obsługi zamówień na pliki.

Z uwagi na to, że System archiwizacji będzie pełnił rolę niezależnego mechanizmu odpowiedzialnego za archiwizację obiektów i metadanych, część przypadków użycia będzie wywoływana przez wewnętrzne zdarzenia w Systemie archiwizacji. Jako że te przypadki użycia będą miały wpływ na Repozytorium, zostały one uwzględnione na poniższym diagramie i oznaczone strzałką skierowaną do Repozytorium.

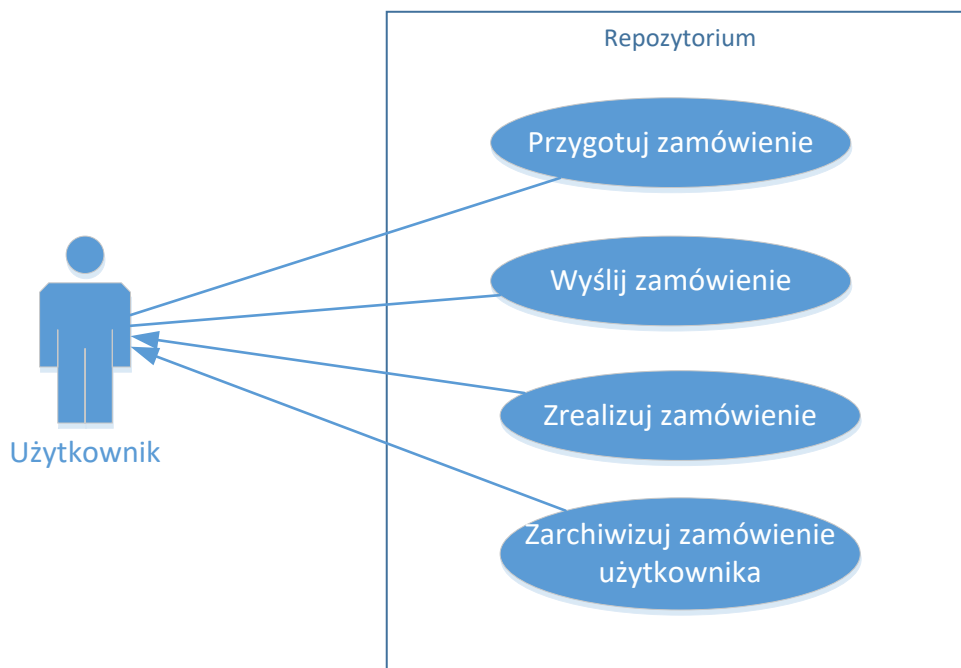
Poniżej zamieszczono diagram przypadków użycia Systemu archiwizacji.



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Poniżej zamieszczono diagram przypadków użycia zmian w Repozytorium w zakresie obsługi zamówień na pliki.

Z uwagi na to, że obsługa zamówień użytkowników będzie asynchronicznie realizować pewne zadania, część zaprezentowanych przypadków użycia będzie wywoływana przez wewnętrzne zdarzenia. Te przypadki użycia będą miały wpływ użytkownika, dlatego zostały one uwzględnione na poniższym diagramie i oznaczone strzałką skierowaną do użytkownika.



### 3.3 Obiekt jako podstawowa encja biznesowa

W niniejszym rozdziale opisano aktualną strukturę obiektu w Repozytorium.

#### 3.3.1 Obiekt

Obiekt jest odzwierciedleniem w systemie konkretnego utworu/produktu wchodzącego w skład zasobów Biblioteki Narodowej. Każdy obiekt składa się z:

1. metadanych – istotnych informacji określających cechy danego utworu/produktu,
2. kontentów – zawierających zdigitalizowaną wersję utworu/produktu.

Obiekt może być powiązany relacją z wieloma innymi obiektami. Występuje kilka rodzajów relacji m.in. „następny obiekt”, „poprzedni obiekt”, „cytowany obiekt”.

Obiekt może posiadać wiele obiektów podrzędnych (tzw. podobiekty).

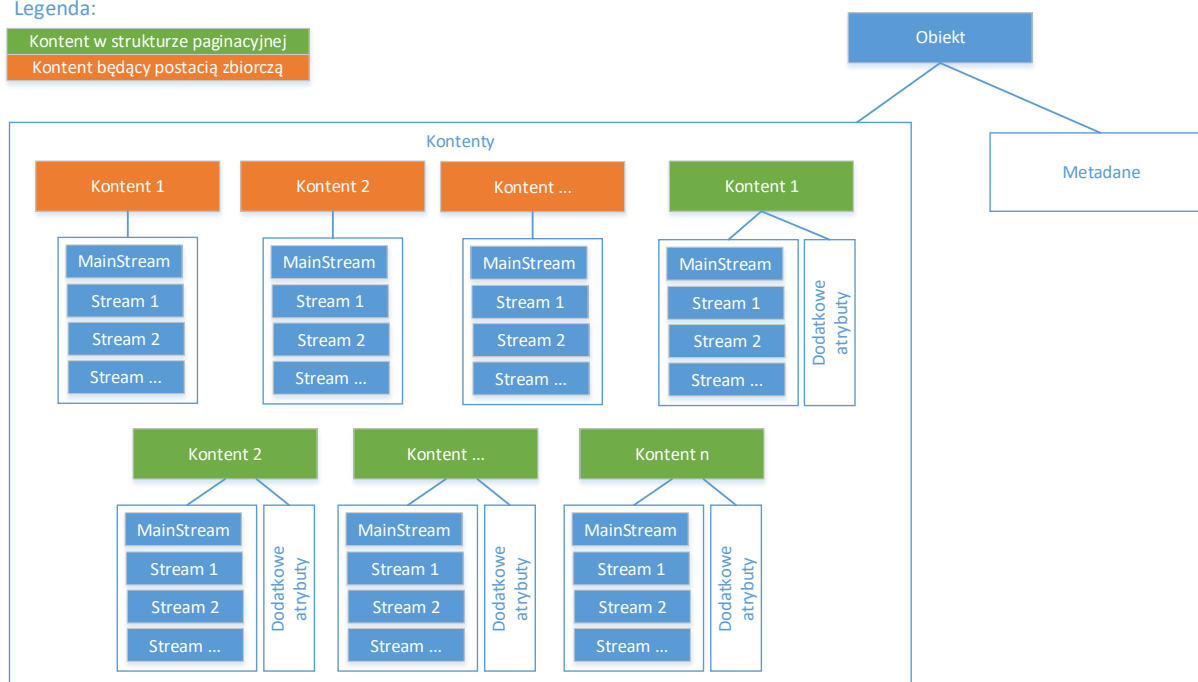
W strukturze obiektu nie są przechowywane miniatury/kafle wygenerowane na jego podstawie – są one jedynie z nim powiązane.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Poniżej zamieszczono koncepcyjny diagram struktury obiektu.

Legenda:

Kontent w strukturze paginacyjnej
Kontent będący postacią zbiorczą



### 3.3.2 Kontent

Każdy kontent może składać się z jednego lub wielu tzw. stream'ów, z czego pierwszy dodany stream jest zawsze tzw. mainStream'em, kolejne streamy są formami pochodnymi mainStreamu, zazwyczaj utworzonymi automatycznie przez system. Każdy stream ma swój jasno określony typ, np. tekst, download\_alto (plik pochodzący z procesu ALTO), download\_fullJPG (wersja JPG tworzona automatycznie dla plików graficznych w formatach bez kompresji).

W szczególnych przypadkach kontent może nie mieć stream'ów, wówczas kontent jest wizualizowany w Repozytorium z użyciem ikony „duszka”. Kontenty takie mogą powstać np. w przypadkach, gdy kontroler stwierdzi brak jednej strony w zdigitalizowanej wersji utworu – tworzy wtedy pusty kontent, który zostanie później uzupełniony przez odpowiednią osobę.

Kontenty, pod względem logicznym, dzielą się na:

1. postaci zbiorcze, ich kolejność zawsze jest zgodna z kolejnością dodawania ich do systemu i nie może być modyfikowana przez użytkownika,
2. kontenty znajdujące się w strukturze paginacyjnej, ich kolejność może być dowolnie zmieniana przez użytkownika; kontenty zbiorcze zawierają dodatkowe atrybuty takie jak układ strony, numer strony, opis strony.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Każdy kontent znajdujący się w strukturze paginacyjnej może posiadać powiązane z nimi kontenty alternatywne. Kontenty te zawierają zazwyczaj tę samą informację, co kontent podstawowy, są jednak ich alternatywną wersją (np. strona książki zawierająca dodatkowo na marginesie ręczne dopiski autora).

W ramach obiektu każdorazowo wybrany musi być dokładnie jeden kontent główny, który używany jest do prezentacji danego obiektu w wynikach wyszukiwania (np. dla książki kontentem głównym zazwyczaj będzie skan okładki). Kontentem głównym nie może być kontent oznaczony jako kontent ukryty.

W szczególnych przypadkach kontent może być przypisany do więcej niż jednego obiektu.

### 3.3.3 Segmentacja

Dla kontentów, które są w postaci graficznej system ma możliwość przeprowadzenia automatycznej segmentacji. Dotyczy to np. skanów stron gazet, które automatycznie dzielone są na artykuły.

W sytuacji, gdy kontent został poddany segmentacji do obiektu dodawany jest dodatkowy podział logiczny na artykuły, a dla każdego artykułu wydzielane są sekcje. Zarówno artykuły jak i sekcje posiadają powiązane metadane opisujące ich zawartość, np. każda sekcja wskazuje konkretne kontenty i koordynaty danej sekcji w plikach graficznych tych kontentów.

## 3.4 Wolumetria danych

Aktualnie produkcyjna baza danych metadanych zrzuconych w postaci SQL zajmuje około 600GB.

Aktualnie wykorzystywane taśmy klasy LTO-5 posiadają pojemność 1,5TB. Planowane jest zastosowanie taśm klasy LTO-7 o pojemności 6TB.

Aktualnie wykorzystywane taśmy klasy LTO-5 posiadają wg informacji od producenta maksymalną prędkość zapisu/odczytu na poziomie 140MB/s. Planowane jest zastosowanie taśm klasy LTO-7 charakteryzujących się wg informacji od producenta maksymalną prędkością zapisu/odczytu na poziomie 300MB/s.

Szacuje się, że na jednej taśmie będzie mieściło się od 2 do 1000 paczek archiwalnych.

Dzienny przyrost danych: średni dzienny wpływ oryginałów do Repozytorium jest na poziomie 0,5 TB, maksymalny zaobserwowany 1,2 TB. Średnia liczona jest razem z weekendami, choć wówczas wykorzystanie systemu przez użytkowników jest mniejsze.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

## 4 Procesy biznesowe

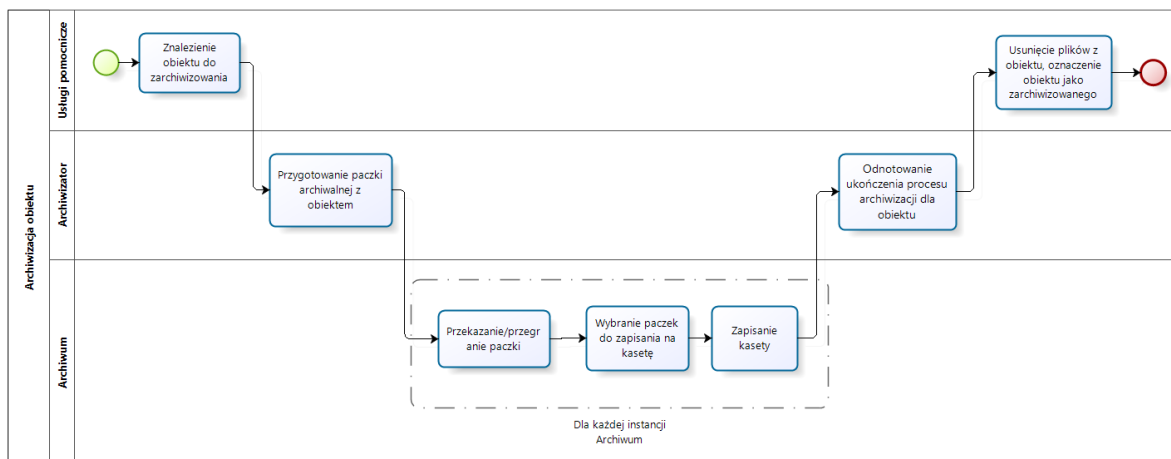
Niniejszy rozdział opisuje procesy biznesowe związane z Systemem archiwizacji.

### 4.1 Archiwizacja obiektu (proces P1)

Poniżej zamieszczono metrykę procesu.

Nazwa:	Archiwizacja obiektu
Kod:	P1
Cel:	Zarchiwizowanie obiektu we wszystkich instancjach Archiwum.
Warunki początkowe:	Dostęp do bazy Repozytorium z danymi obiektów
Warunki końcowe:	Paczka archiwalna z obiektem zapisana na kasecie. Usunięte pliki z obiektu.
Wykorzystywane zasoby	Macierz pomocnicza aby przechować dane obiektów i przygotować paczkę archiwalną. Biblioteka taśmowa aby zapisać paczkę archiwalną na kasetę.
Uwagi:	Brak.

Poniżej zamieszczono diagram procesu.



Powered by  
**bizagi**  
Modeler

### 4.2 Ponowna archiwizacja obiektu (P2)

Poniżej zamieszczono metrykę procesu.

Nazwa:	Ponowna archiwizacja obiektu
--------	------------------------------

Projekt współfinansowany w ramach Programu Operacyjnego Polska Cyfrowa z Europejskiego Funduszu Rozwoju Regionalnego i budżetu państwa na podstawie Umowy o dofinansowanie nr POPC.02.01.00-00-0043/16-00



Fundusze  
Europejskie  
Polska Cyfrowa

**omnis**<sup>BN</sup>

Unia Europejska  
Europejski Fundusz  
Rozwoju Regionalnego

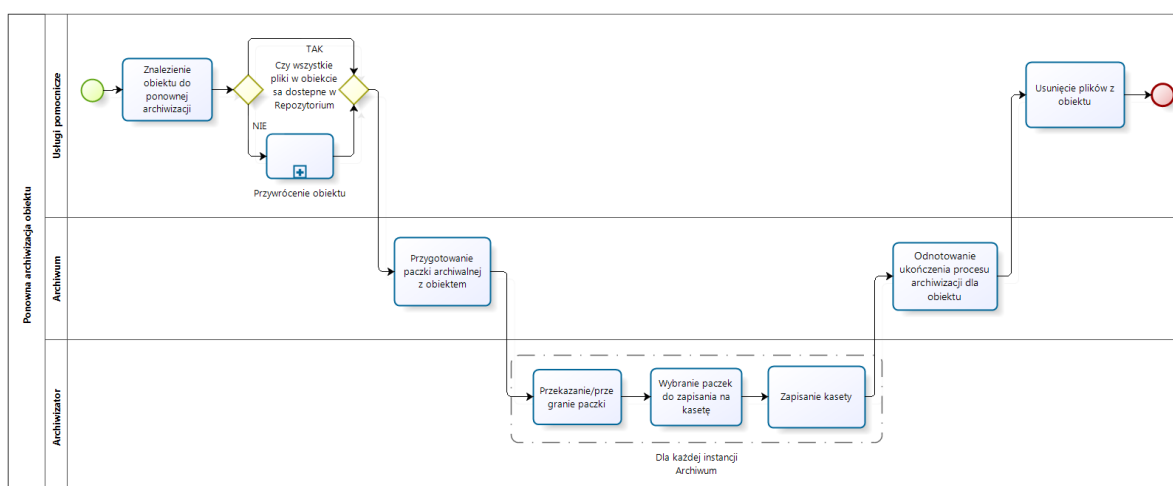


Strona 13 z 70

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Kod:	P2
Cel:	Ponowne zarchiwizowanie obiektu we wszystkich instancjach Archiwum.
Warunki początkowe:	Dostęp do bazy Repozytorium z danymi obiektów
Warunki końcowe:	Paczka archiwalna z obiektem zapisana na kasecie. Usunięte pliki z obiektu.
Wykorzystywane zasoby	Macierz pomocnicza aby przechować dane obiektów i przygotować paczkę archiwalną. Biblioteka taśmowa aby zapisać paczkę archiwalną na kasetę.
Uwagi:	Brak.

Poniżej zamieszczono diagram procesu.



Prepared by  
**bizagi**  
Modeler

### 4.3 Archiwizacja metadanych (P3)

Poniżej zamieszczono metrykę procesu.

Nazwa:	Archiwizacja metadanych
Kod:	P3
Cel:	Zarchiwizowanie metadanych w postaci paczki archiwalnej zawierającej zrzuty baz danych we wszystkich instancjach Archiwum
Warunki początkowe:	Dostęp do bazy Repozytorium z metadanymi obiektów
Warunki końcowe:	Paczka archiwalna z obiektem zapisana na kasecie.

Projekt współfinansowany w ramach Programu Operacyjnego Polska Cyfrowa z Europejskiego Funduszu Rozwoju Regionalnego i budżetu państwa na podstawie Umowy o dofinansowanie nr POPC.02.01.00-00-0043/16-00



Fundusze  
Europejskie  
Polska Cyfrowa



Unia Europejska  
Europejski Fundusz  
Rozwoju Regionalnego

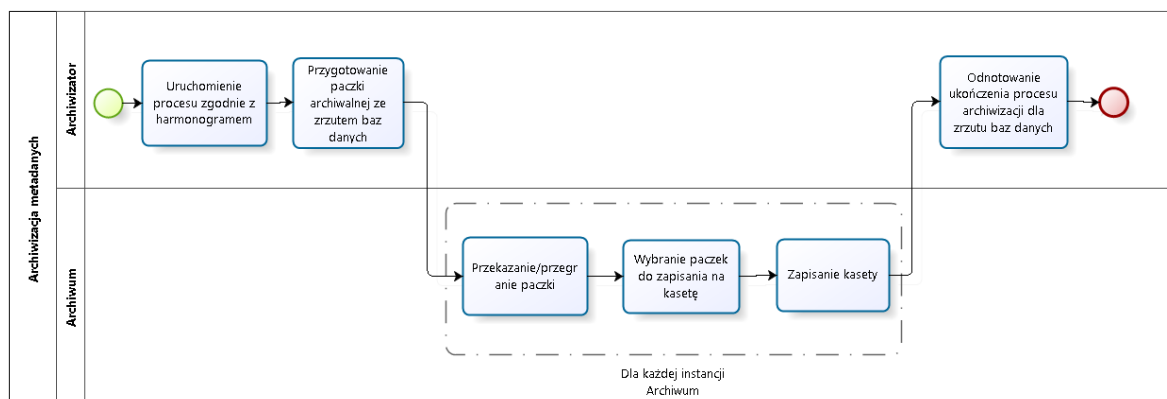


Strona 14 z 70

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Wykorzystywane zasoby	Macierz pomocnicza aby przygotować paczkę archiwalną. Biblioteka taśmowa aby zapisać paczkę archiwalną na kasety.
Uwagi:	Brak.

Poniżej zamieszczono diagram procesu.



Powered by  
**bizagi**  
Modeler


#### 4.4 Weryfikacja stanu kasy (P4)

Poniżej zamieszczono metrykę procesu.

Nazwa:	Weryfikacja stanu kasy
Kod:	P4
Cel:	Weryfikacja stanu kasy w celu oceny czy kasecie wymaga migracji
Warunki początkowe:	Brak.
Warunki końcowe:	Jeżeli kasecie wymaga migracji – przekazanie jej do zmigrowania. Jeżeli kasecie nie wymaga migracji – oznaczenie jako zweryfikowanej.
Wykorzystywane zasoby	Biblioteka taśmowa aby odczytać paczki archiwalne. Macierz pomocnicza aby przechować paczki archiwalne po odczytaniu z kasy.
Uwagi:	Brak.

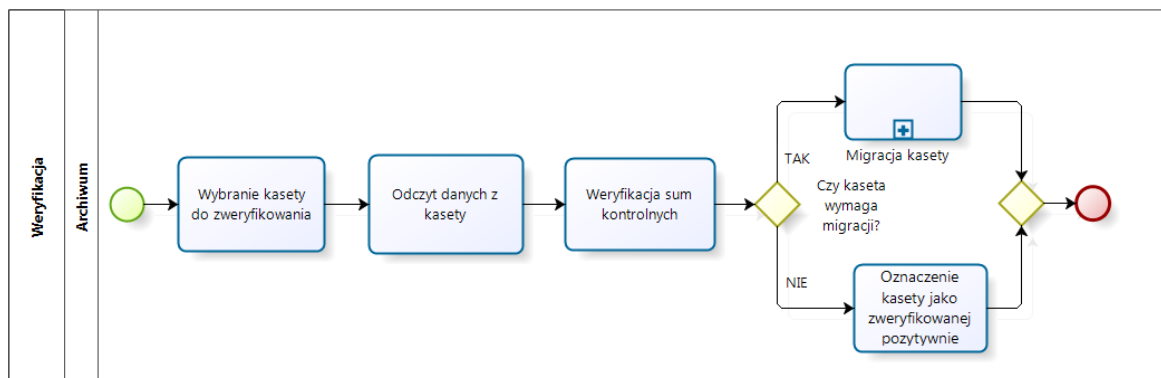
Można wyróżnić dwa rodzaje weryfikacji:

1. Weryfikacja pełna – polegająca na weryfikacji sum kontrolnych plików na kasecie.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

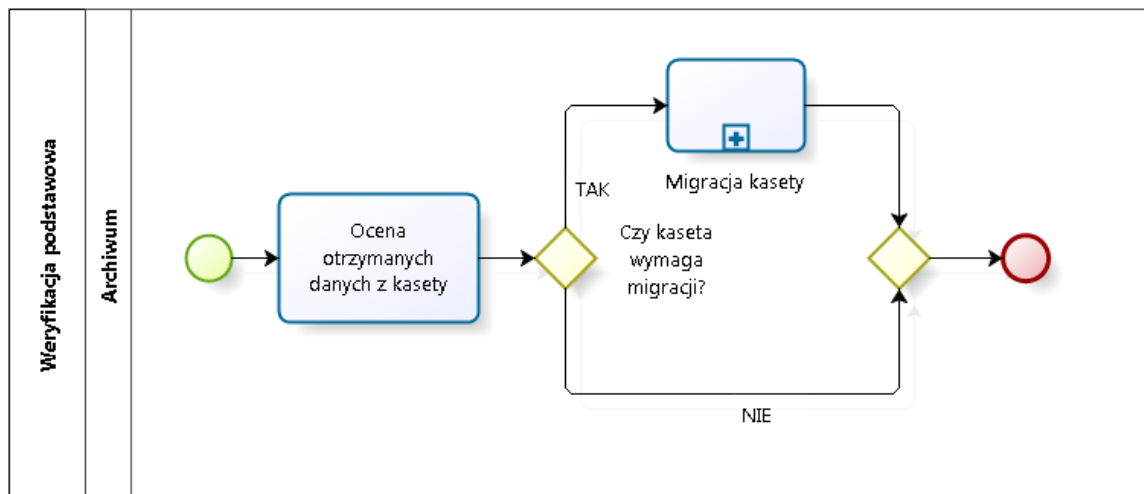
2. Weryfikacja podstawowa – polegająca na ocenie stanu kasety w oparciu o metadane otrzymane od napędu przy okazji odczytu kasety.

Poniżej zamieszczono diagram procesu pełnej weryfikacji.



Powered by  
**bizagi**  
Modeler

Poniżej zamieszczono diagram podstawowej weryfikacji.



Powered by  
**bizagi**  
Modeler

## 4.5 Migracja kasety (P5)

Poniżej zamieszczono metrykę procesu.

Nazwa:	Migracja kasety
--------	-----------------

Projekt współfinansowany w ramach Programu Operacyjnego Polska Cyfrowa z Europejskiego Funduszu Rozwoju Regionalnego i budżetu państwa na podstawie Umowy o dofinansowanie nr POPC.02.01.00-00-0043/16-00



Fundusze  
Europejskie  
Polska Cyfrowa


**omnis**<sup>BN</sup>

Unia Europejska  
Europejski Fundusz  
Rozwoju Regionalnego



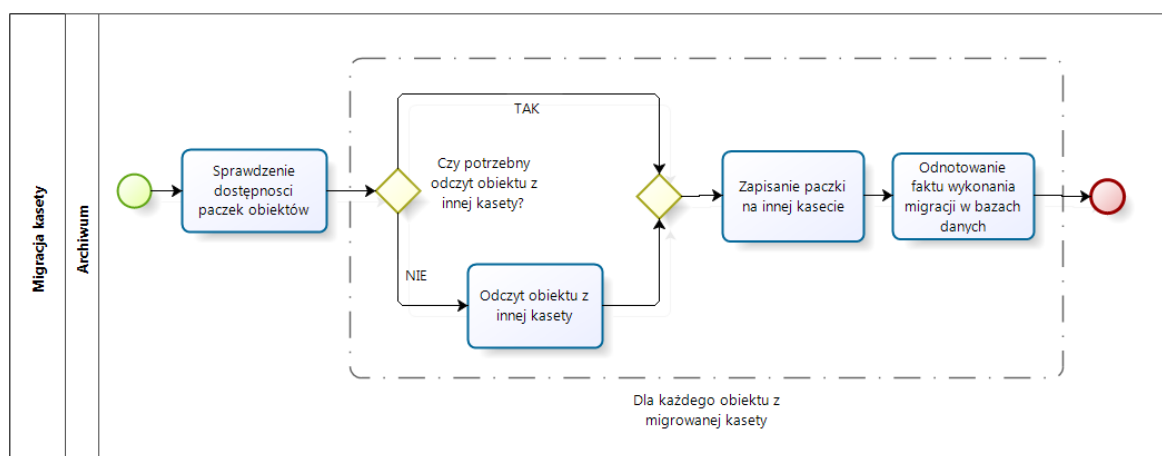
Strona 16 z 70



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Kod:	P5
Cel:	Migracja paczek znajdujących się na kasecie zakwalifikowanej do migracji na nowe kasety.
Warunki początkowe:	Brak.
Warunki końcowe:	Paczki archiwalne zmigrowane na nowe kasety. Migrowana kaseeta oznaczona jako „do usunięcia” z partycji biblioteki taśmowej.
Wykorzystywane zasoby	Biblioteka taśmowa aby odczytać migrowane paczki archiwalne. Macierz pomocnicza aby przechować paczki archiwalne po odczytaniu z kasety. Biblioteka taśmowa aby zapisać migrowane paczki archiwalne na kasetę.
Uwagi:	Brak.

Poniżej zamieszczono diagram procesu.



Powered by  
**bizagi**  
Modeler

## 4.6 Przywrócenie zawartości obiektu (P6)

Poniżej zamieszczono metrykę procesu.

Nazwa:	Przywrócenie zawartości obiektu
Kod:	P6
Cel:	Przywrócenie zawartości obiektu do Repozytorium cyfrowego
Warunki początkowe:	Brak.

Projekt współfinansowany w ramach Programu Operacyjnego Polska Cyfrowa z Europejskiego Funduszu Rozwoju Regionalnego i budżetu państwa na podstawie Umowy o dofinansowanie nr POPC.02.01.00-00-0043/16-00



Fundusze  
Europejskie  
Polska Cyfrowa

**omnis** BN

Unia Europejska  
Europejski Fundusz  
Rozwoju Regionalnego

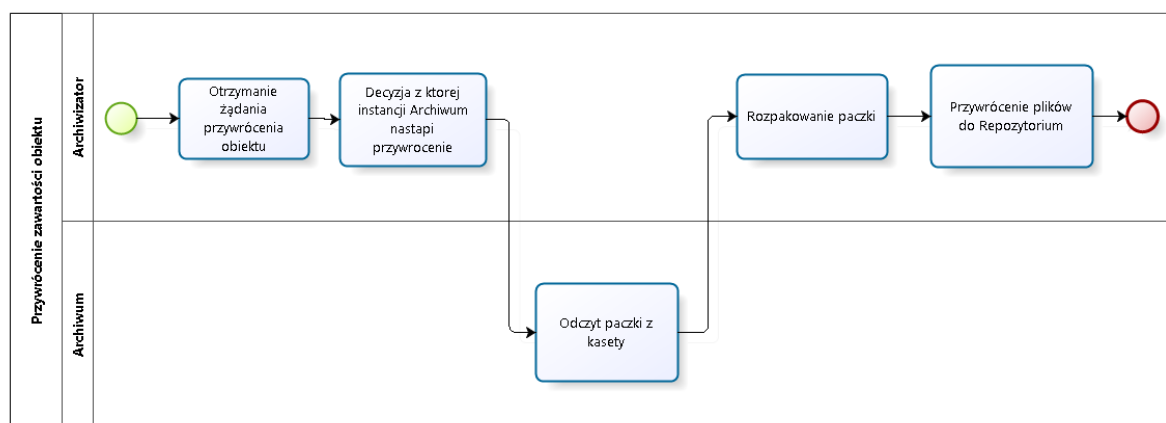


Strona 17 z 70

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Warunki końcowe:	Obiekt w Repozytorium z przywróconymi plikami.
Wykorzystywane zasoby	Biblioteka taśmowa aby odczytać migrowane paczki archiwalne. Macierz pomocnicza aby przechować paczki archiwalne po odczytaniu z kasety.
Uwagi:	Brak.

Poniżej zamieszczono diagram procesu.



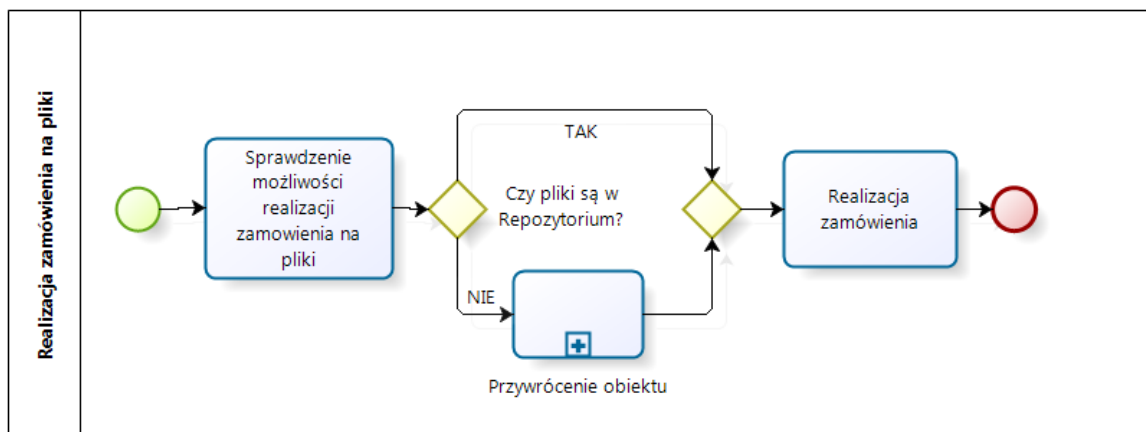
Powered by  
**bizagi**  
Modeler

## 4.7 Pozostałe procesy biznesowe (P7)

Poniżej została zamieszczona lista pozostałych procesów biznesowych, które będą analizowane w ramach tego projektu:

- Usunięcie plików ze zarchiwizowanego obiektu, który został przywrócony.  
Po przywróceniu plików do obiektu oraz po upływie czasu karencji powinno nastąpić usunięcie plików ze zarchiwizowanego obiektu
- Realizacja zamówień użytkowników na pliki z obiektów.  
Użytkownicy powinni móc składać zamówienia na pliki z obiektów. W przypadku kiedy zamówione pliki (lub ich część) nie są dostępne w Repozytorium, powinno następować przywrócenie obiektu, a następnie skompletowanie wszystkich zamówionych plików oraz udostępnienie zamówienia użytkownikowi. Poniżej przedstawiono diagram procesu realizacji zamówienia.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	



Powered by  
**bizagi**  
Modeler

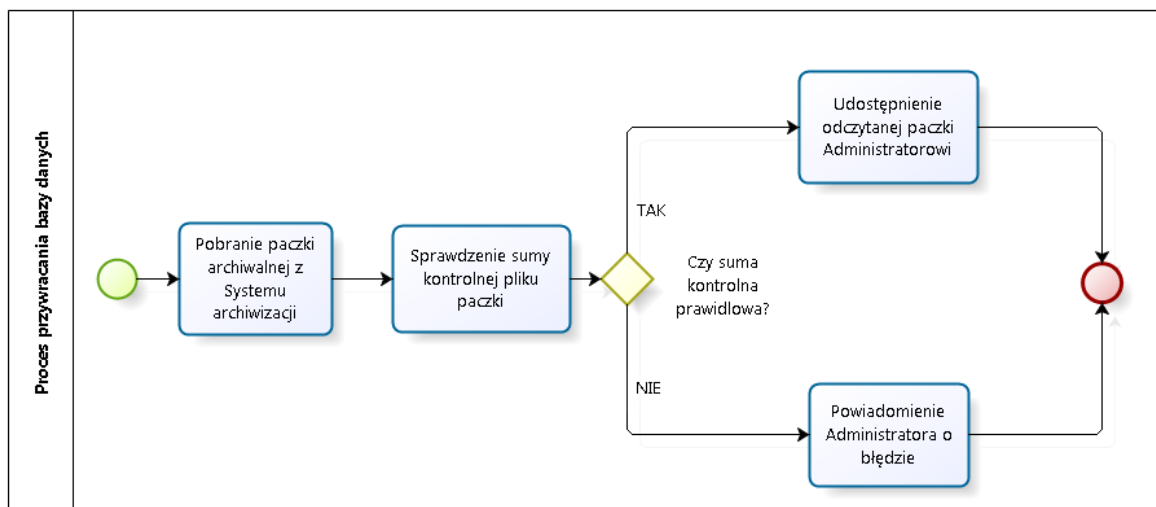
#### 4.8 Proces przywrócenia bazy danych (P8)

Proces archiwizacji powinien obejmować obiekty z Repozytorium oraz okresowo wykonywany zrzut bazy metadanych obiektów.

W przypadku archiwizacji obiektów w ramach projektu powinien zostać zaimplementowany proces przywracania obiektów do Repozytorium.

W przypadku archiwizacji bazy metadanych obiektów proces przywrócenia ich do Repozytorium nie jest przedmiotem niniejszego projektu, ale z uwagi na to, że pojawił się podczas etapu analizy procesowej, został on wstępnie zaprojektowany. Poniżej zamieszczony został koncepcyjny diagram tego procesu.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	



Powered by  
**bizagi**  
Modeler

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

## 5 Architektura

Treść rozdziału nie załączona ze względu na brak związku z przedmiotowym zamówieniem.

## 6 Wymagania funkcjonalne Systemu archiwizacji

### 6.1 Elementy wspólne

Treść rozdziału nie załączona ze względu na brak związku z przedmiotowym zamówieniem.

### 6.2 Archiwum

Treść rozdziału nie załączona ze względu na brak związku z przedmiotowym zamówieniem.

### 6.3 Usługi pomocnicze

W niniejszym rozdziale zostały opisane usługi pomocnicze Systemu archiwizacji, które logicznie mają związek z zadaniami realizowanymi przez Archiwizator.

Mowa o mechanizmach, które powinny uruchamiać: proces archiwizacji obiektu, proces archiwizacji zrzutu bazy metadanych/danych, proces przywracania przywrócenia obiektów oraz proces usuwania plików z obiektu.

Przygotowanie tych mechanizmów jako osobne usługi pomocnicze umożliwia w przyszłości uwzględnienie innego sposobu uruchamiania w/w procesów, przykładowo w sposób ręczny.

Ilekcioć w niniejszym rozdziale mowa o statusie obiektu, to jest to status obiektu w kontekście archiwizacji, którego opis znajduje się w rozdziale **Zmiany encji obiekt**.

#### 6.3.1 Monitorowanie obiektów do zarchiwizowania i inicjowanie archiwizacji

W ramach tego zadania Archiwizator powinien co określony skonfigurowany czas (parametr w konfiguracji: **archivization.scanner.cronExpression**) monitorować obiekty w Repozytorium w celu znalezienia obiektów spełniających kryteria archiwizacji.

Pełna lista kryteriów wyboru obiektów do archiwizacji została przedstawiona poniżej:

1. Obiekt posiada status NEW lub ARCHIVED.
1. Data ostatniej modyfikacji w kontekście archiwizacji jest późniejsza niż data ostatniej archiwizacji LUB data ostatniej archiwizacji jest pusta.
2. Od daty ostatniej modyfikacji w kontekście archiwizacji minął skonfigurowany okres karencji (parametr w konfiguracji: opisano w rozdziale **Zmiany encji obiekt**).
3. inWorkflow=false.

Powyższe warunki wybiorą zarówno obiekty do pierwszej archiwizacji jak i ponownej archiwizacji.

Projekt współfinansowany w ramach Programu Operacyjnego Polska Cyfrowa z Europejskiego Funduszu Rozwoju Regionalnego i budżetu państwa na podstawie Umowy o dofinansowanie nr POPC.02.01.00-00-0043/16-00



Fundusze  
Europejskie  
Polska Cyfrowa



Unia Europejska  
Europejski Fundusz  
Rozwoju Regionalnego



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

W ramach Repozytorium powinna zostać udostępniona metoda `getDocumentIdsToArchive`, która powinna zwracać kolekcję obiektów spełniających powyżej opisane kryteria.

Weryfikację okresu karencji od zakończenia przez obiekt przebiegu workflow planowano pierwotnie wykonać w oparciu o datę wykonania ostatniego kroku workflow wykorzystując encję „activity” w bazie danych. Jednakże z uwagi na dużą liczbę elementów w tej tabeli, wyszukiwanie danych mogłoby znacznie obciążać bazę danych Repozytorium. Dlatego weryfikacja okresu karencji od zakończenia workflow powinna zostać wykonana w taki sposób, aby każde wykonanie kroku w workflow aktualizowało datę ostatniej modyfikacji w kontekście archiwizacji. Wówczas wyszukiwanie obiektów do zarchiwizowania powinno zostać oparte o jeden wspólny okres karencji „od ostatniej modyfikacji w kontekście archiwizacji” oraz warunek `inWorkflow=false`. Dodanie oraz odpowiednia aktualizacja daty ostatniej modyfikacji w kontekście archiwizacji jest przedmiotem prac w ramach zmian w systemie Repozytorium.

Okres karencji od ostatniej modyfikacji w kontekście archiwizacji powinien być konfigurowalny (parametr w konfiguracji: opisano w rozdziale [Zmiany encji obiekt](#)).

Po znalezieniu obiektów spełniających kryteria archiwizacji, powinno nastąpić wywołanie przez interfejs Systemu archiwizacji żądania zarchiwizowania tego obiektu.

Podczas analizy kryteriów wyboru obiektów do archiwizacji pierwotnie rozważano rozdzielanie wykrywania obiektu do pierwszej archiwizacji od wykrywania obiektów do kolejnych archiwizacji. Analiza tego aspektu doprowadziła jednak do wniosku, że można zdefiniować jeden wspólny warunek wyszukiwania obiektów, dzięki czemu nie będzie potrzeby dwukrotnego monitorowania obiektów w poszukiwaniu kandydatów do archiwizacji, tym samym dwukrotnie obciążając bazę danych Repozytorium różnymi zapytaniami.

### 6.3.2 Uruchamianie procesu archiwizacji metadanych/bazy danych

W ramach tego zadania Archiwizator powinien co określony skonfigurowany czas (parametr w konfiguracji: `dbDump.cronExpression`) uruchamiać proces archiwizacji metadanych poprzez wywołanie przez interfejs Systemu archiwizacji tego procesu.

### 6.3.3 Monitorowanie obiektów do przywrócenia i inicjowanie przywrócenia

W ramach tego zadania Archiwizator powinien co określony skonfigurowany czas (parametr w konfiguracji: `restore.scanner.cronExpression`) monitorować obiekty w Repozytorium w celu znalezienia obiektów wymagających przywrócenia.

Obiekt powinien zostać zakwalifikowany do przywrócenia jeżeli spełniony jest chociaż jeden poniższy warunek:

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

1. Nastąpiła modyfikacja zredukowanego obiektu, co powoduje potrzebę przywrócenia oraz ponownej archiwizacji. Warunki zmodyfikowania zredukowanego obiektu są następujące:
  - a. Obiekt posiada status REDUCED
  - b. Obiekt jest zmodyfikowany: data ostatniej modyfikacji w kontekście archiwizacji jest późniejsza niż data ostatniej archiwizacji
2. Użytkownik złożył zamówienie na usunięcie pliki ze zredukowanego obiektu.
3. Proces Repozytorium wymaga usuniętego pliku ze zredukowanego obiektu.

W ramach Repozytorium powinna zostać udostępniona metoda `getDocumentIdsToRestore`, która powinna zwracać kolekcję obiektów spełniających powyżej opisane kryteria. Dodatkowo metoda ta powinna zwracać informację, czy przywrócenie obiektu posiada wysoki priorytet.

Po znalezieniu obiektów spełniających kryteria przywrócenia, powinno nastąpić wywołanie przez interfejs Systemu archiwizacji żądania przywrócenia tego obiektu

#### 6.3.4 Usuwanie plików z przywróconego lub niezredukowanego obiektu

W ramach tego zadania powinno co określony skonfigurowany czas (parametr w konfiguracji: `reduce.process.cronExpression`) być wykonywane monitorowanie obiektów w Repozytorium w celu znalezienia obiektów spełniających kryteria redukcji obiektu oraz powinna być wykonywana próba redukcji tych obiektów.

Pełna lista warunków wyszukiwania obiektów do zredukowania została przedstawiona poniżej:

1. Obiekt posiada status ARCHIVED.
2. Obiekt jest nie jest zablokowany przez Repozytorium (UNLOCKED) oraz nie jest zmodyfikowany (data ostatniej archiwizacji jest późniejsza niż data ostatniej modyfikacji w kontekście archiwizacji).
3. Od daty ostatniego przywrócenia obiektu minął okres karencji lub data ostatniego przywrócenia jest pusta (parametr w konfiguracji: opisano w rozdziale [Zmiany encji obiekt](#)).


W ramach Repozytorium powinna zostać udostępniona metoda `getDocumentIdsToReduce`, która powinna zwracać kolekcję obiektów spełniających powyżej opisane kryteria.

Po znalezieniu obiektów spełniających kryteria redukcji dla każdego znalezionego obiektu powinny zostać wykonane następujące kroki (dodatkowo poniżej zamieszczony został diagram tego procesu):

1. Zmiana statusu obiektu

W ramach tego kroku powinna nastąpić zmiana statusu obiektu na REDUCING.



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

## 2. Usunięcie plików z obiektu w Repozytorium

Przeanalizowane powinny zostać tylko te kontenty, które zawierają streamy spełniające kryteria (klucze stream'ów, mime-types plików) zadane w konfiguracji streamów archiwizowanych i usuwanych (parametr w konfiguracji: opisano w rozdziale **Konfiguracja archiwizowanych i usuwanych kluczy stream'ów z obiektu**).

Plik powinien zostać usunięty ze stream'a jeżeli spełnia następujące kryteria:

- a. Kontent, do którego należy stream nie jest współdzielony pomiędzy różne obiekty; LUB jeżeli jest współdzielony, to wszystkie obiekty do których należy dany kontent są zredukowane:
  - i. Jeżeli wszystkie pozostałe obiekty są zredukowane, to oznacza to, że analizowany obiekt jest ostatnim niezredukowanym obiektem, do którego należy analizowany kontent. W tej sytuacji streamy należące do analizowanego współdzielonego kontentu mogą i powinny zostać usunięte.
  - ii. Jeżeli chociaż jeden obiekt (z pozostałych obiektów) nie jest zredukowany, to streamy należące do analizowanego współdzielonego kontentu nie powinny zostać usunięte.
- b. Stream posiada status ARCHIVED.
- c. Plik posiada mime-type, które znajduje się na skonfigurowanej liście usuwanych mime-types plików – liście skonfigurowanej dla stream'ów o danym kluczu.

W przypadku wykonania usunięcia stream'a, powinien on otrzymać status DELETED.

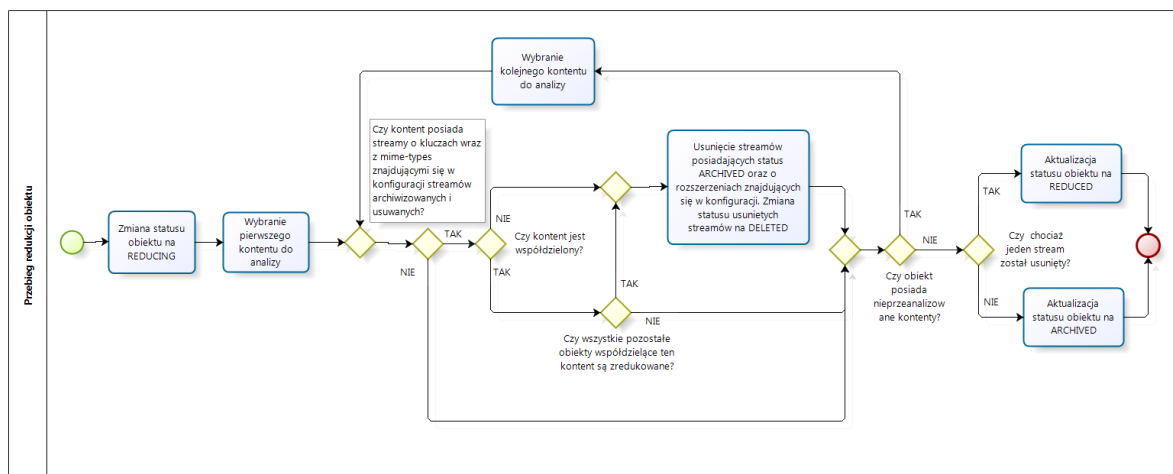
## 3. Aktualizacja statusu obiektu w Repozytorium

W ramach tego punktu powinno nastąpić zaktualizowanie statusu obiektu:

- a. Jeżeli chociaż jeden stream w obiekcie jest usunięty – to zmiana statusu obiektu z REDUCING na REDUCED.
- b. Jeżeli ani jeden stream w obiekcie nie jest usunięty – to zmiana statusu obiektu z REDUCING na ARCHIVED.

Poniżej zamieszczony został diagram tego procesu:





Powered by  
bizagi  
Modeler

### Konfiguracja usuwanych kluczy stream'ów z obiektu

Konfiguracja usuwanych kluczy stream'ów z obiektu została opisana w ramach rozdziału **Konfiguracja archiwizowanych i usuwanych kluczy stream'ów z obiektu**.

## 6.4 Archiwizator

Główne realizowane zadania biznesowe – właściwy proces archiwizacji:

1. Przyjęcie żądania zarchiwizowania obiektu
2. Przyjęcie żądania zarchiwizowania metadanych/bazy danych
3. Przygotowywanie paczki archiwalnej z obiektem
4. Przygotowywanie paczki archiwalnej ze zrzutem metadanych/bazy danych
5. Przekazanie żądania archiwizacji paczki archiwalnej do instancji Archiwum
6. Odebranie informacji o statusie zakończenia zadania archiwizacji paczki w instancji Archiwum wraz z podjęciem kolejnych kroków
7. Obsługa zakończenia procesu archiwizacji paczki z obiektem
8. Obsługa zakończenia procesu archiwizacji paczki z metadanymi

Główne realizowane zadania biznesowe – proces przywracania:

1. Przyjęcie żądania przywrócenia obiektu / udostępnienia paczki archiwalnej
2. Przyjęcie od instancji Archiwum żądania odczytu paczki archiwalnej
3. Odebranie informacji o statusie zadania odczytu paczki archiwalnej wraz z podjęciem kolejnych kroków
4. Monitorowanie statusu przywrócenia obiektu

Wymagania biznesowe do modułu Archiwizator:

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

1. Moduł powinien korzystać z dedykowanych usług realizujących elementarne dające się wydzielić czynności, przykładowo z usługi do pakowania/rozpakowania plików.
2. System powinien dbać o relację listy typów plików archiwizowanych i usuwanych z obiektu.
3. Czas oczekiwania na udostępnienie paczki autorowi żądania powinien być konfigurowalny. Po tym czasie podsystem powinien uznać, że nastąpiła sytuacja nietypowa i powinno to spowodować wysłanie powiadomienia do administratora. Nie powinno to spowodować przerwania realizacji żądania przywracania.

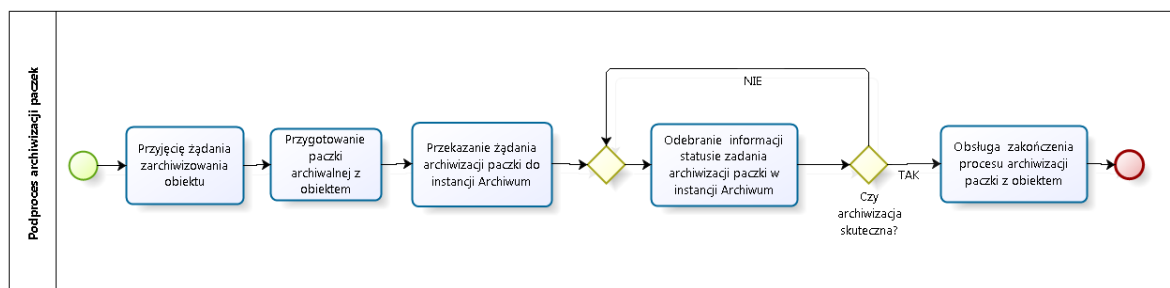
Ilekoć w niniejszym rozdziale mowa o statusie obiektu, to jest to status obiektu w kontekście archiwizacji, którego opis znajduje się w rozdziale **Zmiany encji obiekt**.

#### 6.4.1 Przyjęcie żądania zarchiwizowania obiektu

Po przyjęciu żądania archiwizacji obiektu oprócz uruchomienia procesu archiwizacji powinno nastąpić ustawienie obiektowi statusu ARCHIVING.

W przypadku, w którym system odbierze żądanie archiwizacji obiektu, który jest już archiwizowany – powinien zostać zwrócony odpowiedni komunikat lub kod odrzucenia żądania.

Poniżej zamieszczony został poglądowy diagram procesu archiwizacji obiektu. Diagram ten ma za zadanie przedstawić kolejne kroki procesu bez uwzględniania aspektu asynchroniczności poszczególnych kroków.



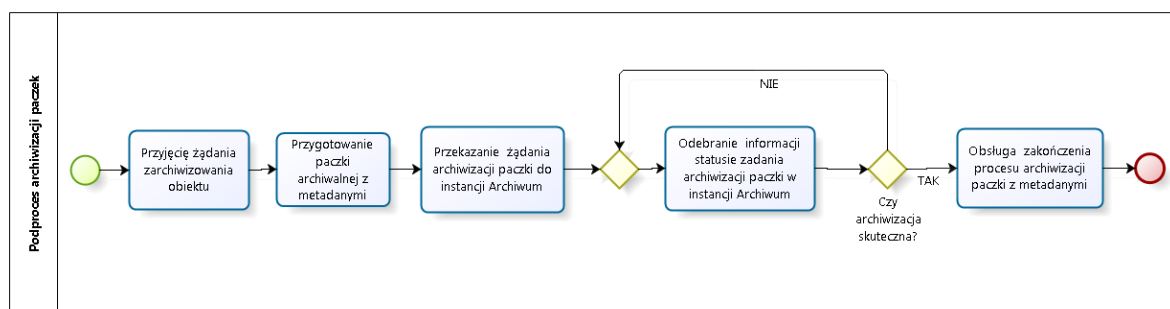
Powered by  
**bizagi**  
Modeler

#### 6.4.2 Przyjęcie żądania zarchiwizowania metadanych/bazy danych

W ramach tego punktu powinno nastąpić rozpoczęcie procesu archiwizacji metadanych.

Poniżej zamieszczony został poglądowy diagram procesu archiwizacji metadanych. Diagram ten ma za zadanie przedstawić kolejne kroki procesu bez uwzględniania aspektu asynchroniczności poszczególnych kroków.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	



Powered by  
bizagi  
Modeler

### 6.4.3 Przygotowywanie paczki archiwalnej z obiektem

W ramach tego zadania powinna zostać przygotowana paczka archiwalna z danymi archiwizowanego obiektu, która powinna zawierać następujące elementy:

1. Metadane obiektu
2. Pozostałe informacje o obiekcie (np. dane z procesu workflow, segmentacja)
3. Strukturę obiektu – kontenty i ich struktura
4. Pliki stream'ów zgodnie z konfiguracją.

Aby wyeliminować sytuacje, kiedy proces przygotowania paczki archiwalnej zostanie rozpoczęty, ale w trakcie jego trwania miejsce na macierzy pomocniczej się wyczerpie, powinien istnieć prosty mechanizm rezerwacji miejsca dla rozpoczynanych zadań mogących wpłynąć na ilość wolnego miejsca na macierzy pomocniczej. Powinien także istnieć mechanizm umożliwiający ustawienie limitu miejsca zajętego (ang. quota) przez proces archiwizacji oraz proces przywracania/udostępniania – w systemie powinien być parametr definiujący, ile procent miejsca na macierzy pomocniczej przeznaczone jest dla procesu archiwizacji (parametr w konfiguracji: **archivization.space.share**); procent miejsca na macierzy pomocniczej przeznaczony dla procesu przywracania/udostępniania powinien wynikać z różnicy 100% - x%, gdzie x to parametr dla procesu archiwizacji. Ten mechanizm pozwoli uniknąć sytuacji zakleszczenia się obu procesów z uwagi na zużycie całego dostępnego miejsca na macierzach pomocniczych.

Pod pojęciem momentu stworzenia paczki archiwalnej (szczególnie w kontekście znacznika czasowego w paczce archiwalnej) rozumie się moment rozpoczęcia pobierania danych z obiektu z Repozytorium do Systemu archiwizacji.

#### 6.4.3.1 Zawartość paczki archiwalnej

Szczegółowa lista elementów obiektu do umieszczenia w paczce archiwalnej została przedstawiona poniżej:

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

1. Wersja Systemu archiwum – jako że w budowaniu baczki archiwalnej bierze udział tylko Archiwizator, to wystarczające jest uwzględnienie w tym miejscu wersji Archiwizatora.
2. Metadane obiektu (w dwóch wersjach, natywnym oraz przekonwertowane do DC – konwersja na DC obecnie dostępnymi gotowymi narzędziami w RCBN, natomiast w celu zapewnienia walidacji zawartości pliku METS, powinien powstać plik XSD, który umożliwi walidację w szczególności metadanych obiektu zapisanych w formacie natywnym Repozytorium jak i w formacie DC)
3. Metadane kontentów (w dwóch wersjach, natywnym oraz przekonwertowane do DC – konwersja na DC obecnie dostępnymi gotowymi narzędziami w RCBN, natomiast w celu zapewnienia walidacji zawartości pliku METS, powinien powstać plik XSD, który umożliwi walidację w szczególności metadanych obiektu zapisanych w formacie natywnym Repozytorium jak i w formacie DC)
4. Pozostałe techniczne atrybuty obiektu: data utworzenia obiektu, data ostatniej modyfikacji obiektu
5. Pozostałe biznesowe atrybuty obiektu: liczba dostępnych w BN fizycznych kopii, główny kontent, kontekst obiektu, informacja czy obiekt jest publiczny czy pod prawem autorskim
6. Historia workflow (ekran log obiektu w GUI):
  - a. Nazwa kroku
  - b. Data wykonania zadania
  - c. Wykonujący (Obiekt przypisany do)
7. Historia skanowania: skanery na których były wykonane skany kontentów (graficznych)
8. Historia modyfikacji (ekran historia modyfikacji w GUI)
9. Lista (struktura) kontentów postaci zbiorczych + ich metadane jak w pkt 3 + lista dostępnych stream'ów + dodatkowe atrybuty (czy ukryty, stan obrotu dla plików graficznych)
10. Lista (struktura) kontentów w strukturze paginacyjnej (wraz z informacją o zachowanej kolejności oraz informacją o kontentach alternatywnych) + ich metadane + lista dostępnych stream'ów + szczegóły paginacyjne (obrot, rozdział, układ, renumeracja, strony, ciekawe, funkcja strony, opis strony) + dodatkowe atrybuty (informacja czy jest ukryty, stan obrotu dla plików graficznych)
11. Segmentacja w obiekcie
12. Pliki (streamy) + ich sumy kontrolne
13. Kategorie, kolekcje – nazwy kategorii i kolekcji przypisanych do obiektu
14. Odnośniki do innych obiektów – identyfikatory podobieństw

Aby było możliwe odtworzenie całego obiektu w bazie danych Repozytorium, jak również przywrócenie zawartości (plików) istniejącego obiektu w Repozytorium, potrzebna jest możliwość precyzyjnego zidentyfikowania, w które miejsce w obiekcie dany plik powinien zostać przywrócony. Dlatego w paczce archiwalnej każdy plik powinien być opisany identyfikatorem kontentu, do którego należy, oraz nazwą stream'u.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

#### 6.4.3.2 Konfiguracja archiwizowanych i usuwanych kluczy stream'ów z obiektu

Powinna być możliwość skonfigurowania typów (kluczy) stream'ów, podlegających archiwizowaniu, oraz w ramach tych kluczy stream'ów tzw. mime-types plików podlegających usuwaniu.

Z uwagi na to, że ta konfiguracja będzie wykorzystywana zarówno przez System archiwizacji, jak i przez Repozytorium, to konfiguracja ta powinna spełniać następujące wymagania:

1. Konfiguracja listy archiwizowanych kluczy stream'ów oraz usuwanych mime-types plików powinna być przechowywana w bazie danych Repozytorium (tabele: [archivable\\_stream](#), [reducible\\_stream\\_type](#)).
2. Powinna zostać przygotowana metoda do pobierania wartości tych konfiguracji z Repozytorium – z której to będzie korzystał System archiwizacji.
3. W Systemie archiwizacji ta konfiguracja będzie cache'owana przez czas zdefiniowany w parametrach Systemu archiwizacji (parametr w konfiguracji: [archivization.config.refresh.delay](#)).

Konfiguracja powinna umożliwiać podanie klucza stream'a (np. `mainStream`, `download_alto`, `text`, `download_fullJPG`) co było by równoznaczne z uwzględnieniem go w archiwizacji, oraz zdefiniowanie dla tego klucza, które tzw. mime-types plików powinny być usuwane. Przykład takiej konfiguracji został zaprezentowany poniżej:

1. Archiwizowany klucz=`"mainStream"`, usuwane mime-types:
  - a. `"image/tiff"`
  - b. `"image/jpeg"`
  - c. `"image/png"`
2. Archiwizowany klucz=`"download_fullJPG"`, usuwane mime-types:
  - a. `"image/jpeg"`
3. Archiwizowany klucz=`"download_przyklad"`, usuwane mime-types: (brak)

Punkt 3 w powyższym przykładzie należy rozumieć w taki sposób, że stream o kluczu `"download_przyklad"` powinien być archiwizowany, ale nie powinno nastąpić usunięcie jakiegokolwiek pliku stream'a.

W związku z konfiguracją wykonywaną w opisany powyżej sposób wykluczającą na poziomie pliku konfiguracyjnego pomyłki polegającej na zdefiniowaniu usunięcia stream'u, który jednocześnie nie jest archiwizowany, nie ma konieczności wprowadzenia dodatkowego mechanizmu weryfikacji w plikach konfiguracyjnych relacji listy kluczy stream'ów archiwizowanych i usuwanych z obiektu.

Mechanizm usuwania stream'ów (plików) został szczegółowo opisany rozdziale dotyczącym usuwania plików z przywróconego/niezredukowanego obiektu ([Usuwanie plików z przywróconego lub niezredukowanego obiektu](#)). Zgodnie z opisem we wspomnianym

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	


rozdziale, podczas usuwania stream'ów (plików) powinna zostać zwrócona uwaga na to, aby nie usunąć stream'ów (plików), które nie są oznaczone jako zarchiwizowane.

#### 6.4.3.3 Struktura danych w kontenerze METS

Dane i informacje o strukturze obiektu powinny zostać umieszczone w kontenerze METS (Metadata Encoding & Transmission Standard), którego struktura została opisane w poniższej tabeli:

Sekcja	Skrót	Elementy	Przykład
METS (korzeń)	mets:mets	OBJID=<Id obiektu (w Repozytorium)> LABEL=<Nazwa obiektu (Tytuł i Iteracja)> ARCHIVISER_VERSION= <Wersja Archiwizatora> OWNER="Biblioteka Narodowa"	<mets:mets xsi:schemaLocation=" ... " OBJID="2648414" LABEL="Bromberger Zeitung R:1894 Nr 057" ARCHIVISER_VERSION="8.234" OWNER="Biblioteka Narodowa">
METS header	metsHdr	CREATEDATE=<Data utworzenia pliku METS> ARCHIVINGDATE=<Data rozpoczęcia zadania kopiowania plików z Repozytorium do Systemu archiwizacji>	<mets:metsHdr CREATEDATE="2012-09-27T08:38:19Z" ARCHIVINGDATE="2012-09-27T08:41:47Z"> </mets:metsHdr>
Descriptive Metadata	dmdSec	1. Metadane obiektu: a) natywnie jak w Repozytorium b) w standardzie DC 2. Metadane kontentów jak w Repozytorium a) postaci zbiorcze - poza podstawowymi metadanymi kontentów: - czy kontent ukryty - stan obrotu dla plików graficznych b) kontenty w strukturze paginacyjnej - poza podstawowymi metadanymi kontentów: - szczegóły paginacyjne (obróć, rozdział, układ, renumeracja, strony, ciekawe, funkcja strony, opis strony) - dodatkowe atrybuty (czy ukryty, stan obrotu dla plików graficznych)	<mets:dmdSec ID="REPO_OBJECT"> ... </mets:dmdSec> <mets:dmdSec ID="DC_OBJECT"> ... </mets:dmdSec> <mets:dmdSec ID="REPO_CONTENT_<Id kontentu 1>"> ... </mets:dmdSec> <mets:dmdSec ID="REPO_CONTENT_<Id kontentu 2>"> ... </mets:dmdSec> <mets:dmdSec ID="REPO_CONTENT_<Id kontentu n>"> ... </mets:dmdSec>



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Sekcja	Skrót	Elementy	Przykład
Administrative Metadata	amdSec	<p>1. Jako <b>techMD</b>:</p> <p>a) Pozostałe techniczne atrybuty obiektu: data utworzenia, data ostatniej modyfikacji</p> <p>b) Pozostałe biznesowe atrybuty obiektu: główny kontent, kontekst obiektu</p> <p>c) Segmentacja w obiekcie wraz z metadanymi artykułów</p> <p>d) Kategorie, kolekcje – nazwy kategorii i kolekcji przypisanych do obiektu</p> <p>e) Odnosińniki do innych obiektów – identyfikatory podobieństw</p> <p>2. Jako <b>rightsMD</b>:</p> <p>a) Liczba dostępnych w BN fizycznych kopii</p> <p>b) czy obiekt jest objęty prawami autorskimi</p> <p>3. Jako <b>sourceMD</b>:</p> <p><i>Sekcja nie zostanie wykorzystana</i></p> <p>4. Jako <b>digiprovMD</b>:</p> <p>a) Historia workflow – log obiektu:</p> <ul style="list-style-type: none"> <li>- Nazwa kroku</li> <li>- Data wykonania zadania</li> <li>- Wykonujący (Obiekt przypisany do)</li> </ul> <p>b) Historia skanowania: skanery na których były wykonane skany kontentów (graficznych)</p> <p>c) Historia modyfikacji (ekran historia modyfikacjiw GUI)</p>	
File Section	fileSec	<p>1. Lista strumieni (plików) wraz z sumami kontrolnymi</p> <p>2. Użycie następujących grup plików:</p> <ul style="list-style-type: none"> <li>- streamy źródłowe (z kluczem mainStream)</li> <li>- streamy pochodne (klucz inny niż mainStream)</li> <li>- pozostałe (serializowane dane obiektu z HCP)</li> </ul> <p>Konwencja nazywania identyfikatorów plików: streamID_streamName</p>	<pre>&lt;fileSec&gt; &lt;fileGrp ID="MAINSTREAMS"&gt; ... &lt;/fileGrp&gt; &lt;fileGrp ID="DERIVEDSTREAMS"&gt; ... &lt;/fileGrp&gt; &lt;fileGrp ID="OTHER"&gt; ... &lt;/fileGrp&gt; &lt;/fileSec&gt;</pre>

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Sekcja	Skrót	Elementy	Przykład
Structural Map	structMap	<p>1. Kontenty w strukturze paginacyjnej: struktura kontentów + lista streamów</p> <p>- Elementy &lt;div&gt; na pierwszym poziomie określają kontenty główne. - Jeśli dana strona ma kontenty alternatywne używamy zagłębionego &lt;div&gt; numerując je kolejno przy pomocy Order zaczynając od 1.</p> <p>- Metadane kontentów linkowane są za pomocą atrybutu DMDID (oczywiście właściwe sekcje metadanych muszą być stworzone w ramach tego samego dokumentu w &lt;dmdSec&gt;).</p> <p>2. Postaci zbiorcze: struktura kontentów + lista streamów</p>	<p>Przykład contentu w strukturze paginacyjnej:</p> <pre>&lt;structMap LABEL="Struktura paginacyjna" TYPE="logical"&gt; &lt;!--content główny dla strony 243--&gt; &lt;mets:div ORDER="243" LABEL="Strona nr 243" TYPE="page" ID="DIV00243" DMDID="CID45654456meta"&gt; &lt;!--pliki contentu głównego--&gt; &lt;mets:fptr FILEID="CID45654456_mainStream"/&gt; &lt;mets:fptr FILEID="CID45654456_downloadALTO"/&gt; &lt;!--content alternatywny strony 243 nr 1--&gt; &lt;mets:div ORDER="1" ID="DIV243_1" TYPE="alternative page" DMDID="CID45654457meta"&gt; &lt;mets:fptr FILEID="CID45654457_mainStream"/&gt; &lt;mets:fptr FILEID="CID45654457_downloadALTO"/&gt; &lt;/mets:div&gt; &lt;!--content alternatywny strony 243 nr 2--&gt; &lt;mets:div ORDER="2" ID="DIV0243_2" TYPE="alternative page" DMDID="CID45654458meta"&gt; &lt;mets:fptr FILEID="CID45654458_mainStream"/&gt; &lt;mets:fptr FILEID="CID45654458_downloadALTO"/&gt; &lt;/mets:div&gt; &lt;/mets:div&gt; &lt;/structMap&gt;</pre> <p>Przykład contentu zbiorczego:</p> <pre>&lt;structMap LABEL="Pliki pozapaginacyjne " TYPE="logical"&gt; &lt;mets:div ORDER="1" LABEL="bulk1" TYPE="document" ID="bulk1" DMDID="CID45654490meta"&gt; &lt;mets:fptr FILEID="CID45654490_mainStream"/&gt; &lt;mets:fptr FILEID="CID45654490_text"/&gt; &lt;mets:fptr FILEID="CID45654490_download_ALTO"/&gt; &lt;/mets:div&gt; &lt;/structMap&gt;</pre> <p>Stringi do wpisywania w wartość TYPE powinny być konfigurowalne.  „nazwa typu dla structMap kontentów w strukturze paginacyjnej”= „logical”  „nazwa typu węzła strony głównej” = „page”  „nazwa typu dla structMap kontentów postaci zbiorczej” = „alternative page”  „nazwa typu dla węzła kontentów postaci zbiorczej” = „logical”  „nazwa typu dla kontentu postaci zbiorczej” = „document”</p>
Structural Links	structLink	<i>Sekcja nie zostanie wykorzystana</i>	
Behavioral	behaviorSec	<i>Sekcja nie zostanie wykorzystana</i>	

Projekt współfinansowany w ramach Programu Operacyjnego Polska Cyfrowa z Europejskiego Funduszu Rozwoju Regionalnego i budżetu państwa na podstawie Umowy o dofinansowanie nr POPC.02.01.00-00-0043/16-00



Fundusze  
Europejskie  
Polska Cyfrowa


**omnis**<sup>BN</sup>

Unia Europejska  
Europejski Fundusz  
Rozwoju Regionalnego



Strona 32 z 70



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Daty w paczce METS powinny być zapisane w standardzie ISO8601.

#### 6.4.3.4 Struktura folderów i nazewnictwo plików w paczce archiwalnej

Wszystkie pliki i podfoldery w paczce archiwalnej powinny znajdować się w folderze, którego nazwą powinien być identyfikator obiektu w Repozytorium. Następnie w tym folderze powinien znaleźć się plik z kontenerem METS oraz podfoldery, których nazwą powinny być poszczególne klucze stream'ów (np. mainStream, download\_alto, text, download\_fullJPG, download\_lowJPG). W tych podfolderach powinny zostać umieszczone stream'y (pliki) archiwizowanych treści należących do obiektu.

Nazewnictwo plików stream'ów zostało przedstawione poniżej:

<identyfikator obiektu z Repozytorium>separator<nazwa typu streamu>separator<kolejność w obiekcie>separator<Identyfikator kontentu z repozytorium>separator<identyfikator streamu>.rozszerzenie pliku

Pod pojęciem kolejności w obiekcie należy rozumieć systemową (w rozumieniu wewnętrznych identyfikatorów kolejności) w Repozytorium kolejność treści w strukturze paginacyjnej.

Numer będący kolejnością w obiekcie powinien być uzupełniany zerami do 4 cyfr, np. 0001, 0002, ..., 0010, 0011... .

W przypadku treści postaci zbiorczych kolejność należy rozumieć jako systemową (w rozumieniu wewnętrznych identyfikatorów kolejności) kolejność w Repozytorium – o ile taka jest dostępna w Repozytorium. Jeżeli taka kolejność w Repozytorium nie jest dostępna, to w nazwach stream'ów postaci zbiorczych zamiast kolejności pobranej z Repozytorium powinna zostać użyta na stałe liczba zero („0”).

Separator zostanie wybrany przez Wykonawcę na etapie implementacji. Może to być przykładowo znak podkreślenia dolnego „\_” lub myślnik „-”.

#### 6.4.3.5 Postać paczki archiwalnej

Paczka archiwalna zawierająca obiekt powinna zostać przygotowana jako pojedynczy plik TAR stworzony bez zastosowania kompresji. Brak zastosowania kompresji argumentuje się tym, że największą część paczki archiwalnej pod kątem zajmowanego rozmiaru będą stanowiły pliki w formacie tiff, który to format już jest skompresowany. Dodatkowo napęd w ramach biblioteki taśmowej zapewnia funkcjonalność kompresji sprzętowej, dlatego wykonanie na etapie tworzenia paczki archiwalnej dodatkowej próby kompresji nie jest zasadne (a może w odległej przyszłości spowodować trudności z odpakowaniem paczki z uwagi na trudności w dostępie do narzędzi służących do odwrócenia procesu kompresji).

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Jedynym istotnym argumentem za podziałem paczki archiwalnej na wiele plików były by problemy sieciowe, których skutkiem w przypadku przerwania transmisji była by potrzeba retransmisji całego dużego pliku paczki archiwalnej. Jednak użycie narzędzia **rsync** zapewnia obsługę tego typu problemów. Natomiast podział paczki archiwalnej na wiele plików powodowałby potrzebę dużo bardziej skomplikowanych algorytmów doboru paczek do zapisania na kasie, jak również dużo bardziej skomplikowanego mechanizmu przekazującego paczkę archiwalną pomiędzy węzłami Systemu archiwizacji.

#### 6.4.3.6 Nazwa pliku paczki archiwalnej

Nazwa paczki archiwalnej powinna składać się z następujących elementów:

1. Identyfikator obiektu z Repozytorium
2. Znacznik czasowy z momentu stworzenia paczki archiwalnej
3. Identyfikator klienta Systemu archiwizacji, który dla Repozytorium będzie cyfrą „1”
4. Stały tekst: „Owner\_BibliotekaNarodowa”

Znacznik czasowy powinien być w formacie „Unix time”, który oznacza liczbę sekund, które upłynęły od 1 stycznia 1970 wykorzystując strefę czasową UTC.

Przykład nazwy pliku został zaprezentowany poniżej:

Id\_201671#Time\_1475139982#Source\_1# Owner\_BibliotekaNarodowa.TAR

#### 6.4.4 Przygotowywanie paczki archiwalnej ze zrzutem metadanych/bazy danych

W ramach tego zadania powinna zostać przygotowana paczka archiwalna metadanych, która powinna zawierać elementy opisane w niniejszym rozdziale.

Aby wyeliminować sytuacje, kiedy proces przygotowania paczki archiwalnej zostanie rozpoczęty, ale w trakcie jego trwania miejsce na macierzy pomocniczej się wyczerpie, powinien istnieć prosty mechanizm rezerwacji miejsca dla rozpoczynanych zadań mogących wpłynąć na ilość wolnego miejsca na macierzy pomocniczej.

##### 6.4.4.1 Zawartość paczki archiwalnej

Szczegółowa lista elementów do umieszczenia w paczce archiwalnej została przedstawiona poniżej:

1. Numer wersji Systemu archiwizacji (zawierający numer wersji poszczególnych instancji Archiwum oraz wersję Archiwizatora), np. w pliku XML lub tekstowym
2. Plik z bazą Repozytorium
3. Pliki z bazami Systemu archiwizacji: osobne pliki dla poszczególnych baz węzłów Systemu archiwizacji

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Należy pamiętać, że pliki ze zrzutami Systemu archiwizacji będą zawierały zadania/żądania do wykonania, których stan będzie aktualny na moment przygotowania paczki archiwalnej. Dlatego ewentualne przywrócenie tych baz do Systemu archiwizacji w szczególności spowoduje utratę spójności danych w Systemie archiwizacji.

#### 6.4.4.2 Format zrzutu baz danych

Formatem zrzutu baz danych powinien być tzw. DUMP (zrzut) zawierający strukturę i zawartość bazy danych, który powinien być w momencie wykonywania zrzutu od razu kompresowany z użyciem GZIP (polecenie `mysqldump .... | gzip > dump.sql.gz`). Takie podejście umożliwia zaoszczędzenie miejsca na macierzy pomocniczej już na tym etapie, a nie dopiero po skompresowaniu całej paczki archiwalnej. Zastosowanie kompresji w stosunku do zrzutów baz danych jest zasadne, ponieważ zrzut typu DUMP posiada postać tekstową, a pliki tekstowe zazwyczaj poddają się przynajmniej częściowej kompresji.

#### 6.4.4.3 Struktura folderów i nazewnictwo plików w paczce archiwalnej

Wszystkie pliki powinny znajdować się w jednym katalogu, którego nazwą będzie data utworzenia katalogu w formacie YYYYMMDD, np. 20160930.

Nazewnictwo plików ze zrzutami baz danych powinno być następujące:

MySQLDump<separator><identyfikator węzła>.SQL

Separator zostanie wybrany przez Wykonawcę na etapie implementacji. Może to być przykładowo znak podkreślenia dolnego „\_”.

#### 6.4.4.4 Postać paczki archiwalnej

Paczka archiwalna zawierająca zrzut metadanych powinna zostać przygotowana jako pojedynczy plik TAR.

Jedynym istotnym argumentem za podziałem paczki archiwalnej na wiele plików były by problemy sieciowe, których skutkiem w przypadku przerwania transmisji była by potrzeba retransmisji całego dużego pliku paczki archiwalnej. Jednak użycie narzędzia `rsync` zapewnia obsługę tego typu problemów. Natomiast podział paczki archiwalnej na wiele plików powodowałby potrzebę dużo bardziej skomplikowanych algorytmów doboru paczek do zapisania na kasie, jak również dużo bardziej skomplikowanego mechanizmu przekazującego paczkę archiwalną pomiędzy węzłami Systemu archiwizacji.

W pierwotnej koncepcji założono, że w pierwszym kroku powinien nastąpić zrzut baz danych w postaci plików DUMP, a następnie powinna nastąpić kompresja tych plików z użyciem GZIP. Jednakże analiza tego aspektu doprowadziła do wniosku, że można wykonać zrzut bazy danych w postaci plików DUMP kompresując go już podczas wykonywania zrzutu za pomocą narzędzia GZIP. Dzięki temu wygenerowany plik zrzutu zostaje zapisany na dysku

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

już w postaci skompresowanej, więc miejsce na macierzy pomocniczej wykorzystywanej przez Archiwizator może być lepiej wykorzystane.

#### 6.4.4.5 Nazwa pliku paczki archiwalnej

Nazwa paczki archiwalnej powinna składać się z następujących elementów:

1. Stały tekst: „MetadataDump”
2. Znacznik czasowy z momentu stworzenia paczki archiwalnej
3. Identyfikator klienta Systemu archiwizacji, który dla Repozytorium będzie cyfrą „1”
4. Stały tekst: „Owner\_BibliotekaNarodowa”

Znacznik czasowy powinien być w formacie „Unix time”, który oznacza liczbę sekund, które upłynęły od 1 stycznia 1970 wykorzystując strefę czasową UTC.

Przykład nazwy pliku został zaprezentowany poniżej:

Id\_MetadataDump#Time\_1475139982#Source\_1#Owner\_BibliotekaNarodowa.TAR

#### 6.4.5 Przekazanie żądania archiwizacji paczki do instancji Archiwum

W ramach tego zadania Archiwizator powinien przekazać żądanie archiwizacji paczki do wszystkich powiązanych z nim instancji Archiwum. Przekazane żądanie może dotyczyć obiektu lub też zrzutu metadanych.

Archiwizator przekazując paczkę archiwalną powinien przekazać informacje, czy paczka powinna podlegać replikacji. Informacja ta powinna być przekazana na podstawie następujących danych:

1. Paczka archiwalna z obiektem – jeżeli metadana obiektu o nazwie „replicate” posiada wartość „true”, to paczka powinna zostać zreplikowana w ramach instancji Archiwum. Założono, że metadana „replicate” będzie typu „string”, dlatego też sprawdzenie jej wartości powinno odbyć się poprzez porównanie tekstów.
2. Paczka archiwalna ze zrzutem metadanych – zgodnie z parametrem zdefiniowanym w konfiguracji Archiwizatora (parametr w konfiguracji: **metadataDump.replication**).

Archiwizator powinien przechowywać informacje, do których instancji Archiwum przekazał żądanie archiwizacji, aby być następnie w stanie określić, czy archiwizacja danej paczki zakończyła się sukcesem i tym samym obsłużyć zdarzenie zakończenia procesu archiwizacji paczki archiwalnej z obiektem oraz z metadanymi.

Zastosowanie systemu kolejkowego do zapewnienia komunikacji między Archiwizatorem i Archiwum wyklucza potrzebę zapewnienia mechanizmu ponowienia wysłania przez Archiwizator żądania w przypadku niedostępności instancji Archiwum.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

#### 6.4.6 Odebranie informacji o statusie zakończenia zadania archiwizacji paczki w instancji Archiwum wraz z podjęciem kolejnych kroków

W ramach tego zadania po odebraniu od instancji Archiwum informacji o statusie zleconego żądania archiwizacji paczki w danej instancji Archiwum, Archiwizator w zależności od statusu żądania powinien:

##### 1. Jeżeli żądania archiwizacji zakończyło się powodzeniem

Archiwizator powinien zweryfikować, czy paczka jest już zarchiwizowana we wszystkich instancjach Archiwum powiązanych z Archiwizatorem w momencie przekazania paczki do instancji Archiwum. Jeżeli paczka jest skutecznie zarchiwizowana, to Archiwizator powinien przejść do czynności zakończeniowych procesu archiwizacji, których przebieg zależy od tego czy:

- Paczka zawiera obiekt – wówczas wykonanie czynności opisanych w rozdziale „Obsługa zakończenia procesu archiwizacji paczki z obiektem”.
- Paczka zawiera metadane – wówczas wykonanie czynności opisanych w rozdziale „Obsługa zakończenia procesu archiwizacji paczki z metadanymi”.

##### 2. Jeżeli żądanie archiwizacji zakończyło się niepowodzeniem

W tym przypadku Archiwizator powinien w zależności od otrzymanej przyczyny niepowodzenia podjąć dalsze kroki opisane w poniższych podpunktach. Zakłada się następujące przyczyny niepowodzenia zadania archiwizacji:

- Plik jest większy niż obsługiwane kasety w tej instancji  
W tym przypadku powinno nastąpić zalogowanie błędu wraz z powiadomieniem administratora.
- Plik jest większy niż całkowite wolne miejsce na macierzy pomocniczej  
W tym przypadku powinno nastąpić zalogowanie błędu wraz z powiadomieniem administratora

W przypadku wystąpienia jednego z wyżej wymienionych problemów, żądanie archiwizacji jest kontynuowane w pozostałych instancjach Archiwum, przy czym całe żądanie archiwizacji kończy się niepowodzeniem oraz ustawieniem w obiekcie statusu PROBLEM, dzięki czemu wyklucza się wybieranie tego obiektu do ponownej archiwizacji dając jednocześnie administratorowi możliwość i czas na weryfikację przyczyny problemu. Przyjęto, że w ramach tego projektu nie zostanie przygotowane narzędzie do wyjścia z tego statusu, dopuszczając tym samym, że wyjście z tego statusu będzie wymagało wykonania modyfikacji statusu obiektu z poziomu bazy danych.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

#### 6.4.7 Obsługa zakończenia procesu archiwizacji paczki z obiektem

Po skutecznej archiwizacji paczki z obiektem Archiwizator powinien odnotować ten fakt w swojej bazie informacji, a następnie powinien przystąpić do następujących kroków:

##### 1. Aktualizacja daty ostatniej archiwizacji

W ramach tego kroku system powinien zmodyfikować w obiekcie datę ostatniej archiwizacji – wypełniając ją datą rozpoczęcia zadania pobrania plików i przygotowania paczki archiwalnej.

##### 2. Aktualizacja statusów stream'ów

W ramach tego kroku system powinien zmodyfikować status tych stream'ów, które zostały zarchiwizowane – status powinien zostać zmieniony na ARCHIVED. Chodzi o właściwą obsługę sytuacji, kiedy podczas trwania archiwizacji w kontencie zostanie dodany nowy stream – wówczas ten stream nie może zostać oznaczony jako ARCHIVED ponieważ nie został zarchiwizowany.

##### 3. Aktualizacja statusu obiektu w Repozytorium

W ramach tego punktu powinno nastąpić zaktualizowanie statusu obiektu z ARCHIVING na ARCHIVED.

Analiza możliwości oraz usunięcie plików z obiektu będzie realizowane przez osobną usługę, która będzie usuwała pliki zarówno z obiektów niedawno zarchiwizowanych, jak i z obiektów przywróconych po upływie okresu karencji od przywrócenia.

#### 6.4.8 Obsługa zakończenia procesu archiwizacji paczki z metadanymi


Po skutecznej archiwizacji paczki ze zrzutem metadanych Archiwizator powinien odnotować ten fakt w swojej bazie informacji oraz usunąć paczkę archiwalną ze swojej macierzy pomocniczej.

#### 6.4.9 Przyjęcie żądania przywrócenia obiektu / udostępnienia paczki archiwalnej

W ramach tego zadania Archiwizator powinien obsłużyć otrzymane żądanie poprzez realizację następujących kroków:

1. Weryfikację, czy obiekt lub zrzut metadanych o podanej wersji w żądaniu jest zarchiwizowany w ramach Systemu archiwizacji. Jeżeli nie jest, to żądanie powinno zostać odrzucone wraz z zalogowaniem zdarzenia w dzienniku zdarzeń oraz poinformowaniem administratora.
2. Ustawienie obiektowi statusu RESTORING w Repozytorium – tylko jeżeli jest to proces przywracania.



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

3. Decyzję, z której instancji Archiwum przywrócić dany obiekt / udostępnić paczkę archiwalną.

Kryterium wyboru instancji Archiwum powinien być współczynnik obciążenia każdej instancji Archiwum. Przyjęto, że na obciążenie instancji Archiwum wpływa bezpośrednio sumaryczny rozmiar plików, które oczekują na archiwizację (zapis) oraz odczyt. Na podstawie tych rozmiarów oraz skonfigurowanych prędkości odczytu i zapisu w danej instancji Archiwum można obliczyć całkowity czas oczekujących zapisów oraz odczytów. Z drugiej strony każda instancja Archiwum może dysponować różną liczbą napędów, dlatego w ramach obliczeń należy wziąć pod uwagę liczbę tych napędów, dzieląc sumaryczny czas operacji do wykonania właśnie przez liczbę napędów.

Przyjęto, że skonfigurowane prędkości odczytu i zapisu będą ręcznie dobierane przez operatora Systemu archiwizacji.

Współczynnik obciążenia instancji Archiwum powinien być liczony w następujący sposób:

**Dane wejściowe:**

- Ln : Liczba napędów
- PpO : Przyjęta prędkość odczytu (parametr w konfiguracji: `drive.average.read.speed`)
- PpZ : Przyjęta prędkość zapisu (parametr w konfiguracji: `drive.average.write.speed`)
- RozO : Sumaryczny rozmiar plików oczekujących na odczytanie
- RozZ : Sumaryczny rozmiar plików oczekujących na zapisanie

**Obliczenia:**

[Odczyty potrważają]:  $[RozO] / [PpO]$

[Zapisy potrważają]:  $[RozZ] / [PpZ]$

W powyższych obliczeniach należy zwrócić uwagę na jednostki, tak aby finalnie uzyskać całkowity czas oczekujących odczytów oraz całkowity czas oczekujących zapisów.

**Wynik:**

Współczynnik obciążenia:  $([Odczyty\ potrważają] + [Zapisy\ potrważają]) / [Ln]$

4. Przekazanie wybranej w poprzednim kroku instancji Archiwum zadania odczytania paczki archiwalnej z obiektem / ze zrzutem metadanych. Instancja Archiwum powinna potwierdzić przyjęcie zadania odczytania. Jeżeli tego nie zrobi w skonfigurowanym czasie (parametr w konfiguracji: `file.retrieval.acceptanceTimeout`), to Archiwizator przekaże zadanie do jednej z

Projekt współfinansowany w ramach Programu Operacyjnego Polska Cyfrowa z Europejskiego Funduszu Rozwoju Regionalnego i budżetu państwa na podstawie Umowy o dofinansowanie nr POPC.02.01.00-00-0043/16-00



Fundusze  
Europejskie  
Polska Cyfrowa



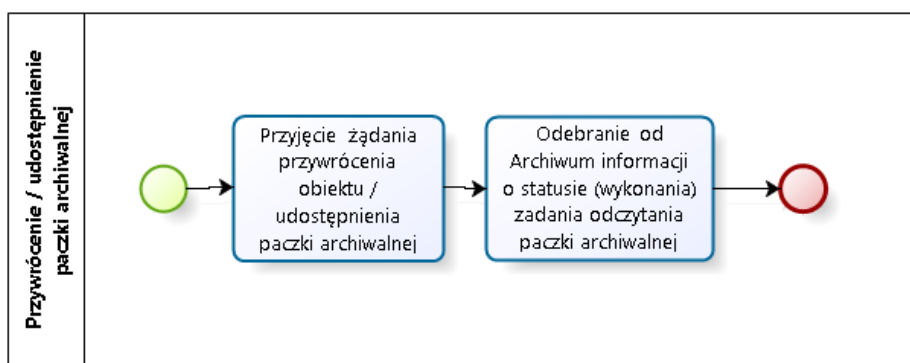
Unia Europejska  
Europejski Fundusz  
Rozwoju Regionalnego



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

pozostałych instancji Archiwum – decydując o wyborze instancji wykorzystując wartość współczynnika obciążenia. Dodatkowo to zdarzenie powinno zostać zapisane w dzienniku zdarzeń wraz z wysłaniem powiadomienia do administratora.

Poniżej zamieszczony został poglądowy diagram procesu przywrócenia obiektu / udostępnienia paczki archiwalnej. Diagram ten ma za zadanie przedstawić kolejne kroki procesu odwzorowane w niniejszym dokumencie jako rozdziały, bez uwzględniania aspektu asynchroniczności poszczególnych kroków.



Powered by  
**bizagi**  
Modeler


#### 6.4.10 Przyjęcie od instancji Archiwum ządania odczytu paczki archiwalnej

W przypadku wystąpienia procesu migracji kasety w ramach instancji Archiwum, może wystąpić potrzeba wykorzystania kopii paczki archiwalnej z innej instancji Archiwum. Wówczas instancja Archiwum powinna wysłać do Archiwizatora ządanie odczytu paczki archiwalnej z innej instancji Archiwum, a Archiwizator powinien przyjąć takie ządanie oraz wykonać następujące kroki:

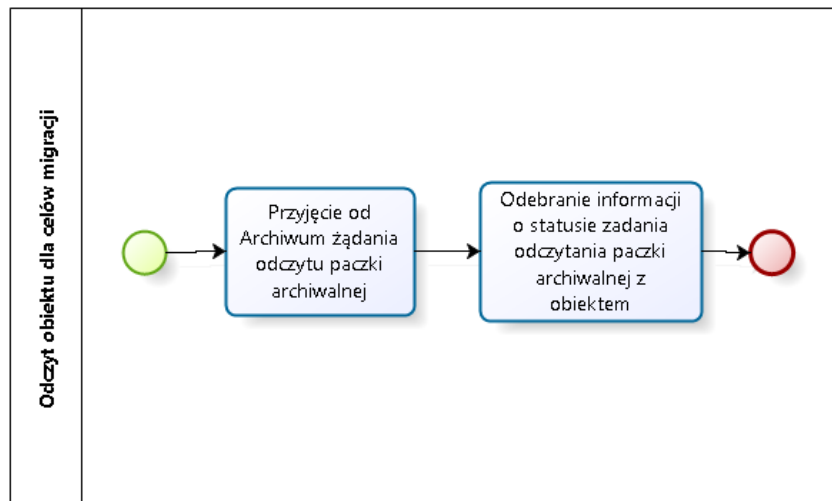
1. Weryfikacja, czy paczka archiwalna w podanej w ządaniu wersji jest zarchiwizowana oraz czy jest aktualnie dostępna w innej instancji Archiwum (ponieważ może zajść sytuacja, w której wszystkie pozostałe kopie nie są dostępne / również są w trakcie migracji).
2. Decyzja, z której instancji Archiwum przywrócić dany obiekt. Decyzja ta powinna zostać podjęta na podstawie współczynników obciążenia instancji Archiwów opisanych w rozdziale dotyczącym przyjęcia ządania przywrócenia obiektu / udostępnienia paczki archiwalnej.
3. Zlecenie odczytania paczki archiwalnej z obiektem do wybranej instancji Archiwum.

Poniżej zamieszczony został poglądowy diagram procesu odczytu obiektu dla celów migracji kasety wykonywanego w jednej z instancji Archiwum. Diagram ten ma za zadanie



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

przedstawić kolejne kroki procesu odwzorowane w niniejszym dokumencie jako rozdziały, bez uwzględniania aspektu asynchroniczności poszczególnych kroków.



Powered by  
**bizagi**  
Modeler

#### 6.4.11 Odebranie informacji o statusie zakończenia zadania odczytu paczki archiwalnej wraz z podjęciem kolejnych kroków

W ramach tego zadania po odebraniu od instancji Archiwum informacji o statusie zleconego zadania odczytu paczki archiwalnej z danej instancji Archiwum, Archiwizator w zależności od statusu zadania powinien:

1. Jeżeli zadanie odczytu zakończyło się powodzeniem i zadanie odczytu było zainicjowane w ramach procesu przywrócenia obiektu do Repozytorium:
  - a. Rozpakowanie paczki archiwalnej  
W ramach tego kroku powinno nastąpić rozpakowanie pliku paczki archiwalnej.
  - b. Zainicjowanie przywrócenia obiektu do Repozytorium  
W ramach tego kroku Archiwizator powinien wykonać procedurę przywrócenia obiektu do Repozytorium wykorzystując funkcjonalność Przywracacza.
2. Jeżeli zadanie odczytu zakończyło się powodzeniem i zadanie odczytu było zainicjowane w ramach procesu udostępnienia paczki archiwalnej, to w ramach tego kroku Archiwizator powinien przegrać paczkę archiwalną na zasób sieciowy wskazany w wywołaniu żądania.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

3. Jeżeli zadanie odczytu zakończyło się powodzeniem i zadanie odczytu było zainicjowane w ramach procesu migracji – Archiwizator powinien przekazać plik paczki archiwalnej do instancji Archiwum, która była autorem zadania odczytu.

4. Jeżeli zadanie odczytu zakończyło się niepowodzeniem

W tym przypadku Archiwizator powinien w zależności od otrzymanej przyczyny niepowodzenia podjąć dalsze kroki opisane w poniższym podpunkcie. Zakłada się następującą przyczynę niepowodzenia zadania odczytu paczki archiwalnej – brak poprawnej kopii paczki o podanym identyfikatorze. W tym przypadku powinno nastąpić zalogowanie zdarzenia w dzienniku zdarzeń wraz z powiadomieniem administratora oraz zlecenie odczytu paczki do innej instancji Archiwum.

#### 6.4.12 Monitorowanie statusu przywrócenia obiektu

W ramach tego zadania Archiwizator powinien dla oczekujących żądań przywrócenia oraz udostępnienia obiektu monitorować okres obsługi przez System archiwizacji tego żądania. W przypadku upływu skonfigurowanego czasu powinno nastąpić zalogowanie zdarzenia w dzienniku zdarzeń oraz wysłanie powiadomienia do administratora (parametr w konfiguracji: **expected.retrieval.time**). Nie powinno to jednak spowodować przerwania realizacji tego żądania.

### 6.5 Interfejs API Systemu archiwizacji

#### 6.5.1 Specyfikacja interfejsu API

Specyfikacja interfejsu API udostępnionego przez System archiwizacji zgodnego ze wzorcem architektonicznym REST jest dostępna w postaci odrębnego pliku, stanowiącego załącznik nr 1 do Projektu technicznego.

Z uwagi na to, że część operacji uruchamianych przez wywołanie interfejsu może trwać dłuższy czas, wynik operacji (np. przywrócenie obiektu) nie będzie mogło zostać wykonane synchronicznie. Dlatego zakłada się, że autor żądania będzie okresowo odpytywał System archiwizacji, używając udostępnionego interfejsu REST, czy zlecone zadanie jest już ukończone:

1. W przypadku ukończenia zadania, zostanie zwrócony:
  - a. Status ukończenia.
  - b. Ewentualne pozostałe dane będące wynikiem zadania.
2. W przypadku, kiedy zadanie nie będzie ukończone, zostanie zwrócona informacja, po jakim czasie autor żądania powinien ponownie sprawdzić status tego zadania. Czas ten będzie konfigurowalny (parametr w konfiguracji: **task.status.checkInterval**). W przyszłości powinna być możliwość dodania funkcjonalności obliczania tego czasu.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Przyjęto, że w przypadku wywołania, które skutkuje przysłaniem danych przez webservice, dane te powinny zostać przekazane w formacie JSON.

### 6.5.2 Uwierzytelnianie

Uwierzytelnianie zostanie zrealizowane poprzez czasowy token wystawiany dla użytkownika. Token będzie wystawiany na skonfigurowany w parametrach czas ważności (parametr w konfiguracji: `api.token.lifespan`). Wywołanie udostępnionego interfejsu REST z wykorzystaniem tokena, pod warunkiem jego ważności w momencie wysłania żądania, będzie przedłużało jego ważność o skonfigurowany w parametrach czas (parametr w konfiguracji: `api.token.lifespan.prolongation`). Token po upływie czasu ważności będzie wygaszany.

W ramach Systemu archiwizacji powinna istnieć baza użytkowników/klientów Systemu archiwizacji, uprawnionych do korzystania z udostępnionego interfejsu REST. Każdy użytkownik powinien posiadać nazwę, hasło oraz identyfikator klienta Systemu archiwizacji. Wykorzystując polecenia w konsoli systemu operacyjnego (z rodziny Linux) powinna być możliwość:

1. Wyświetlania listy użytkowników.
2. Dodawania użytkowników.
3. Edycji użytkowników – ponownego ustawienia hasła.
4. Oznaczania użytkowników jako nieaktywnych.

### 6.5.3 Zapisywanie zdarzeń wywołania interfejsu

Wywołanie metod z interfejsu powinno powodować zapisywanie zdarzeń w dzienniku zdarzeń.

W ramach zdarzenia w dzienniku powinny być zapisywane następujące informacje:

1. Data i godzina UTC
2. Identyfikator hosta
3. Wywołana metoda, np. GET, PUT
4. Poziom zdarzenia
5. Pełna ścieżka/adres URL wywołania
6. Status operacji, np. 200, 403.
7. Zawartość nagłówka http X-NL-AUTH-TOKEN żądania
8. Zawartość body http żądania, w przypadku wywołania `/api/token/` POST „Wygenerowanie i pobranie tokena dla wskazanego użytkownika” gdzie w body jest podany login i hasło, hasło nie powinno być zalogowane
9. Zawartość body http odpowiedzi.

W przypadku wystąpienia statusu 403 Forbidden, zdarzenie to powinno posiadać poziom WARNING. W pozostałych przypadkach zdarzenie powinno otrzymać poziom INFO.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Jeżeli brak jednej z informacji np. parametrów wywołania URL, to zawartość tego segmentu w zdarzeniu należy pozostawić pustą uwzględniając jednak wszystkie separatory pipe „|” przed i po tym segmencie.

Format i kolejność zdarzeń powinna być zgodna z następującą specyfikacją:

```
<Data i godzina UTC>|<Identyfikator hosta>|<Metoda>|<Poziom          zdarzenia>|<Status
operacji>|<Pełna ścieżka/adres URL wywołania>|<Zawartość nagłówka HTTP>|<Zawartość body
HTTP>
```

W przypadku, w którym jeden z powyższych segmentów będzie wielolinijkowy (np. zawartość body HTTP), dopuszcza się, że jedno zdarzenie będzie w pliku zdarzeń zapisane w więcej niż jednej linii.

Wysyłanie powiadomień do administratora powinno występować tylko dla zdarzeń z poziomem WARNING. Powinien zostać zastosowany identyczny mechanizm wysyłania powiadomień, co w przypadku pozostałych modułów Systemu archiwizacji.

## 6.6 Przywracacz

Funkcjonalność Przywracacza ma za zadanie zrealizowanie zadania przywrócenia plików do obiektu w Repozytorium.

Funkcjonalność ta powinna być wykorzystywana przez Archiwizator w procesie przywracania obiektu do Repozytorium.

Przywracacz nie jest osobnym komponentem, a jedynie funkcjonalnością w ramach Archiwizatora. Dlatego nie występuje potrzeba przekazywania, w rozumieniu przykładowo interfejsu API, paczki archiwalnej z obiektem z Archiwizatora do Przywracacza.

Przed rozpoczęciem przywracania paczka obiektu powinna zostać rozpakowana. Aby wyeliminować sytuację, kiedy proces rozpakowania zostanie rozpoczęty, ale w trakcie jego trwania miejsce na macierzy pomocniczej się wyczerpie, powinien istnieć prosty mechanizm rezerwacji miejsca dla rozpoczynanych zadań mogących wpłynąć na ilość wolnego miejsca na macierzy pomocniczej.

### 6.6.1 Przywrócenie plików do obiektu w Repozytorium

Przywrócenie zawartości obiektu powinno polegać na przywróceniu do Repozytorium usuniętych plików stream'ów pochodzących z paczki archiwalnej z uwzględnieniem faktu, że kontent, a więc również stream'y, może należeć do wielu obiektów.

Przed przystąpieniem do przywrócenia najpierw należy sprawdzić, jaki status posiada obiekt:

1. Jeżeli posiada RESTORING, to można przystąpić do wykonania przywrócenia zawartości obiektu.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

2. Jeżeli posiada inny niż RESTORING (np. ARCHIVED), to przywrócenie zawartości obiektu nie powinno zostać wykonane.

Przywrócenie zawartości obiektu powinno składać się z następujących kroków:

1. **Identyfikacja stream'ów znajdujących się w paczce archiwalnej**

W ramach tego kroku na podstawie informacji w paczce METS należy zidentyfikować stream'y dostępne w paczce archiwalnej.

Lista ta powinna zostać zbudowana na podstawie jednego z następujących sposobów, wybranego na etapie implementacji:

- a. Lista wynika z ostatniej sekcji nazwy plików stream'ów w paczce archiwalnej.
- b. Lista wynika z danych w sekcji METS o nazwie fileSec, w której zapisane są relacje nazw plików do identyfikatorów stream'ów i nazw stream'ów – konwencja nazywania identyfikatorów plików jest streamID\_streamName.

Unikalnym identyfikatorem stream'ów jest streamID.

Jeżeli chodzi o pliki stream'ów, to znajdują się one w paczce archiwalnej w podfolderach, których nazwą są poszczególne klucze stream'ów (np. mainStream, download\_alto, text, download\_fullJPG, download\_lowJPG).

Pliki stream'ów posiadają nazwy zgodnie z poniższym schematem:

<identyfikator obiektu z Repozytorium>separator<nazwa typu streamu>separator<kolejność w obiekcie>separator<Identyfikator      kontentu      z repozytorium>separator<**identyfikator streamu**>.rozszerzenie pliku

2. **Identyfikacja stream'ów wymagających przywrócenia**

W ramach tego kroku należy przeanalizować w obiekcie stream'y wymagające przywrócenia.

Lista ta powinna zostać zbudowana na podstawie stream'ów znajdujących się obecnie w obiekcie i posiadających status DELETED. Takie kryteria wykluczają streamy nieusunięte oraz nowe.

Unikalnym identyfikatorem stream'ów jest streamID.

3. **Sprawdzenie możliwości przywrócenia obiektu**

W ramach tego kroku powinno nastąpić sprawdzenie, czy wszystkie stream'y wymagające przywrócenia (lista stworzona w ramach kroku 2) są dostępne w paczce archiwalnej (lista stworzona w ramach kroku 1).

- a. Jeżeli w paczce archiwalnej znajdują się wszystkie wymagane streamy, można przejść do kolejnego kroku – próby przywrócenia stream'ów.

Projekt współfinansowany w ramach Programu Operacyjnego Polska Cyfrowa z Europejskiego Funduszu Rozwoju Regionalnego i budżetu państwa na podstawie Umowy o dofinansowanie nr POPC.02.01.00-00-0043/16-00



Fundusze  
Europejskie  
Polska Cyfrowa



Unia Europejska  
Europejski Fundusz  
Rozwoju Regionalnego



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

- b. Jeżeli w paczce archiwalnej brakuje chociaż jednego streama, należy wykonać następujące czynności:
  - i. Ustawienie w Repozytorium statusu obiektu na PROBLEM.
  - ii. Zalogowanie zdarzenia w dzienniku zdarzeń wraz z poinformowaniem administratora.
  - iii. Zakończenie niepowodzeniem procesu przywrócenia obiektu.
- c. Jeżeli obiekt nie zawiera streamów do przywrócenia (przykładowo w międzyczasie użytkownik usunął z obiektu kontenty zawierające zredukowane streamy), to etap przywracania streamów powinien zostać pominięty i można przejść do kroku aktualizacji statusu przywracanego obiektu.

#### 4. Przywrócenie stream'ów

W ramach tego kroku należy wykonać operacje przywrócenia streamów polegającą na zapisaniu w HCP pliku pod tym samym identyfikatorem jaki znajduje się zarówno w HCP jak i w nazwie pliku w paczce archiwalnej.

Po przywróceniu stream'a jego status powinien zostać zmieniony z DELETED na ARCHIVED.

W przypadku wystąpienia problemu z przywróceniem pliku (np. chwilowa niedostępność HCP), powinno nastąpić zapisanie zdarzenia w dzienniku zdarzeń. Następnie po zdefiniowanym w konfiguracji czasie (parametr w konfiguracji: **restoration.waitForHcpDelay**) powinna nastąpić ponowna próba przywrócenia pliku:

- a. Jeżeli operacja się powiedzie, należy kontynuować przywracanie stream'ów.
- b. Jeżeli ponownie wystąpi błąd, należy wykonać następujące czynności:
  - i. Ustawienie w Repozytorium statusu obiektu na PROBLEM.
  - ii. Zalogowanie zdarzenia w dzienniku zdarzeń wraz z poinformowaniem administratora.
  - iii. Zakończenie niepowodzeniem procesu przywracania obiektu.

#### 5. Aktualizacja statusu przywracanego obiektu


W ramach tego kroku należy zmienić status obiektu na ARCHIVED oraz zaktualizować datę ostatniego przywrócenia obiektu.

Dodatkowo należy zarejestrować w dzienniku zdarzenie zakończone sukcesem przywrócenia obiektu.

Przywrócenie obiektu nie powinno powodować aktualizacji daty ostatniej modyfikacji obiektu w kontekście archiwizacji.

Dzięki wykorzystaniu podczas przywracania identyfikatorów streamów, nie występuje potrzeba uwzględniania i analizowania sytuacji, kiedy to kontenty w Repozytorium są w innej kolejności.



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

W przypadku, w którym występuje współdzielenie kontentów, a więc streamy/pliki nie przynależą tylko do jednego obiektu, to podczas przywracania jednego obiektu może nastąpić przywrócenie części kontentów i streamów należących też do innego lub innych obiektów. Ustalono, że w przypadku wystąpienia takiej sytuacji nie jest zasadne analizowanie statusu pozostałych obiektów.

Podczas analizy zagadnienia przywracania zawartości obiektu wzięto pod uwagę fakt, iż przed rozpoczęciem przywrócenia zawartości obiektu należy zweryfikować, czy wszystkie wymagane w obiekcie streamy znajdują się w paczce archiwalnej. Ta weryfikacja jest wykonywana w powyżej opisanych krokach o numerach 1-3, i w przypadku wystąpienia braku chociażby jednego streamu, obiekt jest odpowiednio oznaczany a informacja o tym zdarzeniu jest zapisywana w dzienniku zdarzeń wraz z poinformowaniem administratora. Dodatkowo podczas analizy przypadków brzegowych okazało się, że może zdarzyć się sytuacja, w której w momencie przywrócenia zawartości obiektu, obiekt nie posiada streamów do przywrócenia. Dlatego w ramach opisanych powyżej kroków przebiegu przywrócenia uwzględniono taki przypadek.

## 6.7 Weryfikator

Treść rozdziału nie załączona ze względu na brak związku z przedmiotowym zamówieniem.

## 7 Wymagania funkcjonalne zmian w Repozytorium

W niniejszym rozdziale zostały umieszczone wymagania dotyczące potrzebnych zmian w Repozytorium.

### 7.1 Zmiany encji obiekt

#### 7.1.1 Status obiektu w kontekście archiwizacji

W ramach realizowanych zmian w Repozytorium powinien zostać dodany nowy status w obiekcie, status mówiący o tym, w jakim stanie pod kątem archiwizacji jest obiekt. Status ten będzie aktualizowany przez System archiwizacji.

Diagram stanów i możliwości przejść między nimi został przedstawiony poniżej. Założenia do poniższego diagramu są następujące:

1. Po uruchomieniu Systemu archiwizacji, wszystkie obiekty będą posiadały status NEW. Po zmianie statusu NEW na inny – obiekt już nigdy go nie osiągnie.
2. Na niebiesko zaznaczono stany, w których obiekt jest zredukowany w Repozytorium, czyli chociaż jeden stream jest usunięty.
3. Na zielono zaznaczono stany, w których obiekt jest dostępny w Repozytorium, czyli ani jeden stream nie jest usunięty.
4. Pod pojęciem MODIFIED\_ON należy rozumieć sytuację, w której „Data ostatniej modyfikacji w kontekście archiwizacji” jest późniejsza niż „Data ostatniej archiwizacji”

Projekt współfinansowany w ramach Programu Operacyjnego Polska Cyfrowa z Europejskiego Funduszu Rozwoju Regionalnego i budżetu państwa na podstawie Umowy o dofinansowanie nr POPC.02.01.00-00-0043/16-00




Fundusze  
Europejskie  
Polska Cyfrowa



Unia Europejska  
Europejski Fundusz  
Rozwoju Regionalnego





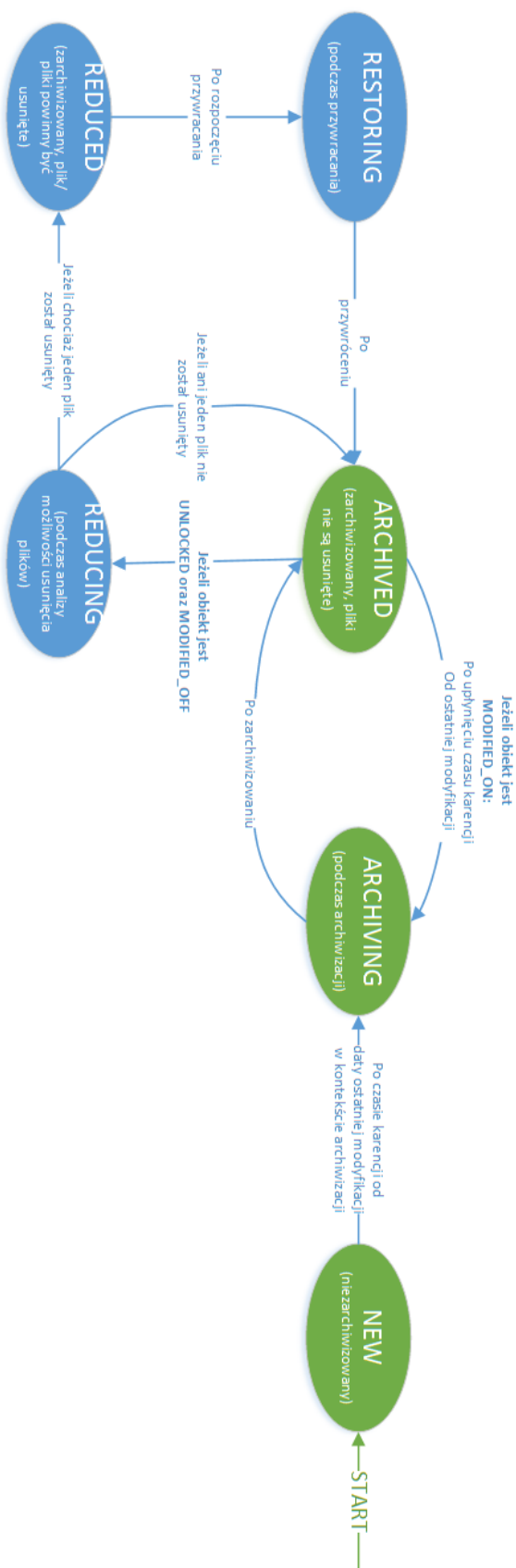
WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

5. Po zakończeniu procesu próby usunięcia plików z obiektu status obiektu powinien zmienić się zgodnie z poniższym opisem:

- a. Jeżeli został usunięty chociażby 1 stream, to obiekt powinien przyjąć status REDUCED.
- b. Jeżeli nie został usunięty ani 1 stream, to obiekt powinien przyjąć status ARCHIVED.

Dodatkowo obiekt będzie mógł przyjąć status PROBLEM oznaczający, że podczas ostatniego procesu archiwizacji lub przywracania wystąpił problem z tym obiektem. Status ten będzie umożliwiał wykluczenie takiego obiektu podczas wybierania obiektów do ponownego procesowania.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

### 7.1.2 Blokowanie obiektu (tzw. LOCK)

W ramach realizowanych zmian w Repozytorium powinien zostać dodany mechanizm blokowania obiektu. Chodzi o sytuacje, kiedy przykładowo proces OCR lub zamówień na pliki będzie za niedługo potrzebował danego obiektu, i chce zapobiec ewentualnej redukcji (usunięciu streamów) tego obiektu.

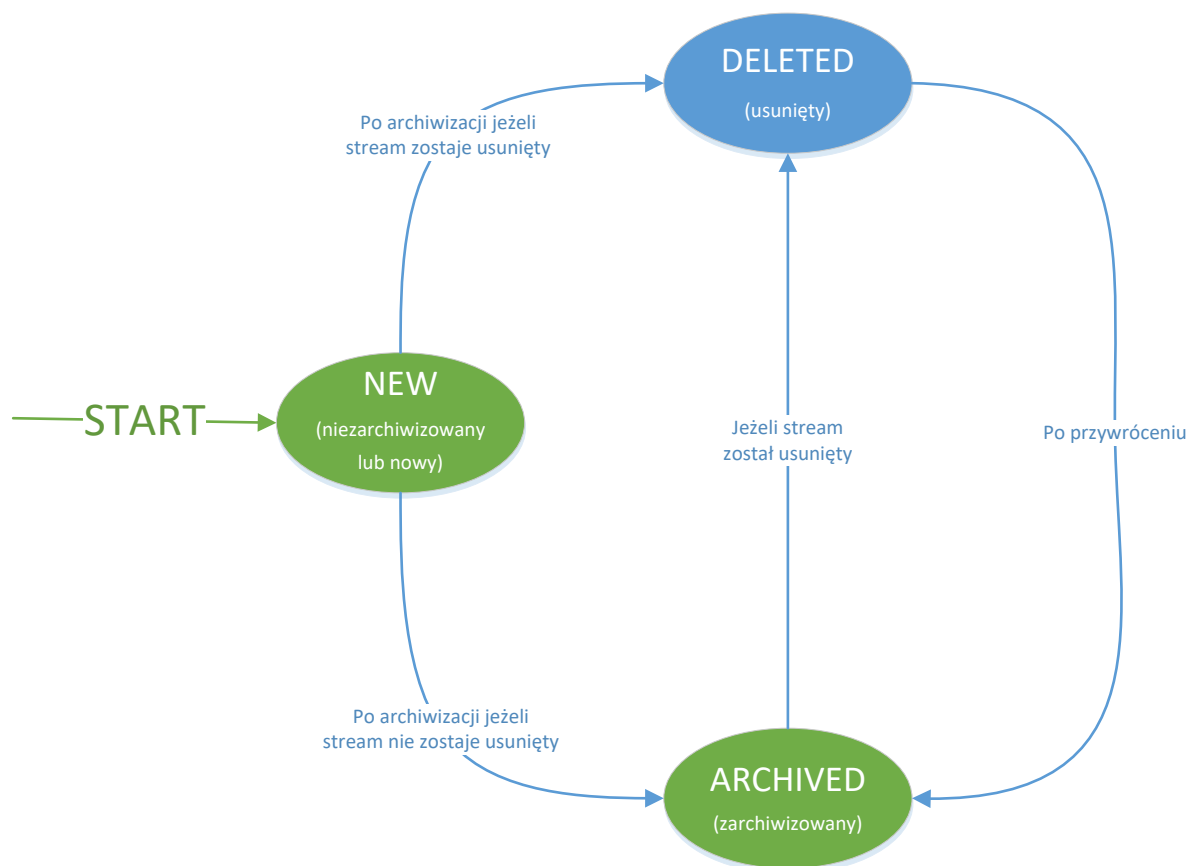
Status ten będzie aktualizowany przez Repozytorium, po stronie Systemu archiwizacji będzie sprawdzanie statusu zablokowania w kilku miejscach zdefiniowanych na diagramie statusów obiektu w miejscach, w których występuje sprawdzenie warunku LOCKED oraz UNCLOCKED.

### 7.1.3 Status streama w kontekście archiwizacji

W ramach realizowanych zmian w Repozytorium powinien zostać dodany nowy status w stream'ie mówiący o tym, w jakim stanie pod kątem archiwizacji jest stream. Status ten będzie aktualizowany przez System archiwizacji.

Diagram stanów i możliwości przejść między nimi został przedstawiony poniżej. Założenia do poniższego diagramu są następujące:

1. Po uruchomieniu Systemu archiwizacji, wszystkie streamy będą posiadały status NEW. Nowy stream (w rozumieniu dodany do istniejącego contentu) również będzie otrzymywał status NEW.
2. Na niebiesko zaznaczono stany, w których stream w Repozytorium jest zredukowany.
3. Na zielono zaznaczono stany, w których stream w Repozytorium jest dostępny.



#### 7.1.4 Modyfikacja obiektu w kontekście archiwizacji

Przez modyfikację obiektu w kontekście archiwizacji rozumiane są następujące zmiany w obiekcie:

1. Dodanie nowego kontentu: postaci zbiorczej lub kontentu w strukturze paginacyjnej z przynajmniej jednym stream'em.
2. Dodanie lub podmiana (np. przez proces OCR) w przynajmniej jednym istniejącym kontencie stream'a znajdującego się na konfigurowalnej liście archiwizowanych stream'ów (konfiguracja brana pod uwagę podczas budowania paczki archiwalnej).

Modyfikacja obiektu w kontekście archiwizacji powinna powodować aktualizację daty ostatniej modyfikacji w kontekście archiwizacji.

Uwaga: przywrócenie plików przez System archiwizacji do zarchiwizowanego obiektu nie może powodować aktualizacji daty ostatniej modyfikacji.

Aby System archiwizacji mógł prawidłowo zakwalifikować obiekt do archiwizacji, powinien zostać sprawdzony warunek czy minął okres karencji od zakończenia przez obiekt workflow. Sprawdzanie daty zakończenia workflow w oparciu o datę wykonania ostatniego kroku workflow (encja „activity” w bazie danych Repozytorium) może powodować nadmierne

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

obciążenie bazy danych Repozytorium, dlatego planowane jest inne rozwiązanie polegające na tym, aby po wykonaniu każdego kroku workflow aktualizowana była data ostatniej modyfikacji w kontekście archiwizacji. Dzięki czemu będzie możliwe sprawdzenie wspólnego okresu karencji od zakończenia workflow i ostatniej modyfikacji w kontekście archiwizacji, bez potrzeby odwoływania się do encji „activity” w bazie danych Repozytorium.

### 7.1.5 Konfiguracja okresów karencji

Powinna być możliwość skonfigurowania okresu karencji od ostatniej modyfikacji w kontekście archiwizacji, po którym powinna zostać wykonana ponowna archiwizacja, oraz okresu karencji od ostatniego przywrócenia, po którym powinna zostać wykonana próba zredukowania obiektu.

Z uwagi na to, że ta konfiguracja będzie wykorzystywana zarówno przez System archiwizacji, jak i przez Repozytorium, to konfiguracja ta powinna spełniać następujące wymagania:

1. Konfiguracja okresu karencji od przywrócenia i od ostatniej modyfikacji w kontekście archiwizacji powinna być przechowywana w bazie danych Repozytorium – tabela **archival\_config**:
  - a. **modification\_timeout** – po jakim czasie od ostatniej modyfikacji obiektu w kontekście archiwizacji, usługa pomocnicza Systemu archiwizacji może przystąpić do archiwizacji obiektu (zarówno pierwszej archiwizacji, jak i kolejnej). Czas ten jest podawany w dniach jako liczba całkowita.
  - b. **restore\_timeout** – po jakim czasie od przywrócenia obiektu do Repozytorium, usługa pomocnicza Systemu archiwizacji można przystąpić do redukcji obiektu. Zostanie przygotowana metoda do pobierania wartości tych konfiguracji z Repozytorium – z której to będzie korzystał System archiwizacji. Czas ten jest podawany w dniach jako liczba całkowita.
2. W Systemie archiwizacji ta konfiguracja będzie cache’owana przez czas zdefiniowany w parametrach Systemu archiwizacji (parametr w konfiguracji: **archivization.config.refresh.delay**).

### 7.2 Pobieranie całej publikacji

Widok ten powinien zostać usunięty/wyłączony, a użytkownik powinien korzystać z zamówień na pliki.

### 7.3 Zamówienia na pliki

Powinna być dostępna funkcjonalność pozwalająca budować zamówienie na pliki z wielu obiektów.

Użytkownik powinien móc budować zamówienia używając następujących funkcjonalności:

1. Budowanie zamówienia z poziomu szczegółów obiektu.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

2. Budowanie zamówienia z poziomu listy obiektów – chodzi o widok „Wszystkie obiekty” w menu Digitalizacji.

### 7.3.1 Budowanie zamówienia z poziomu szczegółów obiektu

Przebieg procesu budowania zamówienia z poziomu szczegółów obiektu powinien być następujący:

1. Użytkownik powinien móc wyświetlić obiekt oraz przejść do widoku „Pobierz”.
2. Użytkownik powinien zobaczyć listę dostępnych plików w obiekcie z możliwością przefiltrowania jej pod względem:
  - a. Typu kontentu z możliwością wyboru jednej lub wielu z tych opcji: zbiorcze, paginacyjne podstawowe, paginacyjne alternatywne.
  - b. Kluczy stream’a z możliwością wyboru jednej lub wielu z tych opcji: np. mainStream, download\_alto, text, download\_fullJPG, download\_lowJPG, RAW. W tym widoku (wybieranie plików w widoku szczegółów obiektu) lista kluczy stream’ów powinna być budowana dynamicznie z kluczy stream’ów występujących w tym obiekcie.
  - c. Wartości z pola typ pliku/mimeType z możliwością wyboru jednej lub wielu z tych opcji. W tym widoku (wybieranie plików w widoku szczegółów obiektu) lista typów plików/mimeType powinna być budowana dynamicznie z plików występujących w tym obiekcie.

Pomiędzy poszczególnymi filtrami (powyższe punkty a, b i c) powinien zostać wykorzystany operator logiczny AND, a pomiędzy wartościami w poszczególnych filtrach operator logiczny OR.

3. Lista plików w obiekcie może być potencjalnie długa, dlatego powinna zostać wyświetlona z zastosowaniem stronicowania.
4. Lista plików powinna zawierać następujące kolumny: id contentu, numer kolejny contentu (dla zbiorczych 0), nazwę klucza streamu, mimetype pliku z bazy Repozytorium, w miarę możliwości nazwę oryginalną pliku, potem kolejność contentów w obiekcie, typ kontentu (postać zbiorcza, content podstawowy, content alternatywny).
5. Kolejność sortowania: postaci zbiorcze, contenty w strukturze paginacyjnej posortowane w wewnętrznej kolejności wyświetlania w systemie. Contenty alternatywne powinny być wyświetlone bezpośrednio pod kontentem pod który podlegają i powinny być posortowane zgodnie z kolejnością wyświetlania ich w Repozytorium.
6. Użytkownik powinien móc zaznaczyć pliki, które chce pobrać. Powinien również móc zaznaczyć i odznaczyć wszystkie pliki w tym widoku.
7. Użytkownik powinien móc potwierdzić wybór zaznaczonych plików oraz wybrać, do którego istniejącego zamówienia (w statusie „Nowe”) dodać wybrane pliki, lub czy utworzyć nowe zamówienie i do niego dodać te pliki. System powinien pamiętać, do którego zamówienia w statusie „Nowe” użytkownik dodawał ostatnio pliki, i to zamówienie powinno być domyślnie wybrane/zaznaczone podczas wybierania

Projekt współfinansowany w ramach Programu Operacyjnego Polska Cyfrowa z Europejskiego Funduszu Rozwoju Regionalnego i budżetu państwa na podstawie Umowy o dofinansowanie nr POPC.02.01.00-00-0043/16-00



Fundusze  
Europejskie  
Polska Cyfrowa

**omnis** BN

Unia Europejska  
Europejski Fundusz  
Rozwoju Regionalnego



Strona 53 z 70

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

zamówienia w tym oknie. Pokazana użytkownikowi lista zamówień powinna obejmować numer i nazwę zamówienia (w jednej linijce). Lista z zamówieniami powinna być listą wybieralną (tzw. combobox). Zamówienia na liście wyboru powinny być posortowane wg numerów rosnąco.

Podczas dodawania plików do zamówienia, system powinien weryfikować czy rozmiar zamówienia nie będzie przekraczać skonfigurowanego maksymalnego limitu rozmiaru zamówienia. Jeżeli rozmiar zamówienia będzie przekraczał skonfigurowany limit, użytkownik powinien dostać powiadomienie a proces dodawania plików z tego obiektu do zamówienia powinien zostać wycofany – dzięki temu unikana jest konieczność „dzielenia” zawartości obiektu pomiędzy kilka zamówień. Rozmiar plików z danego obiektu powinien być określany na podstawie rozmiarów plików przechowywanych w bazie danych.

8. Użytkownik powinien móc wykonać kroki 1-7 dla innych obiektów.

### 7.3.2 Budowanie zamówienia z poziomu listy obiektów

Przebieg procesu budowania zamówienia z poziomu listy obiektów powinien być następujący:

1. Użytkownik w istniejącym widoku Katalogu obiektów powinien móc uruchomić funkcjonalność pobrania plików dla wszystkich obiektów z listy, również tych z kolejnych stron wyników wyszukiwania, uwzględniając zastosowany filtr na listę obiektów. Powinien istnieć skonfigurowany limit obiektów, dla których użytkownik powinien móc wykonać operację pobrania plików, po przekroczeniu którego użytkownik powinien zobaczyć komunikat informujący o braku możliwości wykonania tej operacji dla większej liczby obiektów aniżeli skonfigurowany limit.
2. Użytkownik okno (w „drugim kroku” budowania zamówienia z wielu obiektów) powinien zobaczyć, w którym powinien móc wybrać:
  - a. Typu kontentu z możliwością wyboru jednej lub wielu z tych opcji: zbiorcze, paginacyjne podstawowe, paginacyjne alternatywne.
  - b. Kluczy stream’a z możliwością wyboru jednej lub wielu z tych opcji: np. mainStream, download\_alto, text, download\_fullJPG, download\_lowJPG, RAW. W tym widoku (lista obiektów) lista kluczy stream’ów powinna być budowana dynamicznie.

Pomiędzy poszczególnymi filtrami (powyższe punkty a, b i c) powinien zostać wykorzystany operator logiczny AND, a pomiędzy wartościami w poszczególnych filtrach operator logiczny OR.

Następnie użytkownik powinien potwierdzić wybrane filtry dotyczące stream’ów.
3. Użytkownik powinien móc wybrać, które streamy chce dodać do zamówienia. Dodatkowo użytkownik powinien móc na tym widoku przefiltrować listę streamów zaznaczając/odznaczając obiekty wybrane w poprzednich krokach.
4. Użytkownik powinien dodać streamy do zamówienia wybierając, do którego istniejącego zamówienia (w statusie „Nowe”) dodać wybrane pliki, lub czy utworzyć

Projekt współfinansowany w ramach Programu Operacyjnego Polska Cyfrowa z Europejskiego Funduszu Rozwoju Regionalnego i budżetu państwa na podstawie Umowy o dofinansowanie nr POPC.02.01.00-00-0043/16-00



Fundusze  
Europejskie  
Polska Cyfrowa


**omnis**<sup>BN</sup>

Unia Europejska  
Europejski Fundusz  
Rozwoju Regionalnego



Strona 54 z 70



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

nowe zamówienie i do niego dodać te pliki. System powinien pamiętać, do którego zamówienia w statusie „Nowe” użytkownik dodawał ostatnio pliki, i to zamówienie powinno być domyślnie wybrane/zaznaczone podczas wybierania zamówienia w tym oknie. Pokazana użytkownikowi lista zamówień powinna obejmować numer i nazwę zamówienia (w jednej linijce). Lista z zamówieniami powinna być listą wybieralną (tzw. combobox). Zamówienia na liście wyboru powinny być posortowane wg numerów rosnąco.

Podczas dodawania plików do zamówienia, system powinien weryfikować czy rozmiar zamówienia nie będzie przekraczać skonfigurowanego maksymalnego limitu rozmiaru zamówienia. Jeżeli rozmiar zamówienia będzie przekraczał skonfigurowany limit, użytkownik powinien dostać powiadomienie a proces dodawania plików z tego obiektu do zamówienia powinien zostać wycofany, a system powinien przejść do analizy możliwości dodania do zamówienia kolejnego wybranego przez użytkownika obiektu wycofany – dzięki temu unikana jest konieczność „dzielenia” zawartości obiektu pomiędzy kilka zamówień. Rozmiar plików z danego obiektu powinien być określany na podstawie rozmiarów plików przechowywanych w bazie danych.

### 7.3.3 Budowanie zamówienia a uprawnienia

W systemie istnieją uprawnienia, które regulują dostęp do obiektów i ich zawartości (w tym do pobierania) na różnych poziomach – uprawnienia te powinny być respektowane podczas dodawania plików do zamówienia. Zarówno podczas dodawania plików z poziomu pojedynczego obiektu jak i z poziomu listy obiektów system podczas dodawania plików do zamówienia powinien sprawdzić czy użytkownik ma prawo pobrać ten plik, jeżeli nie to nie powinien on zostać dodany do listy. Na koniec powinna zostać wyświetlona lista z podsumowaniem, które pliki nie mogły zostać dodane do zamówienia z powodu braku uprawnień.

W ramach zmian w RCBN w zakresie wprowadzenia nowego sposobu pobierania hurtowego plików z systemu – zamówienia na pliki – istniejący system uprawnień zostanie dostosowany zgodnie z poniżej umieszczonym opisem, specyfikującym uprawnienia mające wpływ na przeglądanie i pobieranie obiektów i ich zawartości.

Rodzaje uprawnień, które określają dostęp do obiektu (ŁCO):

- A. Ładowanie Cyfrowych Obiektów (ŁCO) – pozwala na dostęp do obiektu, zakres dostępu jest określony przez dodatkowy cel uprawnienia:
  - a. Metadane – nadaje dostęp jedynie do metadanych obiektu (możliwość wyszukania i przeglądania metadanych)
  - b. Contenty paginacyjne - daje dostęp do podstawowej zawartości jedynie contentów paginacyjnych
  - c. Contenty zbiorcze - daje dostęp do podstawowej zawartości jedynie contentów zbiorczych

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

d. \* - daje dostęp do podstawowej zawartości wszystkich contentów, bez względu na rodzaj

B. Ładowanie Kontekstów – pozwala na dostęp do wszystkich obiektów znajdujących się w danym kontekście równoważny z nadaniem każdemu obiektowi w tym kontekście uprawnienia Ad (czyli ŁCO:\*)

WAŻNE: Nadane przez uprawnienia ŁCO dostępy pozwalają zarówno na wyświetlenie tej zawartości jak i jej pobranie. Nie ma oddzielnych uprawnień definiujących pobieranie zawartości.

Powyższe uprawnienia są bazowymi uprawnieniami określającymi dostęp do obiektów. Działanie wszystkich pozostałych uprawnień jest od nich zależne – mówiąc inaczej użytkownik musi posiadać w stosunku do danego obiektu „ŁCO” aby zadziałały pozostałe uprawnienia dotyczące operacji na obiekcie.

Dodatkowo należy zaznaczyć iż uprawnienia te (ŁCO) dające dostęp do zawartości contentu mają zastosowanie tylko do stream’u „mainStream”, z wyłączeniem plików graficznych. Wyłączone są również contenty które są oznaczone jako ukryte.

Poprzez określenie „tylko do stream’u mainStream z wyłączeniem plików graficznych” – rozumie się, że użytkownik nie ma prawa pobrania/wyświetlenia pliku graficznego (mimetype: image/\*) ze stream’u mainStream, pomimo iż ma dostęp do zawartości tego streamu, zamiast tego ma dostęp do wersji pochodnej „miniaturki” oraz „kafelkowca” które są trzymane poza repozytorium plików i są postaciami prezentacyjnymi tego contentu – w takiej sytuacji wyświetlenie zawartości graficznej realizowane jest przez viewer seadragon i kafelki a nie poprzez wyświetlenie pliku źródłowego dostępnego w tym stream’ie. Aby się do niego dostać trzeba posiadać dodatkowe uprawnienia. Wyłączenie to stosuje się do wszystkich typów contentów, nie jest ważne czy to content z grupy paginacyjnej czy zbiorczej.

Poza opisanymi uprawnieniami ŁCO istnieją uprawnienia, które rozszerzają powyższe uprawnienia, czyli zadziałają pod warunkiem że są nadane wpierw odpowiednie uprawnienia z pozycji A lub B:

1. Dostęp do ukrytych contentów – rozszerza dostęp o contenty które są oznaczone jako ukryte
2. Dostęp do alternatywnych streamów – rozszerza dostęp o wszystkie rodzaje streamów, włącznie z dostępem do plików graficznych w stream’ie mainStream
3. Pobieranie pliku ALTO - rozszerza dostęp o stream „download\_alto”
4. Pobieranie pliku pochodnego fullJPG – rozszerza dostęp o stream „download\_fullJPG”
5. Pobieranie pliku oryginalnego – rozszerza dostęp o pełny stream „mainStream”, tzn. włącznie z plikami graficznymi się w nim znajdującymi
6. Dostęp do alternatywnych streamów text i alto – to uprawnienie pozwala na pobranie zawartości stream’u „download\_alto” oraz „text” nawet jeżeli użytkownik nie posiada

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

żadnego dostępu do contentów – wystarczy że posiada dostęp do obiektu na poziomie metadanych.

7. Rozszerzone ładowanie kontekstów – uprawnienie które można dodać tylko do kontekstu – daje ono użytkownikowi rozszerzony dostęp dla wszystkich obiektów znajdującym się w danym kontekście do wszystkich streamów alternatywnych i wszystkich streamów mainStream włącznie z graficznymi i wszystkich ukrytych contentów (czyli tak jakby dla wszystkich obiektów z kontekstu użytkownik miał uprawnienia: "Dostęp do alternatywnych streamów" + "Dostęp do ukrytych contentów"). Uprawnienie zadziała dla danego obiektu tylko jeżeli użytkownik ma prawo ŁCO dla danego obiektu (bezpośrednio lub przez kontekst).

Do realizowania zamówień służy uprawnienie „tworzenie i obsługa zamówień”.

### 7.3.4 Zmiany w uprawnieniach

Poniżej umieszczona została finalna lista uprawnień mających wpływ na przeglądanie i pobieranie obiektów i ich zawartości:

1. Ładowanie cyfrowych obiektów - zgodnie z opisem w rozdziale "Budowanie zamówienia a uprawnienia"
2. Ładowanie Kontekstów - zgodnie z opisem w rozdziale "Budowanie zamówienia a uprawnienia"
3. Pobieranie kontekstu – przemianować na „Rozszerzone ładowanie kontekstów”, działanie zgodnie z opisem w rozdziale "Budowanie zamówienia a uprawnienia"
4. Dostęp do alternatywnych streamów – zgodnie z opisem w rozdziale "Budowanie zamówienia a uprawnienia"
5. Pobieranie pliku ALTO – zgodnie z opisem w rozdziale "Budowanie zamówienia a uprawnienia", dodatkowo przemianować na "dostęp do streamu download\_alto"
6. Pobieranie pliku pochodnego fullJPG – zgodnie z opisem w rozdziale "Budowanie zamówienia a uprawnienia", dodatkowo przemianować na "dostęp do streamu download\_fullJPG"
7. Pobieranie pliku oryginalnego – zgodnie z opisem w rozdziale "Budowanie zamówienia a uprawnienia", dodatkowo przemianować na "pełny dostęp do streamu mainStream"
8. Dostęp do alternatywnych streamów text i alto – zgodnie z opisem w rozdziale "Budowanie zamówienia a uprawnienia"
9. Tworzenie i obsługa zamówień – uprawnienie będzie pozwalało tworzyć zamówienia, realizować je oraz pobierać zawartość przygotowaną dla zamówienia.
10. Uprawnienie, które nadane użytkownikowi pozwala w pełni zarządzać wszystkimi zamówieniami, nie tylko swoimi.

### 7.3.5 Zarządzanie zamówieniami

Funkcjonalność obsługi zamówień powinna działać w następujący sposób:

Projekt współfinansowany w ramach Programu Operacyjnego Polska Cyfrowa z Europejskiego Funduszu Rozwoju Regionalnego i budżetu państwa na podstawie Umowy o dofinansowanie nr POPC.02.01.00-00-0043/16-00



Fundusze  
Europejskie  
Polska Cyfrowa

**omnis**<sup>BN</sup>

Unia Europejska  
Europejski Fundusz  
Rozwoju Regionalnego



Strona 57 z 70

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

1. Użytkownik powinien móc przejść do widoku „Moje zamówienia”, który powinien zawierać listę jego zamówień wraz ze statusem. Powinny być następujące statusy zamówień:
  - a. Nowe – oznacza nowe zamówienie, użytkownik powinien mieć możliwość utworzenia dowolnej liczby nowych zamówień.
  - b. W trakcie realizacji – oznacza zamówienie wysłane do realizacji.
  - c. Zrealizowane – oznacza, że pliki w formie pliku TAR zostały przygotowane dla użytkownika.
  - d. Zarchiwizowane – oznacza, że plik TAR z zamówionymi plikami został usunięty.
2. Lista zamówień powinna nie wyświetlać domyślnie zamówień ze statusem Zarchiwizowane. Powinna być jednak możliwość wyświetlenia wszystkich zamówień tego użytkownika łącznie z zarchiwizowanymi.
3. Użytkownik powinien móc wykonać następujące akcje:
  - a. Powinien móc dodać zamówienie o statusie Nowe
  - b. Dla zamówienia o statusie Nowe:
    - i. Wysłać do realizacji
    - ii. Usunąć (np. w przypadku pomyłki)
  - c. Dla zamówienia o statusie Zrealizowane:
    - i. Pobrać plik zawierający zamówione pliki
    - ii. Zarchiwizować, czyli usunąć plik zawierający zamówione pliki, co powinno również skutkować zmianą statusu zamówienia na Zarchiwizowane

W przypadku usunięcia zamówienia zostanie ono trwale usunięte z listy zamówień. Funkcjonalność usunięcia zamówienia powinna być wykorzystywana tylko dla przypadku kiedy użytkownik nie chce/nie potrzebuje realizować tego zamówienia.

4. Użytkownik powinien móc wyświetlić szczegóły zamówienia obejmujące:
  - a. Nagłówek zamówienia – zawierające wspólne informacje w zamówieniu: numer zamówienia, nazwa zamówienia (opisowe pole tekstowe), status, datę utworzenia, datę wysłania do realizacji, datę realizacji, datę zarchiwizowania oraz flagę czy zamówienie posiada wysoki priorytet. Użytkownik powinien móc zmodyfikować nazwę zamówienia, pole to nie powinno być jednak obowiązkowe. Użytkownik powinien móc zmodyfikować flagę czy zamówienie posiada wysoki priorytet, o ile zamówienie posiada status Nowe.
  - b. Tabelkę z listą zamówionych plików – zawierające informacje o wybranych do zamówienia plikach, każdy wiersz zamówienia powinna zawierać: nazwę obiektu (tytuł i iteracja), nazwę pliku, rozszerzenie pliku, rozmiar pliku. Lista plików powinna zostać wyświetlona z zastosowaniem stronicowania.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Lista plików powinna zawierać następujące kolumny: id obiektu, id contentu, numer kolejny contentu (dla zbiorczych 0), nazwę klucza streamu, mimetype pliku z bazy z Repozytorium, w miarę możliwości nazwę oryginalną pliku, potem kolejność contentów w obiekcie, typ kontentu (postać zbiorcza, content podstawowy, content alternatywny).

Kolejność sortowania: id obiektu, postaci zbiorcze, contenty w strukturze paginacyjnej posortowane w wewnętrznej kolejności wyświetlania w systemie. Contenty alternatywne powinny być wyświetlone bezpośrednio pod kontentem pod który podlegają i powinny być posortowane zgodnie z kolejnością wyświetlania ich w Repozytorium.

5. Użytkownik powinien móc usunąć wybrane wiersze z zamówienia. Z uwagi na to, że zamówienie może potencjalnie dotyczyć wielu plików, użytkownik powinien móc przefiltrować wyświetlaną listę plików wybierając:
  - a. Typu kontentu z możliwością wyboru jednej lub wielu z tych opcji: zbiorcze, paginacyjne podstawowe, paginacyjne alternatywne.
  - b. Kluczy stream'a z możliwością wyboru jednej lub wielu z tych opcji: np. mainStream, download\_alto, text, download\_fullJPG, download\_lowJPG, RAW. W tym widoku (szczegóły zamówienia) lista kluczy stream'ów powinna być budowana dynamicznie z kluczy stream'ów występujących w tym zamówieniu.
  - c. Wartości z pola typ pliku/mimeType z możliwością wyboru jednej lub wielu z tych opcji. W tym widoku (szczegóły zamówienia) lista typów plików/mimeType powinna być budowana dynamicznie z plików występujących w tym zamówieniu.

Pomiędzy poszczególnymi filtrami (powyższe punkty a, b i c) powinien zostać wykorzystany operator logiczny AND, a pomiędzy wartościami w poszczególnych filtrach operator logiczny OR.


### 7.3.6 Limit miejsca na zamówienia użytkownika

Przed wysłaniem zamówienia system powinien sprawdzić, czy sumaryczny rozmiar produktów (plików źródłowych) ze zrealizowanych a niezarchiwizowanych zamówień użytkownika nie przekracza skonfigurowanego limitu:

1. Jeżeli przekracza, to użytkownik powinien otrzymać komunikat "Twoja przestrzeń robocza jest zajęta, pobierz pliki i zarchiwizuj pozostałe zamówienia, aby wysłać nowe zamówienie", a zamówienie nie powinno zostać wysłane do realizacji.
2. Jeżeli nie przekracza, to zamówienie powinno zostać wysłane do realizacji.

Rozmiar zamówień powinien być obliczony w analogiczny sposób jak podczas weryfikacji rozmiaru zamówienia podczas dodawania obiektów do zamówienia – opis tego mechanizmu znajduje się w rozdziale **Budowanie zamówienia z poziomu szczegółów obiektu** w punkcie 6.



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

### 7.3.7 Struktura folderów i nazewnictwo plików

Struktura folderów w paczce z plikami z zamówienia:

1. Folder z numerem zamówienia
2. Podfoldery obiektów, których nazwa powinna składać się z identyfikatora obiektu + początek nazwy obiektu – do 60 znaków. Znaki zastrzeżone w nazwach plików w systemie Windows powinny zostać zastąpione znakiem „\_”.
3. Pliki z obiektów

Nazewnictwo plików stream’ów zostało przedstawione poniżej:

<identyfikator obiektu z Repozytorium>separator<kolejność w obiekcie>separator<nazwa typu streamu>separator<Identyfikator kontentu z repozytorium>separator<identyfikator streamu>.rozszerzenie pliku

Pod pojęciem kolejności w obiekcie należy rozumieć systemową (w rozumieniu wewnętrznych identyfikatorów kolejności) w Repozytorium kolejność kontentów w strukturze paginacyjnej.

Numer będący kolejnością w obiekcie powinien być uzupełniany zerami do 4 cyfr, np. 0001, 0002, ..., 0010, 0011... .

W przypadku kontentów postaci zbiorczych kolejność należy rozumieć jako systemową (w rozumieniu wewnętrznych identyfikatorów kolejności) kolejność w Repozytorium – o ile taka jest dostępna w Repozytorium. Jeżeli taka kolejność w Repozytorium nie jest dostępna, w nazwach stream’ów postaci zbiorcze zamiast kolejności powinna zostać użyta na stałe liczba zero („0”).

Separator zostanie wybrany przez Wykonawcę na etapie implementacji. Może to być przykładowo znak podkreślenia dolnego „\_”.

### 7.3.8 Realizacja zamówienia

Po wysłaniu przez użytkownika zamówienia do realizacji, system powinien sprawdzić czy wszystkie zamówione pliki są dostępne, tj. czy nie są usunięte z uwagi na proces archiwizacji.

System powinien zablokować (LOCK) obiekty, których pliki są w zamówieniu. Blokada powinna zostać zdjęta w momencie zakończenia zrealizowania zamówienia – przygotowania paczki z plikami z zamówienia.

Jeżeli część plików nie jest dostępnych, system powinien uruchomić proces przywracania obiektów wybranych do realizowanego zamówienia. Realizacja zamówienia powinna zostać wznowiona po zakończeniu przywracania brakujących obiektów.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

W momencie, kiedy wszystkie zamówione pliki będą już dostępne, system za pomocą narzędzia 7zip powinien stworzyć paczkę ZIP – produkt zamówienia, który powinien zawierać wszystkie zamówione pliki w danym zamówieniu. Rozważana jest zmiana zakresu opisanego w niniejszym dokumencie polegająca na tym, że w przypadku kiedy rozmiar zamówienia przekracza określony w konfiguracji limit, system powinien podzielić pliki na kilka plików ZIP. Dzięki temu użytkownik będzie miał do pobrania kilka plików składających się na zamówienie, np. jeżeli całe zamówienie jest oszacowane na 100GB to plik archiwum powinien zostać podzielony na kilka części np. po 10GB – tak aby w razie zerwania połączenia podczas pobierania jednego pliku, użytkownik musiał pobrać jeszcze raz tylko ten plik.

Po stworzeniu paczki system powinien zmienić status zamówienia na „Zrealizowane” oraz powinien wysłać do użytkownika powiadomienie o możliwości pobrania produktu zamówienia wraz z informacją, do kiedy trzeba pobrać plik z zamówienia, ponieważ po tym czasie paczka z plikami z zamówienia zostanie automatycznie usunięta. Data do kiedy należy pobrać paczkę powinna zostać obliczana na podstawie skonfigurowanego okresu karencji na pobranie paczki z zamówienia od momentu realizacji zamówienia.

Użytkownik powinien móc pobrać produkt zamówienia zarówno z poziomu listy zamówień jak i po wyświetleniu szczegółów zamówienia.

Pobranie pliku powinno być zabezpieczone w taki sposób, aby inny użytkownik nie był w stanie odgadnąć ścieżki URL, pod którą znajduje się plik z zamówieniem innego użytkownika. Przykładem zabezpieczenia jest umieszczenie w ścieżce URL sekcji z losowym GUID.

### 7.3.9 Archiwizacja zamówienia

Przez archiwizację rozumie się usunięcie produktu zamówienia oraz zmianę statusu zamówienia na „Zarchiwizowane”. Jej celem jest zwolnienie miejsca zajmowanego przez potencjalnie duże paczki z zamówionymi plikami.

Powinny istnieć następujące ścieżki archiwizacji zamówienia:


#### 1. Automatyczna archiwizacja po zdefiniowanym czasie

Zamówienie powinno zostać automatycznie zarchiwizowane po upływie skonfigurowanego okresu karencji od na pobranie paczki z zamówienia.

#### 2. Ręczna archiwizacja przez użytkownika

Użytkownik powinien móc wywołać ręcznie akcję zarchiwizowania zamówienia (opisano w sekcji [Zarządzanie zamówieniami](#)).



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

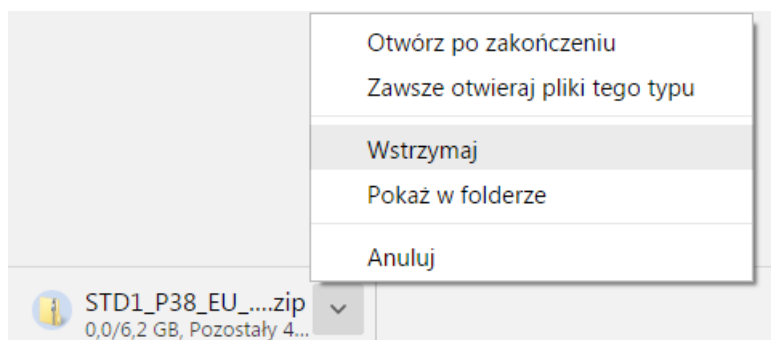
### 7.3.10 Pobieranie dużych plików

Aplet java do pobierania wielu plików był do tej pory wykorzystywany jedynie w widoku pobierania całej publikacji. Widok ten jednak po wprowadzeniu Systemu archiwizacji nie będzie już dostępny/będzie następowało przekierowanie do funkcjonalności zamówień na pliki.

Z drugiej strony zamówione przez użytkownika pliki będą udostępnione użytkownikowi w formie pliku/plików archiwalnych ZIP zawierającego zamówione pliki. Dlatego użytkownik pobierając zamówienie będzie pobierał jeden plik, a nie wiele.

Przeanalizowano możliwe rozwiązania sposobu pobrania potencjalnie dużego pliku z zamówienia i zastosowanie komponentu zapewniającego funkcjonalność tzw. download managera spowodowała by dodatkowe obciążenie serwera. Z drugiej strony najbardziej popularne aktualnie przeglądarki internetowe (np. Chrome) zapewniają funkcjonalność wstrzymania oraz wznowienia pobierania plików. Dlatego proponuje się zastosowanie natywnych funkcjonalności przeglądarek internetowych.

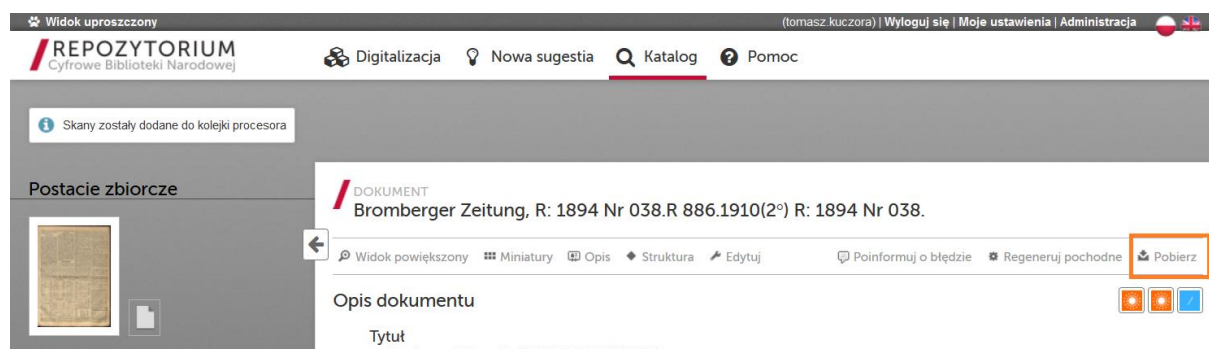
Prezentacja dostępnej opcji wstrzymaj pobierania pliku w przeglądarce Chrome:



### 7.4 Pobieranie plików z obiektu

Funkcjonalność pobierania plików umożliwia pobieranie wybranych elementów z obiektu.

Użytkownik może uruchomić funkcjonalność pobrania plików w widoku podglądu obiektu, używając w menu przycisku „Pobierz”, jak na poniższym przykładzie.



Projekt współfinansowany w ramach Programu Operacyjnego Polska Cyfrowa z Europejskiego Funduszu Rozwoju Regionalnego i budżetu państwa na podstawie Umowy o dofinansowanie nr POPC.02.01.00-00-0043/16-00



Fundusze  
Europejskie  
Polska Cyfrowa



Unia Europejska  
Europejski Fundusz  
Rozwoju Regionalnego



Strona 62 z 70

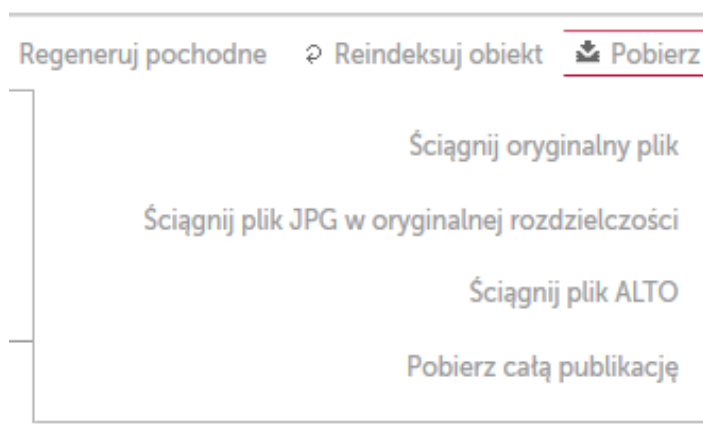
WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Użycie przycisku „Pobierz” rozwija dodatkowe menu pozwalające na uruchomienie pobrania wybranych elementów publikacji. Elementy w tym menu są dynamiczne i zależą od uprawnień użytkownika oraz dostępności zawartości:

1. "Ściągnij oryginalny plik" - pobiera plik źródłowy tego kontentu, pobierany jest stream o kluczu "mainStream"
2. "Pobierz całą publikację" – przekierowuje użytkownika do widoku z możliwością utworzenia zamówienia na pliki
3. "Ściągnij plik JPG w oryginalnej rozdzielczości" - pobranie pliku JPG (poprzez klucz "download\_fullJPG")
4. "Ściągnij plik ALTO" - pobranie pliku ALTO wyświetlanego kontentu (poprzez klucz "download\_alto")

W przypadku kiedy użytkownik nie znajduje się w widoku podglądu pojedynczego kontentu, ale przykładowo w widoku Opis, o ile użytkownik posiada stosowne uprawnienie w rozwijanym menu funkcjonalności „Pobierz” dostępna jest tylko jedna opcja o nazwie „Pobierz całą publikację”.

Przykład rozwiniętego menu na widoku podglądu pojedynczego kontentu został zaprezentowany poniżej.



Dostępność poszczególnych pozycji w tym menu jest zależna od uprawnień danego użytkownika.

Po wybraniu jednego z elementów (oprócz „Pobierz całą publikację”) do pobrania powinno nastąpić sprawdzenie, czy wymagany plik/pliki nie są usunięte z uwagi na archiwizację. Jeżeli plik/pliki są usunięte, to:

1. Jeżeli użytkownik posiada uprawnienie do zamawiania plików – to powinien zostać poinformowany o tym, że musi zamówić pliki używając funkcjonalności „Zamówienia na pliki” (wraz z zapewnieniem użytkownikowi możliwości przejścia do tej funkcjonalności).

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

2. Jeżeli użytkownik nie posiada uprawnień do zamawiania plików – to powinien zostać poinformowany o tym, że dany plik jest niedostępny.

W przypadku wybrania elementu „Pobierz całą publikację” użytkownik powinien zostać przekierowany do widoku budowania zamówienia z poziomu szczegółów obiektu (opisany w rozdziale [Zamówienia na pliki](#)).

## 7.5 Ponowna regeneracja pochodnych

Regeneracja pochodnych uruchamia proces ponownego wygenerowania formatów pochodnych dla skanów w strukturze paginacyjnej danego obiektu.

Użytkownik może uruchomić funkcjonalność regeneracji pochodnych z poziomu widoku podglądu projektu, używając w menu przycisku „Regeneruj pochodne” jak na poniższym przykładzie.



W przypadku użycia przycisku „Regeneruj pochodne” powinno nastąpić sprawdzenie, czy wymagane pliki są usunięte z uwagi na archiwizację:

1. Jeżeli nie są usunięte, to procesowanie powinno przebiegać jak dotychczas.
2. Jeżeli są usunięte, to powinno zostać wysłane żądanie przywrócenia, a po zakończeniu przywrócenia powinno zostać wykonane zadanie ponownej regeneracji pochodnych danego obiektu. Dodatkowo użytkownik powinien zostać powiadomiony o tym, że proces regeneracji pochodnych będzie trwał dłużej niż zwykle z uwagi na potrzebę przywrócenia zawartości obiektu.

Założono, że wykonanie regeneracji pochodnych może być modyfikacją w kontekście archiwizacji i po upływie czasu karencji może skutkować wykonaniem ponownej archiwizacji. Argumentuje się to tym, że proces obrócenia może posiadać inne parametry (np. może być dokładniejszy), dlatego zasadne wydaje się w takiej sytuacji ponowne zarchiwizowanie obiektu. To, czy regeneracja pochodnych będzie modyfikacją w kontekście archiwizacji, powinno zależeć o skonfigurowanej liście archiwizowanych kluczy stream'ów (konfiguracja wykorzystywana podczas tworzenia paczki archiwalnej z obiektem). Lista ta jest wykorzystywana w celu oceny, czy modyfikacja obiektu powoduje zmodyfikowanie obiektu w kontekście archiwizacji. Mechanizm ten został opisany w rozdziale dotyczącym zmian w encji obiektu.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

## 7.6 Obracanie skanów

Użytkownik może uruchomić funkcjonalność obrócenia skanu z następujących poziomów:

1. Widoku „Paginacja” w edycji obiektu, w zakładce „Obracanie obrazków”: możliwość zbiorczego zastosowania funkcji obrotu dla wybranych kontentów.



2. Widoku „Paginacja” w kontrolce wyświetlania miniaturki i edycji kontentu: obrót w prawo i w lewo



Wykonanie akcji "Obróć obrazki" powoduje wizualny obrót miniaturek, ale ta zmiana zapisze się dopiero po użyciu przycisku "Zapisz wszystko".

Zapisanie zmiany dotyczącej obrotu poprzez „Zapisz wszystko” powoduje wysłanie zadania do procesorów graficznych obrotu wszystkich formatów pochodnych danego kontentu, za wyjątkiem pliku źródłowego (mainStream).

W przypadku użycia funkcjonalności obrócenia (wraz z zapisaniem zmian), powinno nastąpić sprawdzenie czy pliki wymagane do przeprowadzenia procesu obrócenia skanów są usunięte z uwagi na archiwizację:

1. Jeżeli nie są usunięte, to obrócenie skanów powinno przebiegać jak dotychczas.
2. Jeżeli są usunięte, to powinno zostać wysłane żądanie przywrócenia obiektu, a po zakończeniu przywrócenia powinno zostać wykonane zadanie obrócenia skanów.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

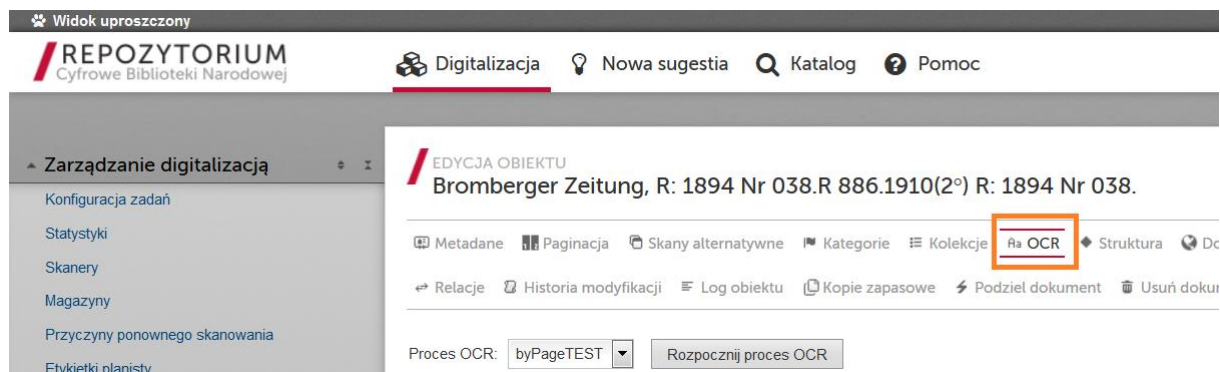
Dodatkowo użytkownik powinien zostać powiadomiony o tym, że proces będzie trwał dłużej niż zwykle z uwagi na potrzebę przywrócenia zawartości obiektu.

Założono, że wykonanie obrócenia skanów może być modyfikacją w kontekście archiwizacji i po upływie czasu karencji może skutkować wykonaniem ponownej archiwizacji. Argumentuje się to tym, że proces obrócenia może przykładowo poprawiać orientację skanu (np. pion/poziom), dlatego zasadne wydaje się w takiej sytuacji ponowne zarchiwizowanie obiektu. To, czy obrócenie skanów będzie modyfikacją w kontekście archiwizacji, powinno zależeć o skonfigurowanej liście archiwizowanych kluczy stream'ów (konfiguracja wykorzystywana podczas tworzenia paczki archiwalnej z obiektem). Lista ta jest wykorzystywana w celu oceny, czy modyfikacja obiektu powoduje zmodyfikowanie obiektu w kontekście archiwizacji. Mechanizm ten został opisany w rozdziale dotyczącym zmian w encji obiektu.

Na czas wykonania operacji obracania skanów system powinien zablokować (LOCK) ten obiekt dla tego procesu. Blokada powinna zostać zdjęta w momencie zakończenia realizowania tej operacji.


## 7.7 Wysyłanie obiektu do procesu OCR

Po przejściu obiektu przez proces workflow tylko użytkownik może uruchomić funkcjonalność przetworzenia obiektu przez OCR, i jest to możliwe z poziomu widoku edycji projektu, używając w menu przycisku „OCR” jak na poniższym przykładzie.



Po użyciu przycisku „Rozpocznij proces OCR” pojawia się potwierdzenie, że zadanie przetworzenia obiektu zostało wysłane do OCR, jak na poniższym przykładzie:



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

Widok uproszczony

**REPOZYTORIUM**  
Cyfrowe Biblioteki Narodowej

Digitalizacja Nowa sugestia Katalog Pomoc

Wysłano dokument do OCR Dokument jest już w trakcie procesu OCR

**Zarządzanie digitalizacją**

- Konfiguracja zadań
- Statystyki
- Skanery
- Magazyny
- Przyczyny ponownego skanowania
- Etykiety planisty
- Ustawienia pola wyszukiwania (duplikaty obiektu)
- Konfiguracja statusów
- Stan kolejek komunikatów
- Zadania grupowej edycji obiektów

**EDYCJA OBIEKTU**  
Bromberger Zeitung, R: 1894 Nr 038.R 886.1910(2°) R: 1894 Nr 039

Metadane | Pagine | Skany alternatywne | Kategorie | Kolekcje | **OCR**

Relacje | Historia modyfikacji | Log obiektu | Kopie zapasowe | Podziel dokument

Proces OCR: byPageTEST Rozpocznij proces OCR

Zmień status ostatniego (aktualnego) procesu OCR na FAIL Zmień status

**Historia wysyłania dokumentu do OCR**

L.p.	W ramach workflow?	Status	Data wysłania do OCR	Data powrotu z OCR
1	Nie	IN_PROGRESS	13-09-2016 11:48:45	


Po użyciu przycisku „Rozpocznij proces OCR” powinno nastąpić sprawdzenie, czy pliki wymagane do przeprowadzenia procesu OCR są usunięte z uwagi na archiwizację:

1. Jeżeli nie są usunięte, to procesowanie powinno przebiegać jak dotychczas.
2. Jeżeli są usunięte, to powinno zostać wysłane żądanie przywrócenia obiektu, a po zakończeniu przywrócenia powinno zostać wykonane zadanie ponownej regeneracji pochodnych danego obiektu. Dodatkowo użytkownik powinien zostać powiadomiony o tym, że proces regeneracji pochodnych będzie trwał dłużej niż zwykle z uwagi na potrzebę przywrócenia zawartości obiektu.

Założono, że wykonanie ponownego procesu OCR-owania obiektu może być modyfikacją w kontekście archiwizacji i po upływie czasu karencji może skutkować wykonaniem ponownej archiwizacji. Argumentuje się to tym, że nowy proces OCR-owania może posiadać inne parametry (np. może być dokładniejszy), dlatego zasadne wydaje się w takiej sytuacji ponowne zarchiwizowanie obiektu. To, czy przetworzenie obiektu przez proces OCR będzie modyfikacją w kontekście archiwizacji, powinno zależeć o skonfigurowanej listy archiwizowanych kluczy stream’ów (konfiguracja wykorzystywana podczas tworzenia paczki archiwalnej z obiektem). Lista ta jest wykorzystywana w celu oceny, czy modyfikacja obiektu powoduje zmodyfikowanie obiektu w kontekście archiwizacji. Mechanizm ten został opisany w rozdziale dotyczącym zmian w encji obiektu.

Na czas wykonania operacji przetwarzania OCR system powinien zablokować (LOCK) ten obiekt dla tego procesu. Blokada powinna zostać zdjęta w momencie zakończenia realizowania tej operacji.



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

## 7.8 Zmiana interfejsu API

Udostępniony przez Repozytorium interfejs API służy aktualnie przykładowo systemowi Polona do pobierania elementów obiektu.

W istniejącej obsłudze zapytań otrzymanych przez webservice powinno zostać dodane sprawdzenie czy plik nie jest usunięty z uwagi na archiwizację.

Jeżeli obiekt, do którego należy weryfikowany plik, jest zarchiwizowany oraz ponadto weryfikowany plik jest usunięty, to w takim przypadku powinno nastąpić zwrócenie nowego kodu błędu oznaczającego brak pliku z uwagi na proces archiwizacji.

Rozbudowa interfejsu Repozytorium o możliwość wywołania przywrócenia obiektu. Możliwość wywołania przywrócenia powinno być ograniczone nowym uprawnieniem oraz dodatkowo powinien być sprawdzany skonfigurowany dzienny limit wywołanych przywróceń, aby systemy zewnętrzne nie były w stanie wygenerować nieograniczonej liczby żądań przywrócenia.

W interfejsie graficznym systemu Repozytorium oraz Academica należy w „Widoku powiększonym” w momencie wyświetlenia kontentu sprawdzić czy kontent ma format graficzny:

1. Jeżeli tak, to wyświetlany jest tzw. „kafelkowiec” i w tym zakresie nie jest potrzebna żadna zmiana.
2. Jeżeli nie, to powinno zostać dodane sprawdzenie czy plik w mainStream nie jest usunięty z uwagi na archiwizację, i jeżeli jest usunięty to zamiast podglądu powinna zostać wyświetlona informacja lub ikonka, że plik niedostępny.

## 7.9 Przycisk do przywracania obiektu

Użytkownik powinien mieć możliwość użycia przycisku, który spowoduje uruchomienie procesu przywrócenia danego obiektu. Przycisk ten powinien być widoczny tylko jeżeli obiekt posiada status REDUCED oraz jeżeli użytkownik posiada uprawnienie edycji tego obiektu oraz dodatkowo ma nadane nowe uprawnienie umożliwiające przywracanie obiektów.

## 7.10 Funkcjonalność współdzielenia i wydzielania contentów

W systemie Repozytorium istnieje funkcja, która pozwala współdzielić content pomiędzy inne obiekty albo wręcz przenosić content do innego obiektu. Aktualnie nie ma ograniczeń co do możliwości współdzielenia i wydzielania contentów pomiędzy obiektami.

Możliwości przeniesienia/współdzielenia kontentu powinna zostać zablokowana, jeżeli stream’y w tym kontencie są usunięte z uwagi na archiwizację. Operacja przeniesienia/współdzielenia zarchiwizowanego kontentu z usuniętą częścią streamów powinna poprzedzona realizowaniem zamówienia na ten obiekt. To rozwiązanie wydaje się odpowiednie, ponieważ użytkownik, który będzie chciał ustawić współdzielenie/przeniesienie

Projekt współfinansowany w ramach Programu Operacyjnego Polska Cyfrowa z Europejskiego Funduszu Rozwoju Regionalnego i budżetu państwa na podstawie Umowy o dofinansowanie nr POPC.02.01.00-00-0043/16-00



Fundusze  
Europejskie  
Polska Cyfrowa



Unia Europejska  
Europejski Fundusz  
Rozwoju Regionalnego



WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

kontentu najpierw będzie chciał się upewnić, czy na pewno próbuje współdzielić/przenieść odpowiedni контент. A aby to sprawdzić wyśle zapewne zamówienie na najważniejsze pliki tego kontentu w celu ich weryfikacji merytorycznej.

Jeżeli użytkownik chciałby współdzielić/przenieść content, który jest zarchiwizowany ale nie jest usunięty, działanie systemu powinno być następujące:

1. Ustawić współdzielenie tego kontentu z obiektem, który jest zarchiwizowany: w drugim obiekcie należy zaktualizować datę ostatniej aktualizacji w kontekście archiwizacji i tym samym pobudzić go do ponownej archiwizacji
2. Ustawić przeniesienie tego kontentu do obiektu zarchiwizowanego, który jest zarchiwizowany: należy zmienić status wszystkich streamów na „NEW” i tym samym pobudzić go do ponownej archiwizacji
3. Ustawić współdzielenie tego kontentu z obiektem, który nie jest zarchiwizowany: nie trzeba podejmować kolejnych kroków
4. Ustawić przeniesienie tego kontentu do obiektu, który nie jest zarchiwizowany: zmienić status wszystkich streamów na „NEW” i nie trzeba podejmować dalszych kroków.

Dodatkowo powinny zostać wykonane opisane poniżej zmiany w mechanizmie wydzielania/współdzielenia obejmujące również zmiany udostępniającego tę funkcjonalność widoku w systemie Repozytorium. Repozytorium powinno wykrywać, które contenty są zablokowane do przenoszenia/współdzielenia – powinny one być oznaczone, aby użytkownik widział, że nie mogą być one przenoszone/współdzielone (np. duży czerwony X na miniaturce/ikonce) i nie powinno być możliwości zaznaczenia tych contentów. W przypadku nieprawidłowego zadziałania blokady możliwości przeniesienia/współdzielenia contentów na widoku, mechanizm przenoszenia/współdzielenia nie powinien umożliwić (wraz z wyświetleniem komunikatu dla użytkownika) przeniesienia/współdzielenia contentów, jeżeli wybrane zostały contenty zawierające usunięte streamy.

## 8 Monitorowanie usług Systemu archiwizacji

Treść rozdziału nie załączona ze względu na brak związku z przedmiotowym zamówieniem.

## 9 Procedura wyłączenia i włączenia Systemu archiwizacji

Treść rozdziału nie załączona ze względu na brak związku z przedmiotowym zamówieniem.

## 10 Schemat i opis struktury bazy danych

Treść rozdziału nie załączona ze względu na brak związku z przedmiotowym zamówieniem.

WARSZAWA 2016	Biblioteka Narodowa	 BIBLIOTEKA NARODOWA
	System archiwizacji – projekt techniczny	

## 11 Założenia

Treść rozdziału nie załączona ze względu na brak związku z przedmiotowym zamówieniem.

## 12 Produkty końcowe projektu

Treść rozdziału nie załączona ze względu na brak związku z przedmiotowym zamówieniem.