



DInGO: „Digitise and Go!”

Pakiet oprogramowania do kompleksowej obsługi projektów digitalizacyjnych

Tomasz Parkoła

Plan prezentacji

Wprowadzenie

- Pakiet DIInGO „Digitise and Go!”



dLibra od 2011 do teraz

- Zmiany w wersji 5
- Nowości w wersji 6



dMuseion dla muzeów i galerii cyfrowych

- Funkcje i najważniejsze cechy
- Przykład wdrożenia: Muzeum Narodowe w Warszawie
- Przykład wdrożenia: Muzeum Historii Żydów Polskich



dLab do systematyzacji i automatyzacji digitalizacji

- Funkcje i zastosowanie
- Przykład wdrożenia: Ośrodek Brama Grodzka – Teatr NN

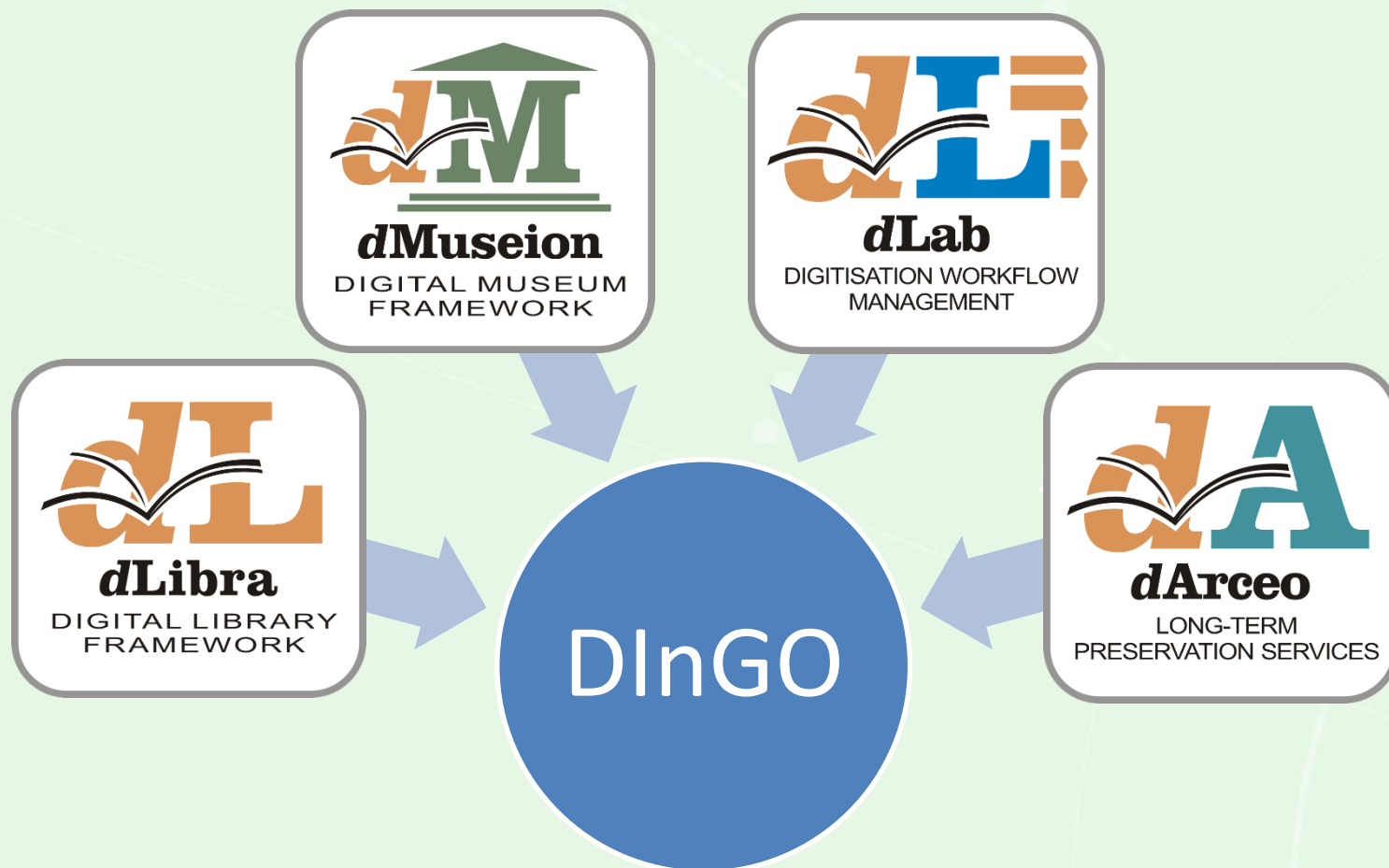


dArceo

- Funkcje i model OAIS
- Zgodność z wytycznymi Biblioteki Narodowej w zakresie programu WPR Kultura+

Pakiet oprogramowania DInGO

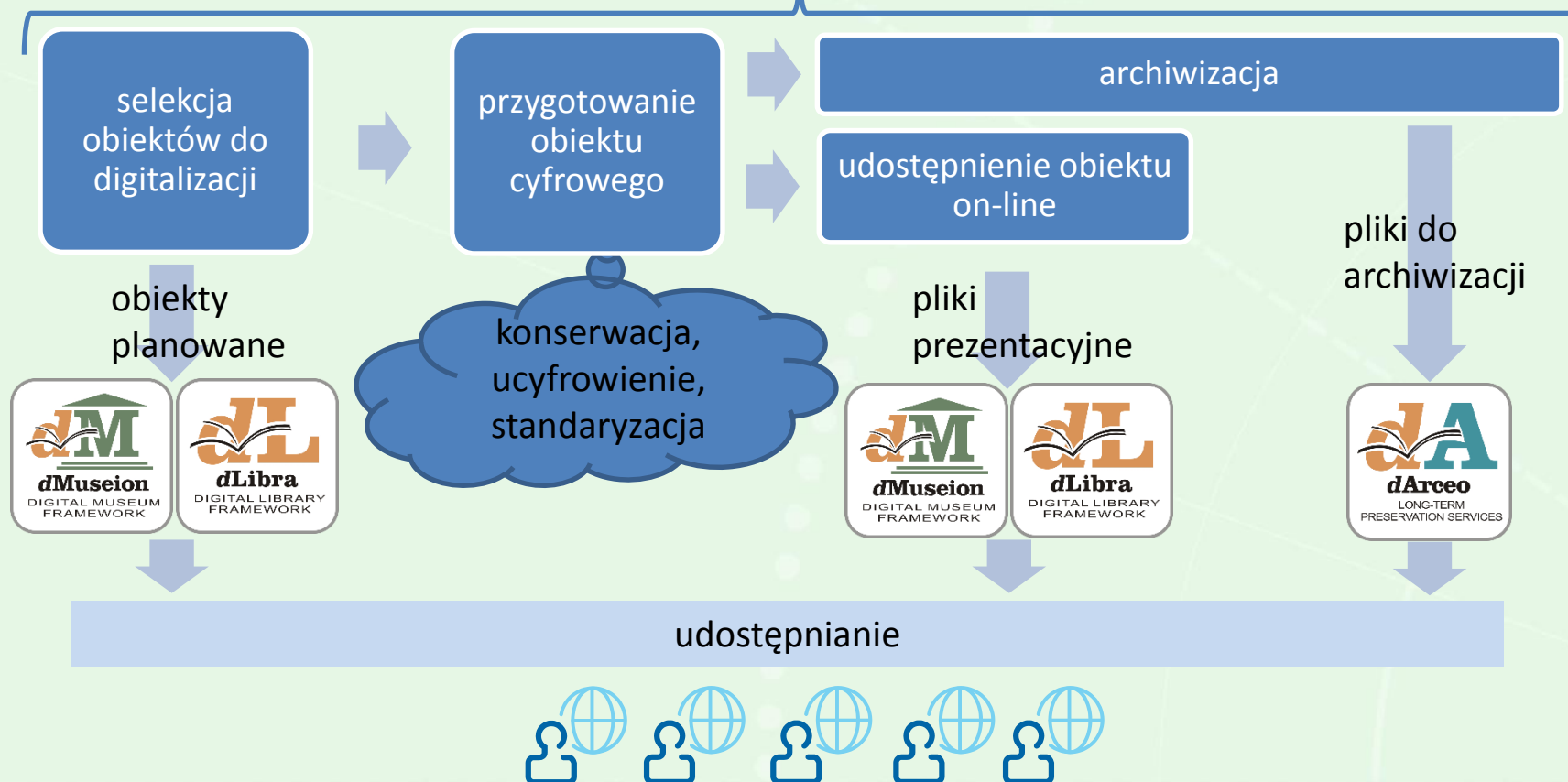
Digitise and Go!



DInGO: proces digitalizacji



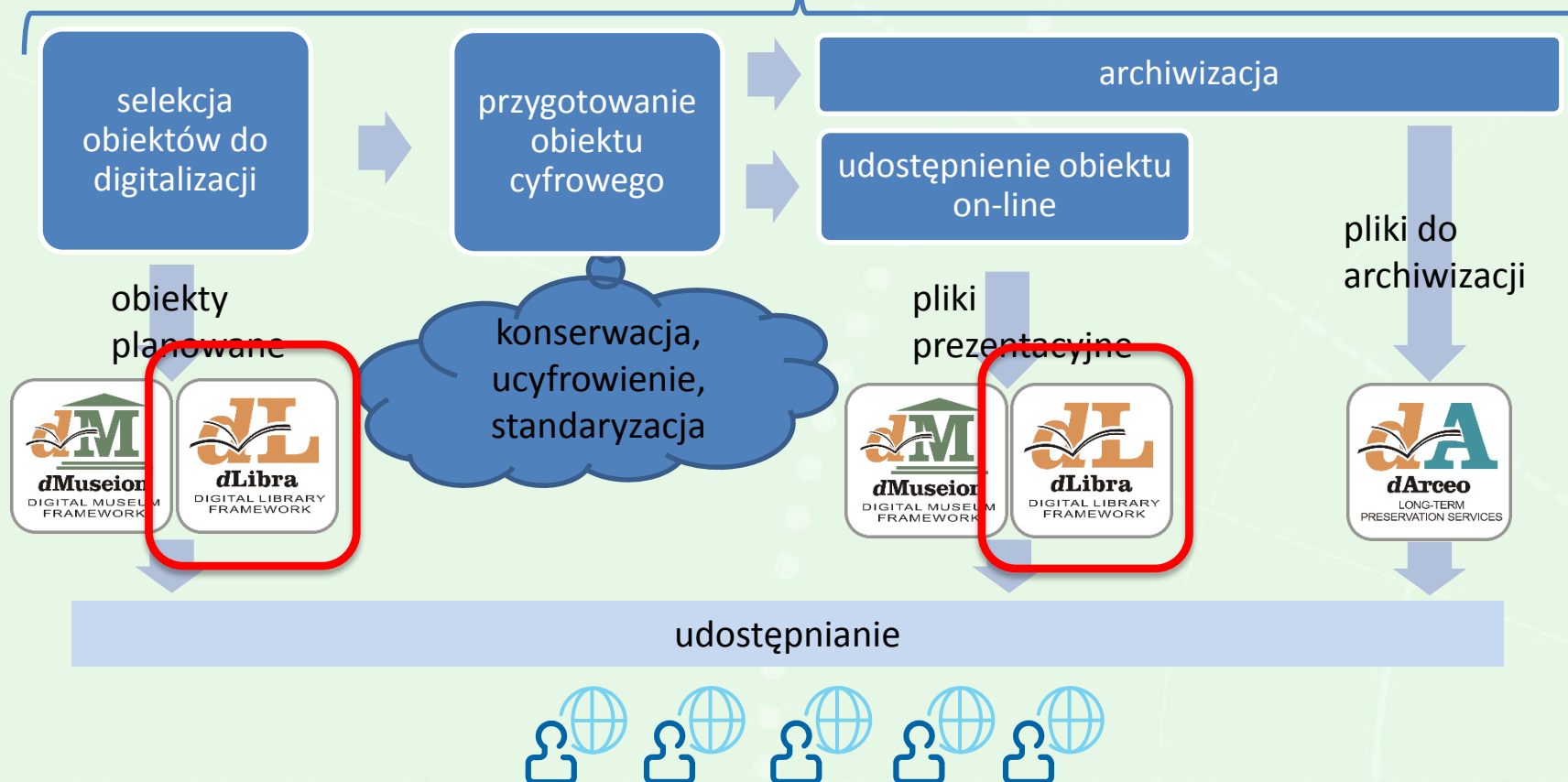
DInGO: proces digitalizacji



DInGO: dLibra



DInGO: proces digitalizacji (dLibra)





DInGO: dLibra

Informacje podstawowe

- Rozwijany od 1999 roku
- Pierwsze polskiej produkcji oprogramowanie dla bibliotek i repozytoriów cyfrowych
- Kluczowy element w stymulowaniu rozwoju polskich bibliotek cyfrowych
- Obecnie ponad 100 wdrożeń
- Ponad 2 miliony obiektów udostępnianych w całej Polsce





DInGO: dLibra

Wersja 6.0.0

Nowe
funkcje w
wersji 5

Nowe
funkcje tylko
w wersji 6

dLibra 6
(2011-2015)



DInGO: dLibra

Wersja 6.0.0

Nowe
funkcje w
wersji 5

Nowe
funkcje tylko
w wersji 6

- 8 dużych wersji 5.x
- 9 wersji z poprawkami

dLibra 6
(2011-2015)

- Ułatwienie dostępu do zasobów



DInGO: dLibra

Zmiany w wersji 5

W sumie



- ☐ 117 poprawek błędów
- ☐ 45 ulepszeń
- ☐ 15 nowych funkcji

Ulepszenia



- ☐ Przyspieszenie działania całego systemu (ARIa, WWW, serwer)
- ☐ Konfigurowalny komponent z kolekcjami
- ☐ Usprawnienia w mechanizmie self-archiving
- ☐ Usprawnienia w interfejsie OAI-PMH
- ☐ Zmiany pod kątem lepszego indeksowania przez wyszukiwarki (np. metatagi PDF)
- ☐ Ulepszenia mechanizmu importu słowników
- ☐ Przeniesienie statystyk do serwera w celu zwiększenia niezawodności
- ☐ Optymalizacja indeksowania plików DjVu (znaczne przyspieszenie)

Nowe funkcje

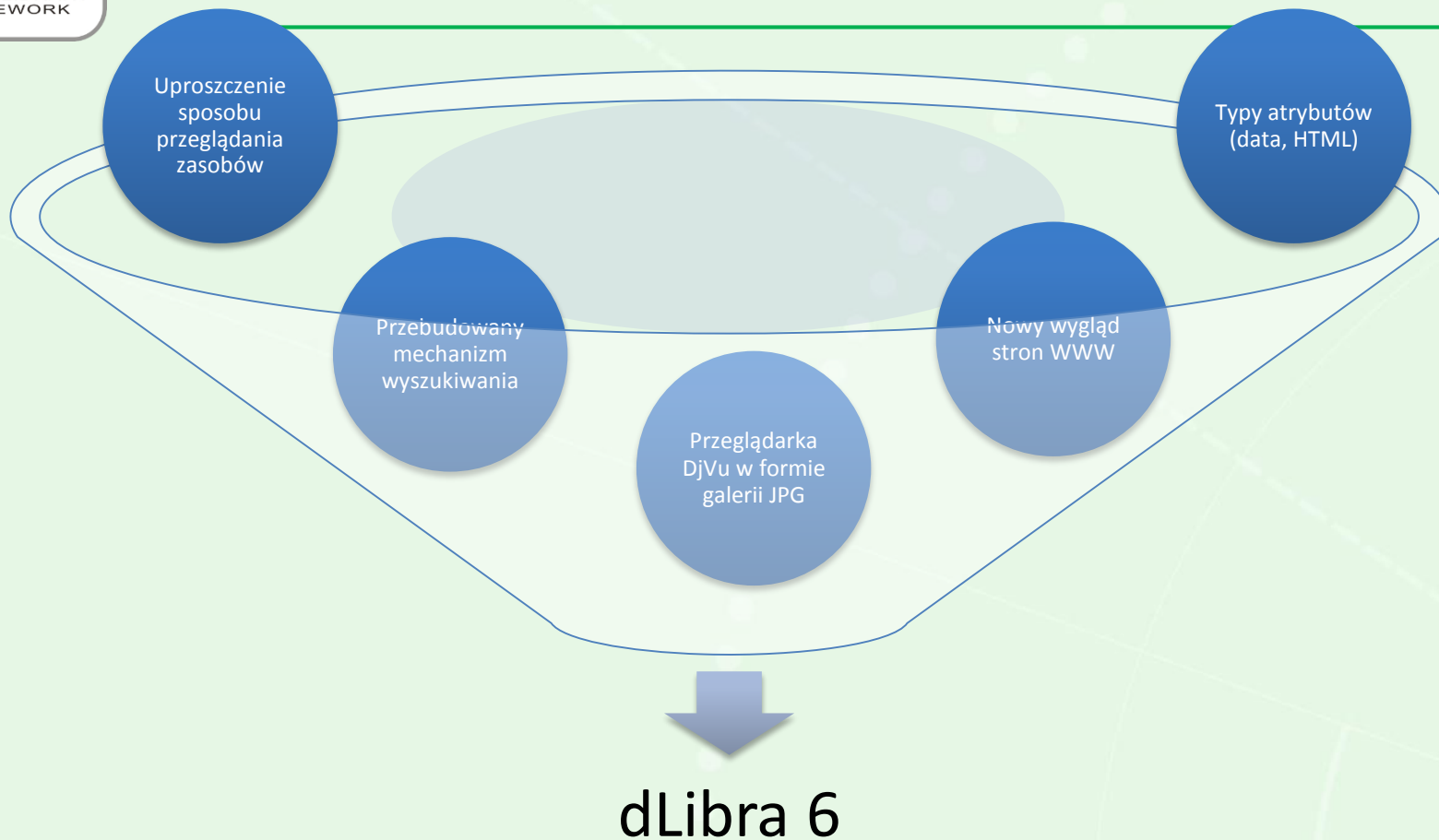


- ☐ Narzędzie do monitorowania pracy serwera w Aplikacji Redaktora i Administratora
- ☐ Możliwość uruchamiania kilku portali WWW na jednej instalacji
- ☐ Możliwość tworzenia publikacji grupowej z publikacji planowanej
- ☐ Masowe ładowanie publikacji planowanych
- ☐ Integracja z systemem dLab (publikacje planowane, self-archiving)
- ☐ Publikacje linkujące (link zamiast treści publikacji)
- ☐ Wprowadzenie interfejsu OAI-ORE



DInGO: dLibra

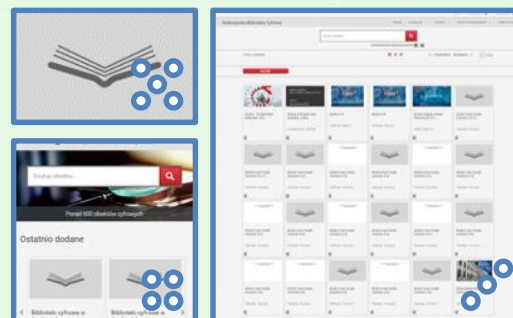
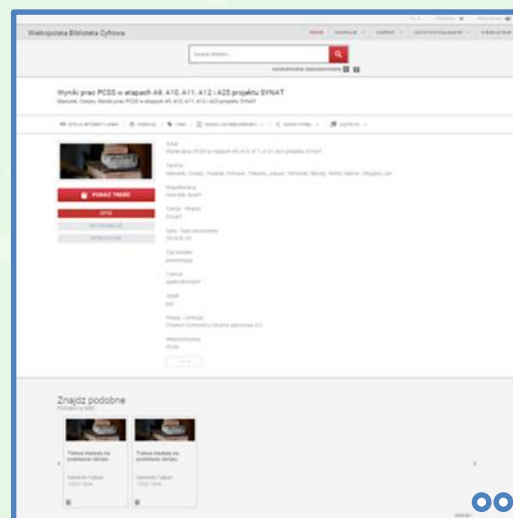
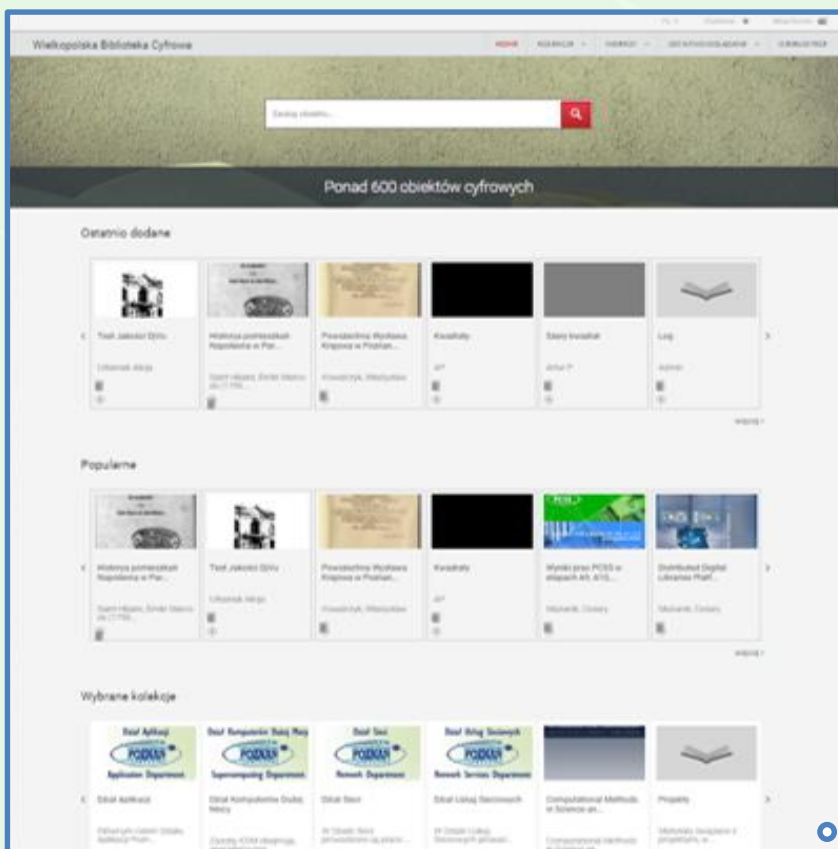
Nowe funkcje w wersji 6





DInGO: dLibra

Nowe funkcje w wersji 6: wygląd stron WWW



Strona
główna

Strona
obiektu

Wyniki
wyszukiwania

Wersja
mobilna

Miniaturki
obiektów i kolekcji



DInGO: dLibra

Nowe funkcje w wersji 6: przebudowany mechanizm indeksowania i wyszukiwania

Ulepszony mechanizm podpowiadania

- Wbudowany mechanizm technologii wykorzystywanej do wyszukiwania

Rozmyte wyszukiwanie synonimów

- Dostateczne jest wykorzystanie słowa lub frazy z grupy synonimów, nie ma konieczności definiowania pełnej frazy
- Mechanizm importu słownika z pliku csv

Konfiguracja rankowania wyników wyszukiwania

- Możliwość zmiany konfiguracji wag dot. wyszukiwanych elementów (np. tytuł, autor, słowa kluczowe, metadane, tekst)

Filtrowanie wyników wyszukiwania względem metadanych

- Filtrowanie obiektów o takich samych wartościach w określonych elementach metadanych

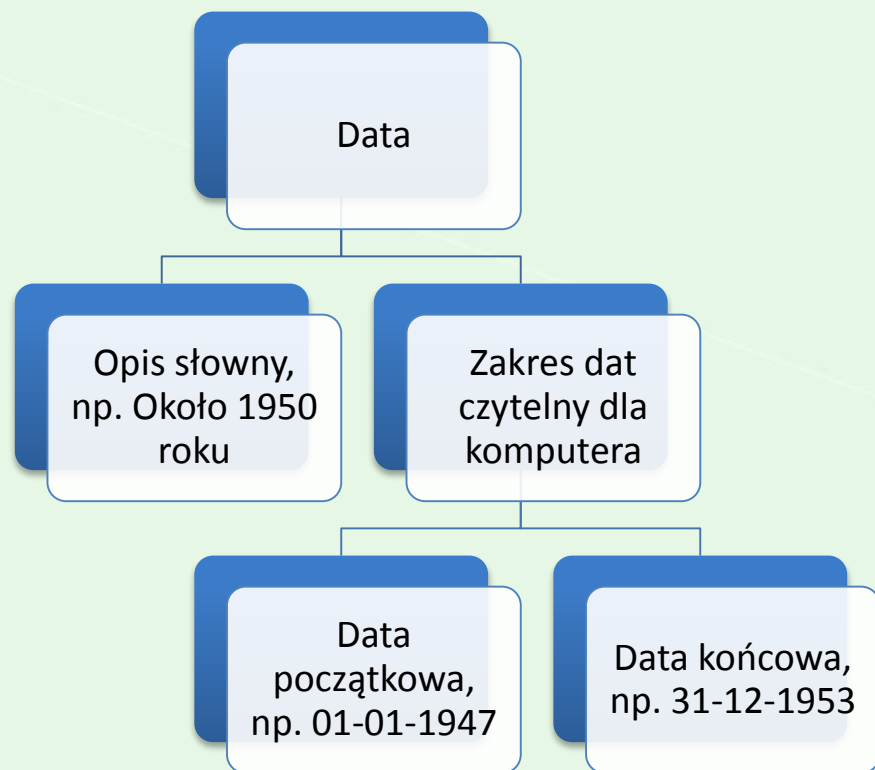
Filtrowanie wyników wyszukiwania względem publikacji grupowych, kolekcji, metadanych i tekstu

- Dodatkowy mechanizm filtrowania, wykraczający poza standardowe możliwości filtrowania metadanych



DInGO: dLibra

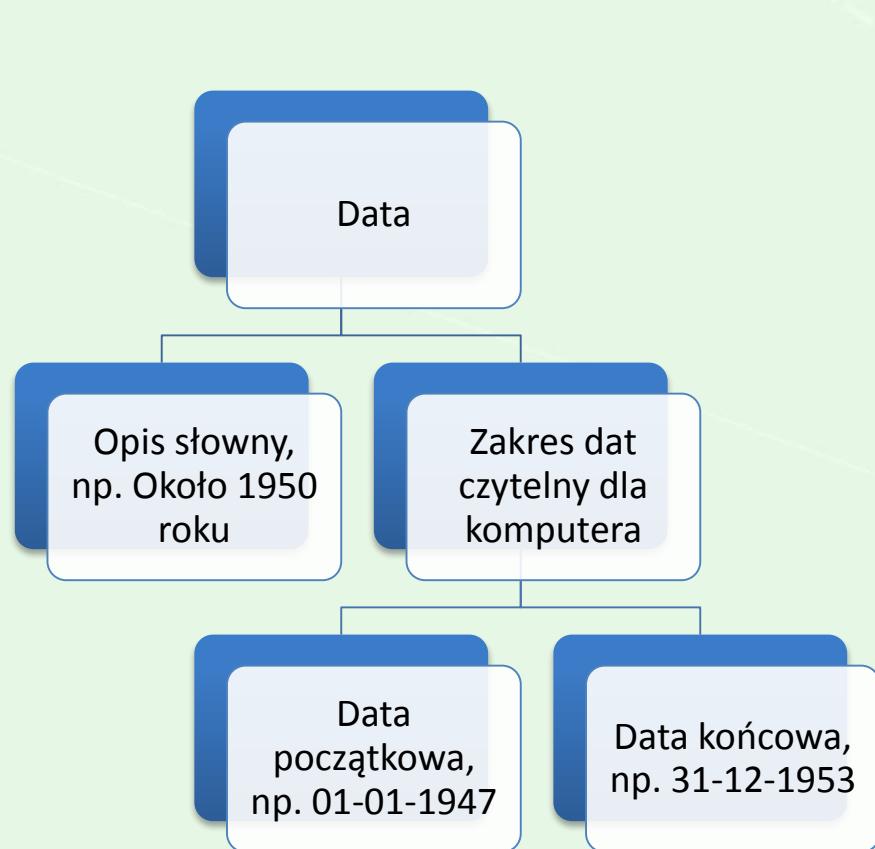
Nowe funkcje w wersji 6: typy atrybutów – data w opisie obiektu





DInGO: dLibra

Nowe funkcje w wersji 6: typy atrybutów – data w opisie obiektu



Mechanizm automatycznego rozpoznawania zakresów dat

- Migracja do wersji 6
- Wprowadzanie daty (ręczne lub import)

Dedykowany panel w Aplikacji Redaktora i Administratora

- Możliwość korekty rozpoznanych dat
- Możliwość wprowadzenia zakresu bezpośrednio do systemu

Nowe możliwości w wyszukiwaniu i przeglądaniu obiektów

- Sortowanie wyników wyszukiwania po dacie
- Zawężanie wyników sortowania względem daty
- Automatyczne kategorie uwzględniające datę, np. epoki literackie, wydarzenia/okresy w dziejach kraju



DInGO: dLibra

Nowe funkcje w wersji 6: przeglądarka DjVu w formie galerii JPG

Wykorzystuje istniejące mechanizmy konwersji w systemie (konwersja w ramach udostępniania profilu mets_exp poprzez OAI-PMH)

Pozwala na przeglądanie dokumentów DjVu bez konieczności instalacji wtyczek

Wykorzystuje bardzo popularne narzędzie BookReader (wykorzystywane np. przez Internet Archive)

Wiele ciekawych opcji przeglądania (pojedyncza strona, dwie strony, galeria stron, automat)

Wbudowany widok metadanych oraz opcji społecznościowych/konta.



DInGO: dLibra

Nowe funkcje w wersji 6: uproszczenie sposobu przeglądania zasobów

Łatwiejsze
przeglądanie indeksów
wartości
(dwupoziomowa
struktura przeglądania)

Publikacje bez
uprawnień do
wyświetlania wskazują
taki stan miniaturką
(strona obiektu, listy
obiektów,
wyszukiwanie)

Listy obiektów w
kolekcji posiadają
możliwość filtrowania
(podobnie jak wyniki
wyszukiwania)



DInGO: dLibra

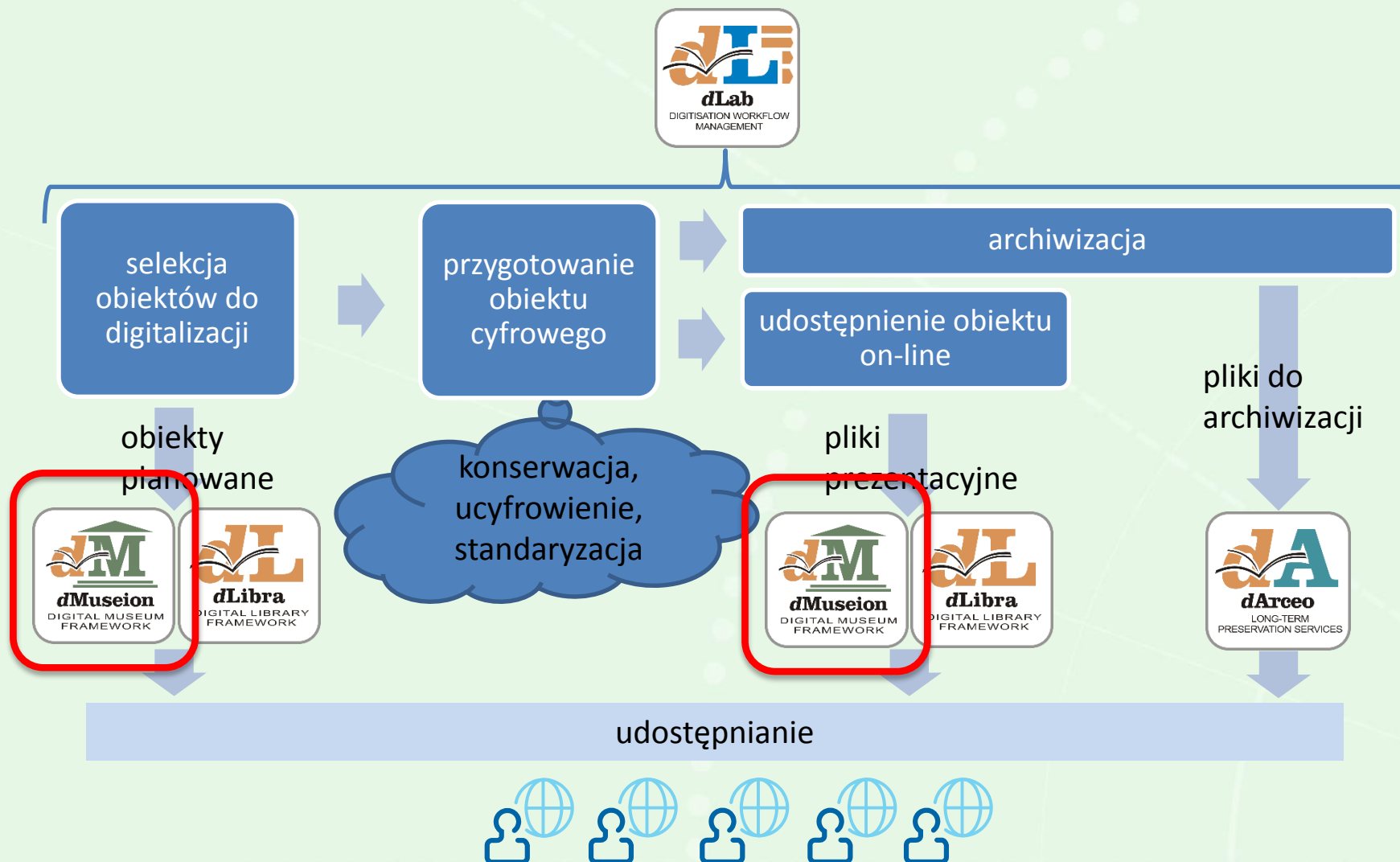
Nowe funkcje w wersji 6: demonstracja



DInGO: dMuseion



DInGO: proces digitalizacji (dMuseion)





DInGO: dMuseion

Informacje podstawowe

- Rozwijany od 2009 roku
- Wspólny serwer z systemem dLibra (synergia)
 - Nowe funkcje, udoskonalenia, doświadczenie
- Realizuje specyficzne potrzeby muzeów i galerii cyfrowych
- Wdrożenia mają z reguły dedykowany charakter
- Obecnie dwa publicznie dostępne wdrożenia: Cyfrowe Muzeum Narodowe w Warszawie oraz Centralna Baza Judaików





DInGO: dMuseion

Kluczowe funkcje i właściwości

Dedykowany wygląd WWW

- Terminologia
- Specyficzne potrzeby

Kategoryzacja zasobów

- Kolekcje zdefiniowane na stałe
- Wystawy kreowane na bieżąco
- Struktura muzeum (działy)

Prezentacja zabytków złożonych

- Wskazywanie powiązanych obiektów

Schowek użytkownika

- Drukowanie
- Współdzielenie

Współpraca z systemem inwentaryzacji zabytków

- Metadane
- Pliki
- Automatyzacja



DInGO: dMuseion

Przykład wdrożenia w Muzeum Narodowym w Warszawie (2009-2015): proces wprowadzania danych do systemu

System inwentaryzacji MONA

Pliki eksportowe

Narzędzie importujące do dMuseion

dMuseion

DInGO: dMuseion

Przykład wdrożenia w Muzeum Narodowym w Warszawie (2009-2015): proces wprowadzania danych do systemu

Narzędzie importujące do dMuseion

Wsadowy tryb działania

Mapowanie metadanych

Aktualizacja oraz tworzenie kolekcji i obiektów

Aktualizacja tylko plików lub tylko metadanych

Przypisanie do kolekcji i folderu na podstawie metadanych

Raportowanie dot. opublikowanych obiektów/pokazów

DInGO: dMuseion

Przykład wdrożenia w Muzeum Narodowym w Warszawie (2009-2015): nowe funkcje w DInGO



Usprawnienia w
przeglądaniu indeksów



Filtrowanie w
wyszukiwaniu



Tymczasowy schowek



DInGO: dMuseion

Przykład wdrożenia w Muzeum Narodowym w Warszawie (2009-2015): strona WWW



[Pomoc](#) | [Informacje](#) | [Strona główna MNW](#)

Zaloguj się

Zarejestruj się

Zapomniałeś hasła?

[Share / Save](#)

Szukanie zaawansowane

☐ Ze zdjęciem

☐ Z glossą



Zbiory muzeum



Pokazy tematyczne



Indeks artystów



Rodzaje zabytków

Czym są Cyfrowe Zbiory MNW?

Muzeum Narodowe w Warszawie gromadzi nie tylko dzieła sztuki, ale także ich cyfrowe wizerunki oraz dane dzieł sztuki. Na tej stronie udostępniamy te informacje. Jest ich jeszcze niewiele w porównaniu z wielkością kolekcji MNW, która liczy ponad 800 tysięcy muzealiów, ale liczba udostępnionych zabytków systematycznie rośnie. Zachęcamy do częstego odwiedzania naszej strony!

Aktualności

Ostatnio dodane



DInGO: dMuseion

Przykład wdrożenia w Muzeum Narodowym w Warszawie (2009-2015): strona WWW



Szukanie zaawansowane



Zbiory muzeum



Rodzaje zabytków



Kobieta z dzieckiem

0 1cm 2 3 4 5

Opis obiektu | Pokaż wszystkie wizerunki

DInGO: dMuseion



Wyczyść schowek
Przejdź do listy
Czym jest schowek

WW/Cy

Szukanie zaawansowane



Zbiory muzeum



Rodzaje zabytków

Aktualności

- Zbiory muzeum
- Pokazy tematyczne
- Indeks artystów
- Rodzaje zabytków



Zbiory muzeum

Zbiory Muzeum Narodowego w Warszawie liczą ok. 830 tys. dzieł sztuki polskiej i światowej, od antyku do współczesności, i obejmują malarstwo, rzeźbę, rysunki oraz ryciny, fotografie, numizmaty, a także przedmioty sztuki użytkowej i wzornictwo. Muzeum ma cztery oddziały: Muzeum Plakatu w Wilanowie, Muzeum Rzeźby im. Xawerego Dunikowskiego w Królikarni, Muzeum Wnętrz w Otwocku Wielkim oraz Muzeum w Nieborowie i Arkadii. Na tej stronie można je przeglądać według kolekcji w skład których wchodzi.

Szewczyk
1885

Nieborów
Gorazdowski
1884

Szukanie zaawansowane ▼

☐ Ze zdjęciem ☐ Z glosą



Zbiory Sztuki Średniowiecznej

Liczba zabytków w pokazach: 23



Zbiory Polskiej Sztuki Nowożytnej

Liczba zabytków w pokazach: 3171



Zbiory Obcej Sztuki Nowożytnej

Liczba zabytków w pokazach: 912



Zbiory Sztuki Orientalnej

Liczba zabytków w pokazach: 879



Gabinet Monet i Medali

Liczba zabytków w pokazach:



Zbiory Sztuki Nowoczesnej

Liczba zabytków w pokazach: 682

DInGO: dMuseion



Wyczyść schowek
Przejdź do listy
Czym jest schowek

WW/Cy

Szukanie zaawansowane



Zbiory muzeum



Rodzaje zabytków

Aktualności

- Zbiory muzeum
- Pokazy tematyczne
- Indeks artystów
- Rodzaje zabytków

Szukanie zaawansowane

☐ Ze zdjęciem ☐ Z glosą

DEMO

Szewczyk
1885

Nieborów
Gorazdowski
1884

numizmaty, a także przedmioty sztuki użytkowej i wzornictwo. Muzeum ma cztery oddziały: Muzeum Plakatu w Wilanowie, Muzeum Rzeźby im. Xawerego Dunikowskiego w Królikarni, Muzeum Wnętrz w Otwocku Wielkim oraz Muzeum w Nieborowie i Arkadii. Na tej stronie można je przeglądać według kolekcji w skład których wchodzi.

Zbiory Sztuki Orientalnej

Liczba zabytków w pokazach: 879

Gabinet Monet i Medali

Liczba zabytków w pokazach:

Zbiory Sztuki Nowoczesnej

Liczba zabytków w pokazach: 682



DInGO: dMuseion

Przykład wdrożenia w Muzeum Historii Żydów Polskich (2014/2015): proces wprowadzania danych do systemu

System inwentaryzacji zabytków
Integra/Mediastore

Moduł importujący dane do
dMuseion

dMuseion

DInGO: dMuseion

Przykład wdrożenia w Muzeum Historii Żydów Polskich (2014/2015): proces wprowadzania danych do systemu

Moduł importujący do dMuseion

W pełni automatyczny tryb działania (aktualizacja co godzinę)

Mapowanie i ujednolicanie metadanych

Generowanie wersji prezentacyjnej (ograniczenie rozmiaru/rozdzielczości)

Opcjonalne nanoszenie znaku wodnego

Przypisanie do kolekcji i folderu na podstawie metadanych

DInGO: dMuseion

Przykład wdrożenia w Muzeum Historii Żydów Polskich (2014/2015): nowe funkcje w DInGO



Import słownika z pliku csv



Ulepszone wyszukiwanie z wykorzystaniem synonimów

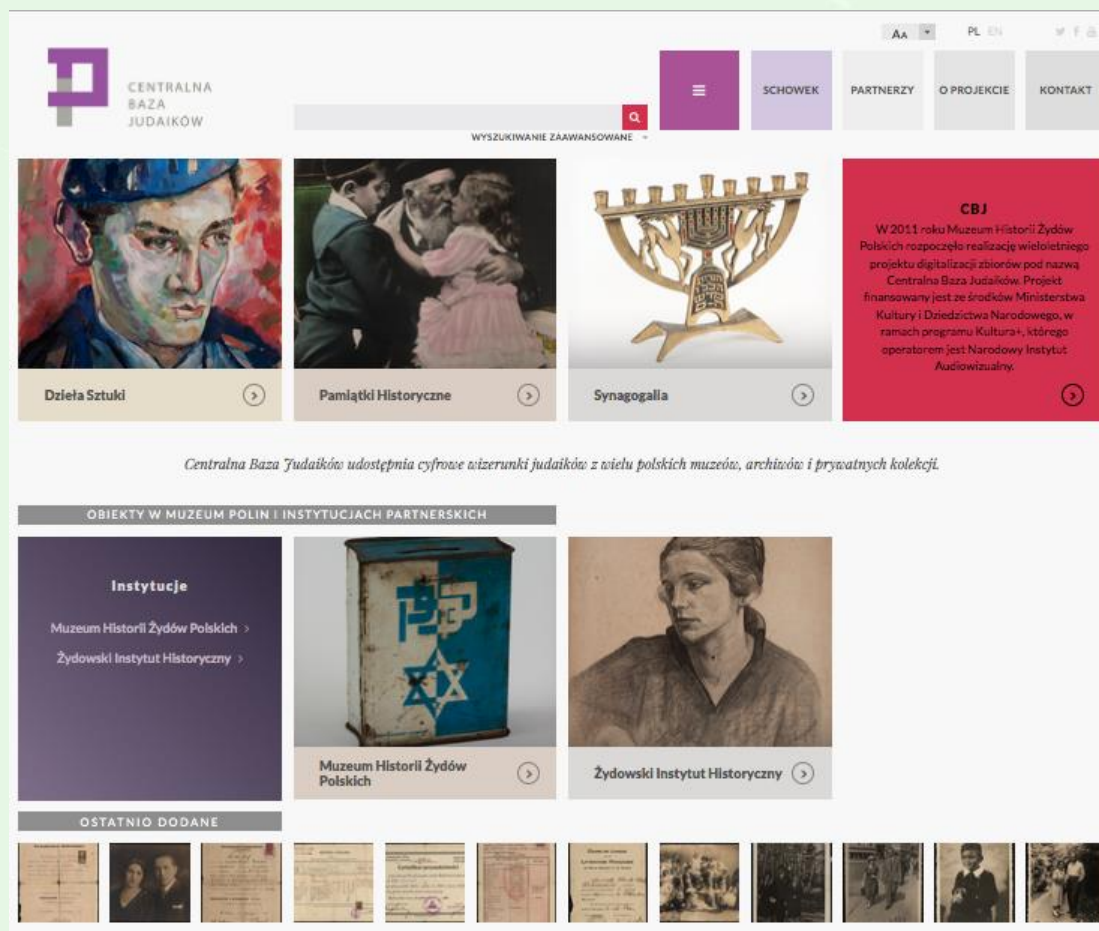


Listy obiektów tworzone na podstawie wyszukiwania (filtrowanie)



DInGO: dMuseion

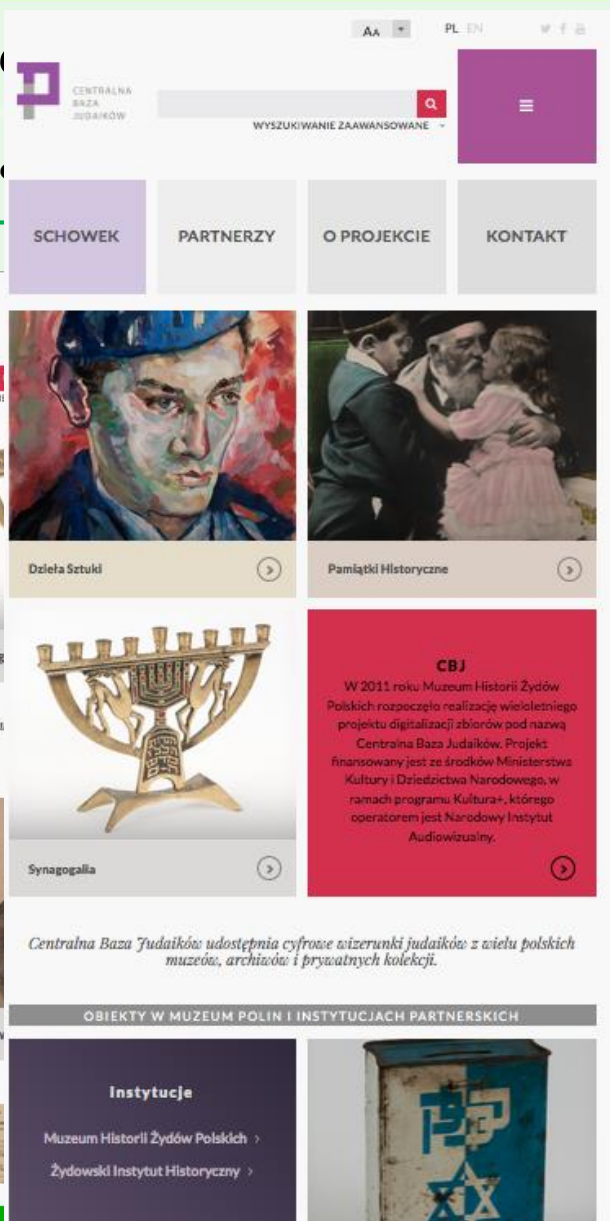
Przykład wdrożenia w Muzeum Historii Żydów Polskich (2014/2015)





DInGO: dMuseion

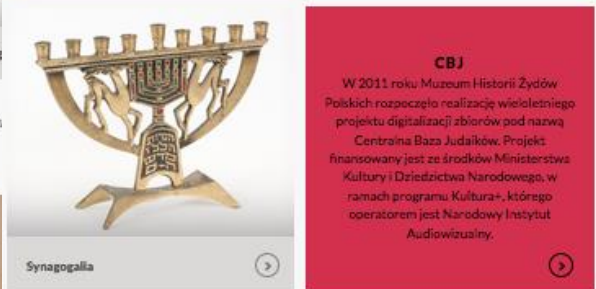
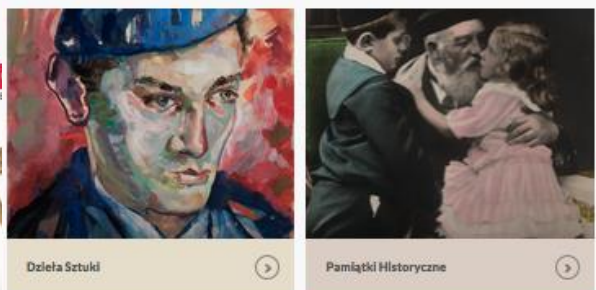
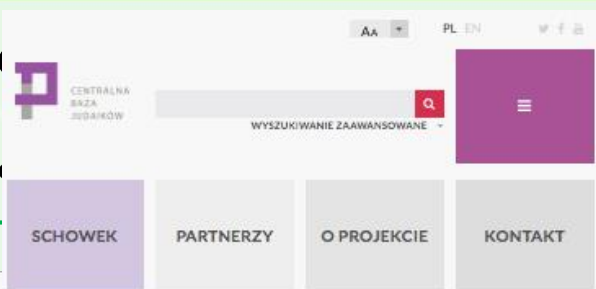
Przykład wdrożenia w Muzeum



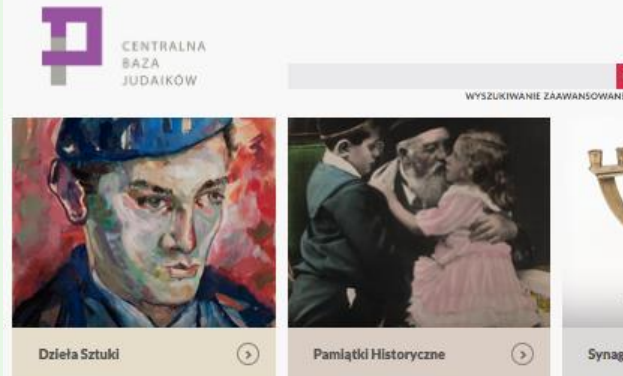
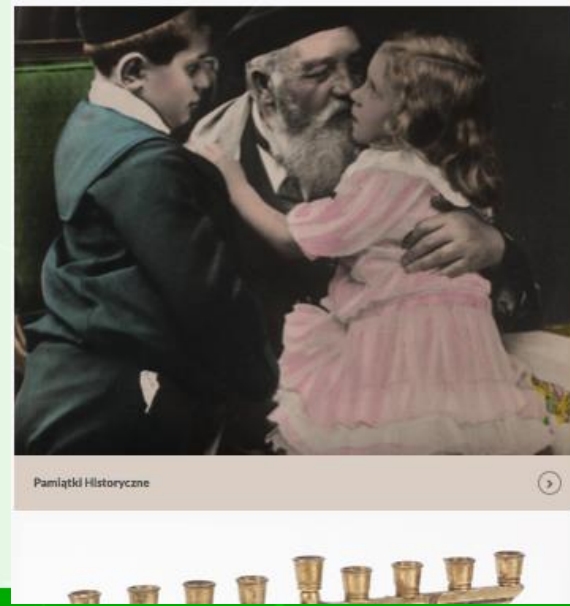
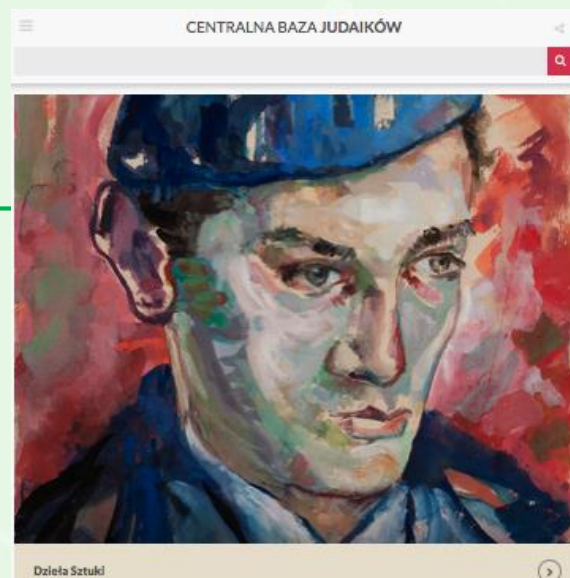


DInGO: dMus

Przykład wdrożenia w Muzeum



Centralna Baza Judaików udostępnia cyfrowe wizerunki judaików z wielu polskich muzeów, archiwów i prywatnych kolekcji.



Centralna Baza Judaików udostępnia cyfrowe wizerunki judaików z wielu





CENTRALNA
BAZA
JUDAIKÓW



CENTRALNA
BAZA
JUDAIKÓW

PL

CENTRALNA BAZA JUDAIKÓW

SCHOWEK

PARTNERZY

O PROJEKCIE

KONTAKT

Aa

PL EN



SCHOWEK

PARTNERZY

O PROJEKCIE

KONTAKT

WYSZUKIWANIE ZA AWANISOWANIE

STRONA GŁÓWNA

MUZEUM HISTORII ŻYDÓW POLSKICH

ŻYDOWSKI INSTYTUT HISTORYCZNY

SCHOWEK

PARTNERZY

O PROJEKCIE

KONTAKT

DZIEŁA SZTUKI

PAMIĄTKI HISTORYCZNE

SYNAGOGALIA

MUZEUM HISTORII ŻYDÓW POLSKICH

ŻYDOWSKI INSTYTUT HISTORYCZNY

- Dzieła Sztuki
- Pamiątki Historyczne
- Synagogalia

- Dzieła Sztuki
- Pamiątki Historyczne
- Synagogalia

czba obiektów: 982

Instytucje

Muzeum Historii Żydów Polskich
Żydowski Instytut Historyczny

OSTATNIO DODANE



PODOBNE OBIEKTY



Zdjęcie osoby

Autor nieznany (Brak informacji)
1922
1942P-08.01.12
Muzeum Historii Żydów Polskich



Zdjęcie osoby

Autor nieznany (Brak informacji)
1928
1942P-08.01.12
Muzeum Historii Żydów Polskich



Zdjęcie osoby

Autor nieznany (Brak informacji)
1928
1942P-08.01.12
Muzeum Historii Żydów Polskich



Zdjęcie osoby

Grupa/Wytwórnia Artykuł Kłania
1928
1942P-08.01.12
Muzeum Historii Żydów Polskich



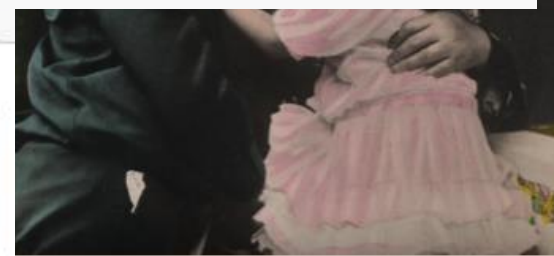
Zdjęcie osoby

Grupa/Wytwórnia Artykuł Kłania
1928
1942P-08.01.12
Muzeum Historii Żydów Polskich



Zdjęcie osoby

Autor nieznany (Brak informacji)
1928
1942P-08.01.12
Muzeum Historii Żydów Polskich



Pamiątki Historyczne





CENTRALNA
BAZA
JUDAIKÓW

PL

CENTRALNA BAZA JUDAIKÓW

SCHOWEK

PARTNERZY

O PROJEKCIE

KONTAKT

AA

PL EN



CENTRALNA
BAZA
JUDAIKÓW



SCHOWEK

PARTNERZY

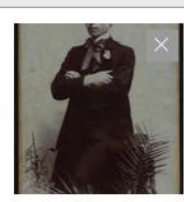
O PROJEKCIE

KONTAKT

WYSZUKIWANIE ZA AWANCOVANIE



Autoportret
malarski / Autoportret



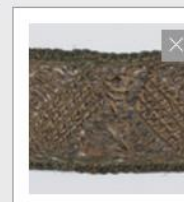
Zdjęcie osób



Zdjęcie osób



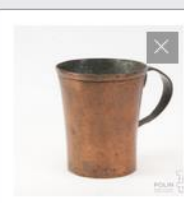
Zdjęcie osób



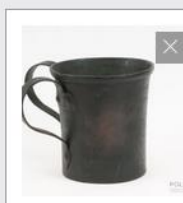
Atara



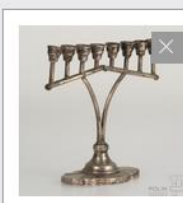
Jarmułka (kipa)



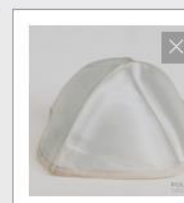
Kwarta do
rytualnych ablucji



Kwarta do
rytualnych ablucji



Lampka chanukowa
(chanukija)



Jarmułka (kipa)



Zdjęcie osób



OSTATNIO DODANE



Zdjęcie osób

Autor nieznany (Brak informacji)
1922
104229-08.01.12
Muzeum Historyczne Żydów Polaków



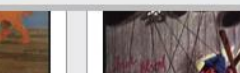
Zdjęcie osób

Autor nieznany (Brak informacji)
1928
104229-08.01.12
Muzeum Historyczne Żydów Polaków



Zdjęcie osób

Autor nieznany (Brak informacji)
1928
104229-08.01.12
Muzeum Historyczne Żydów Polaków



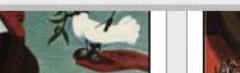
Zdjęcie osób

Grupa Wytwórców Artykułów
1928
104229-08.01.12
Muzeum Historyczne Żydów Polaków



Zdjęcie osób

Grupa Wytwórców Artykułów
1928
104229-08.01.12
Muzeum Historyczne Żydów Polaków

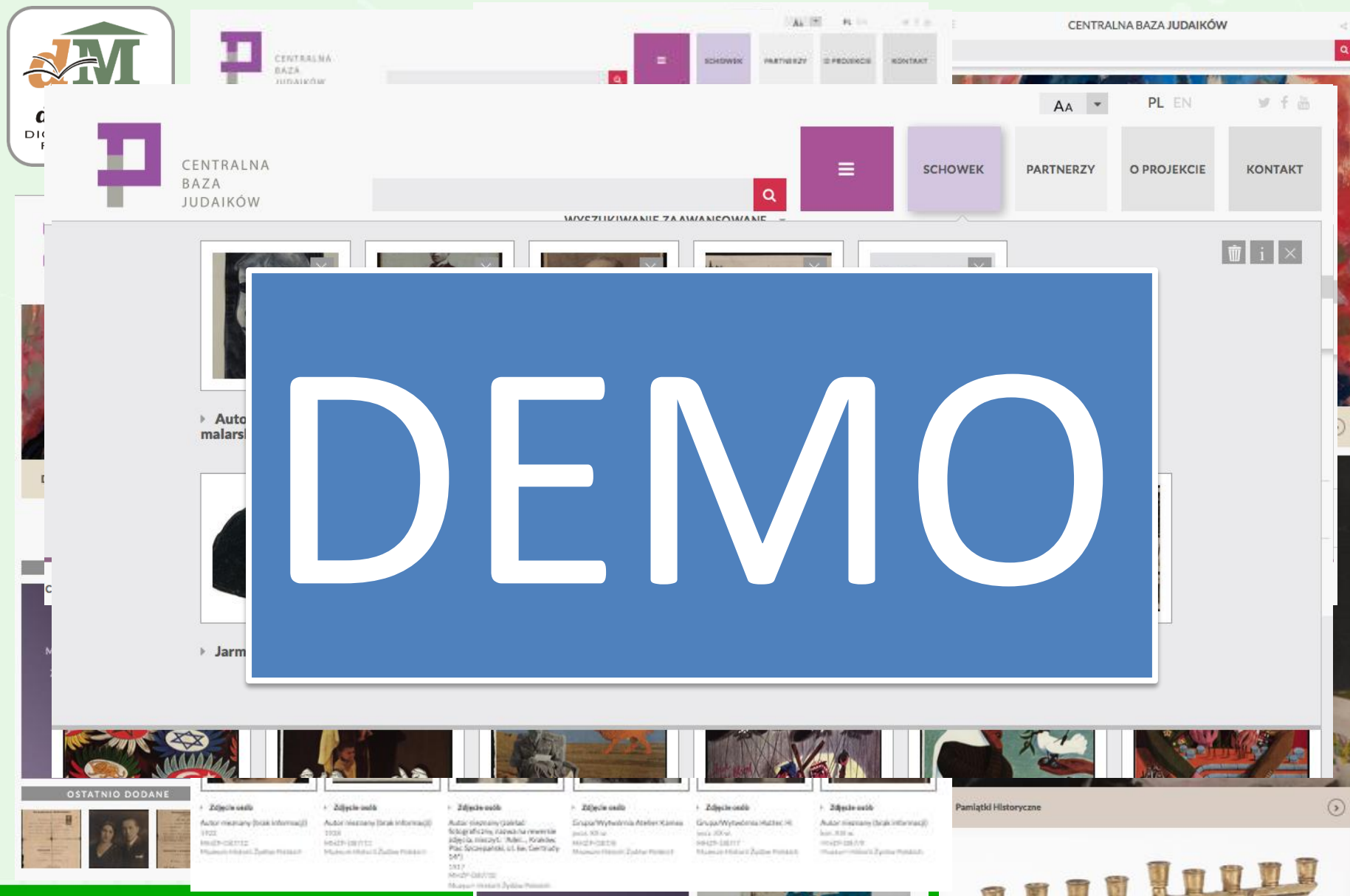


Zdjęcie osób

Autor nieznany (Brak informacji)
1928
104229-08.01.12
Muzeum Historyczne Żydów Polaków

Pamiętki Historyczne







DInGO: dMuseion

Wdrożenia w Muzeum Narodowym w Warszawie i Muzeum Historii Żydów Polskich: podsumowanie

Cyfrowe Muzeum Narodowe w Warszawie

System MONA

Narzędzie importujące uruchamiane na żądanie

Mapowanie danych

Wersja prezentacyjna pobierana na podstawie
wskazań w importowanych metadanych

Brak nanoszenia znaku wodnego

Przypisanie do kolekcji i folderu na podstawie
wskazań redaktora lub automatyczne

Cyfrowa Baza Judaików

System Integra i Mediastore

W pełni automatyczny tryb działania (aktualizacja co
godzinę)

Mapowanie i ujednolicanie metadanych

Generowanie wersji prezentacyjnej (ograniczenie
rozmiaru/rozdzielczości)

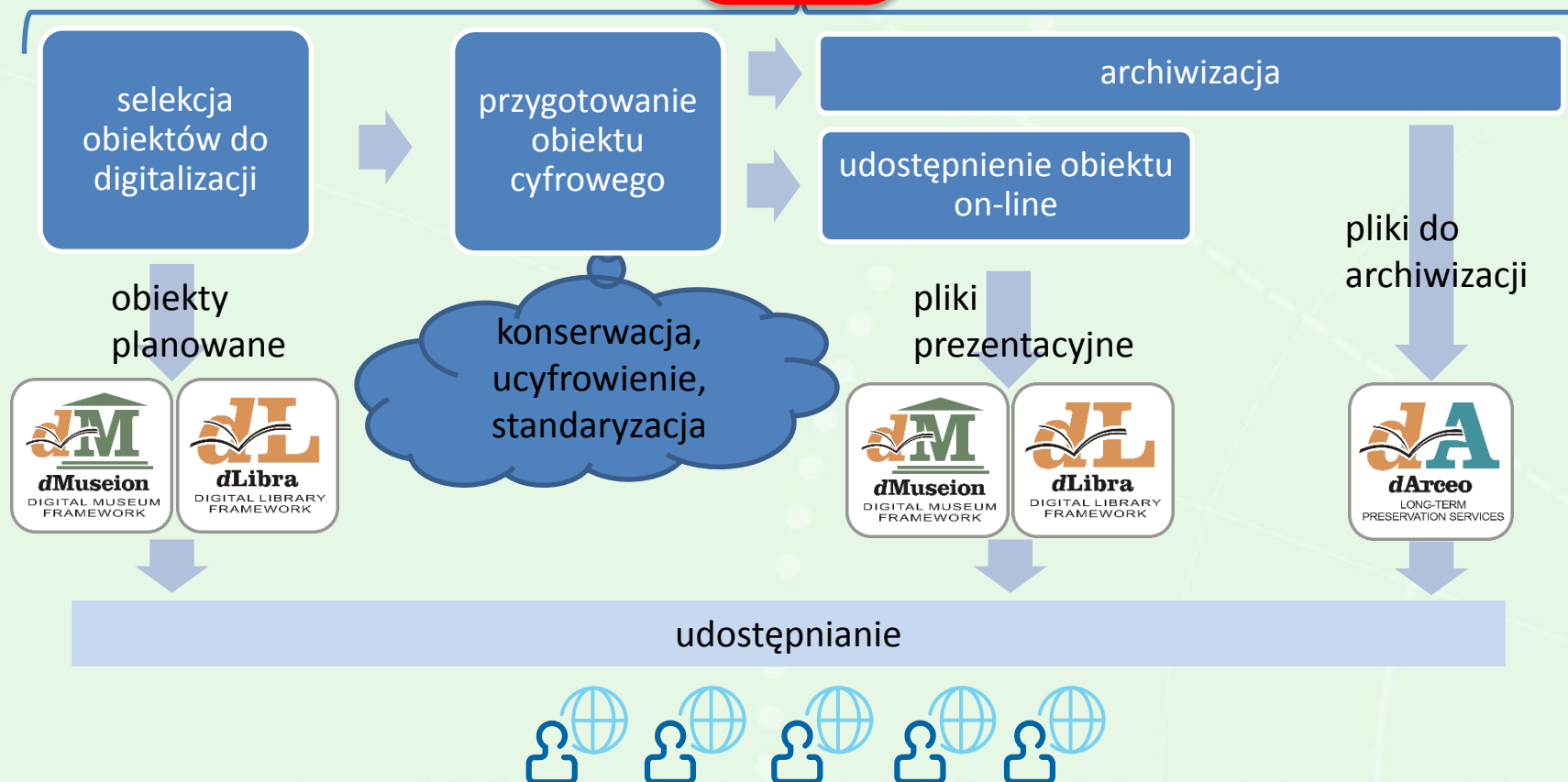
Opcjonalne nanoszenie znaku wodnego

Przypisanie do kolekcji i folderu automatycznie na
podstawie metadanych

DInGO: dLab



DInGO: proces digitalizacji (dLab)





DInGO: dLab

Informacje podstawowe

- Rozwijany od 2010 roku
- System pozwalający na pełną profesjonalizację projektów digitalizacyjnych
- Bardzo rozbudowane możliwości i duża elastyczność
- Dedykowane wdrożenie w kontekście potrzeb konkretnej instytucji lub projektu
- Obecnie 8 wdrożeń (1 w trakcie realizacji)



Proces digitalizacji

Zarządzanie przepływem prac

- Parametryzacja i planowanie
- Alternatywne procesy digitalizacji
- Podział prac na etapy
- Stan realizacji poszczególnych etapów pracy

Kontrolowanie i monitorowanie

- Historia zmian
- Monitorowanie postępu prac
- Punkty weryfikacji prac
- Kontrola jakości
- Raporty i podsumowania

Automatyzacja prac

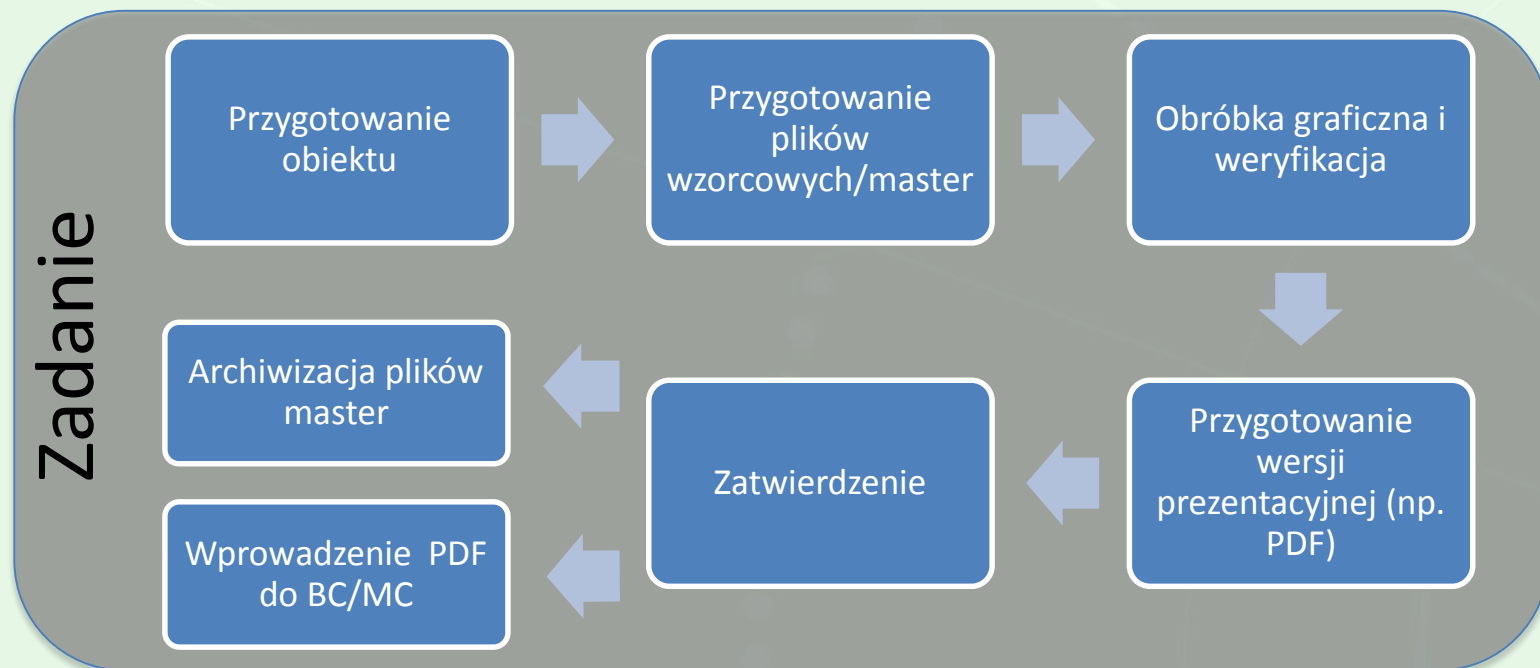
- Integracja zewnętrznych narzędzi (OCR, korekta graficzna, konwersja, zapewnianie jakości, ...)
- Komunikacja z zewnętrznymi systemami (udostępnianie, archiwizacja, ...)
- Eksport i import danych



DInGO: dLab

Podstawowe informacje

- Zadanie: podstawowy element w systemie, powizany z digitalizacją danego obiektu, np. książki, obrazu, rzeźby, monety
- Czynność: składowa zadania, reprezentująca działanie do wykonania w ramach procesu digitalizacji danego obiektu





DInGO: dLab

Zadanie, czynności i użytkownicy

Redaktor

Skanujący

Automat

Nadzorca

Zadanie

Przygotowanie
obiektu

Przygotowanie
plików
wzorcowych/master

Obróbka graficzna i
weryfikacja

Archiwizacja plików
master

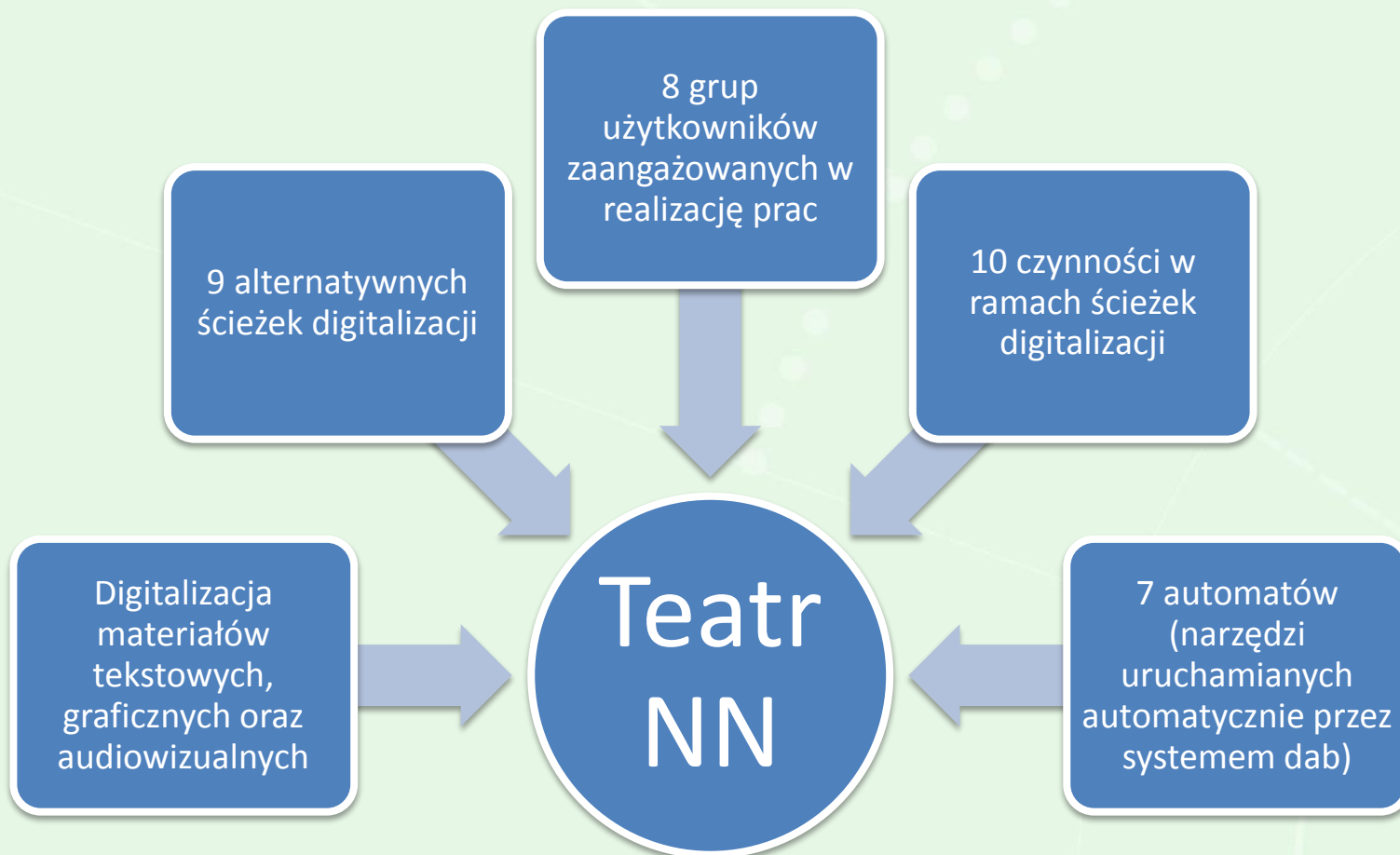
Wprowadzenie PDF
do BC/MC

Zatwierdzenie

Przygotowanie wersji
prezentacyjnej (np.
PDF)

DInGO: dLab

Przykłady wdrożeń: Ośrodek Brama Grodzka – Teatr NN – informacje ogólne





DInGO: dLab

Przykłady wdrożeń: Ośrodek Brama Grodzka – Teatr NN – czynności

Przygotowanie plików master

Weryfikacja plików master

Archiwizacja plików master

Archiwizacja na taśmie

Przygotowanie wersji prezentacyjnej

Weryfikacja wersji prezentacyjnej

Archiwizacja wersji prezentacyjnej

Przygotowanie metadanych

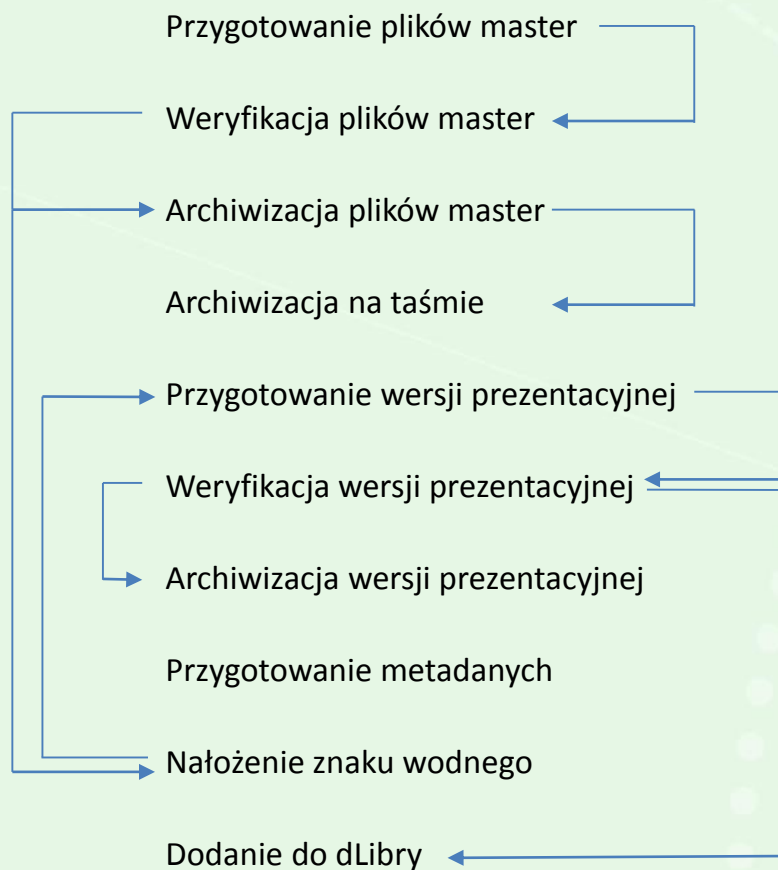
Nałożenie znaku wodnego

Dodanie do dLibry



DInGO: dLab

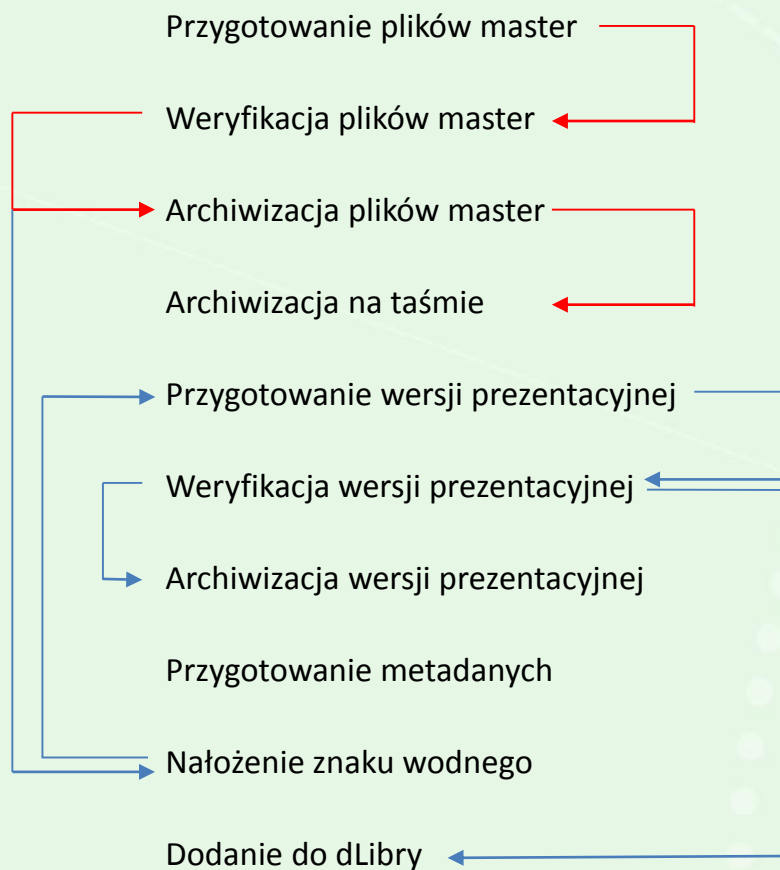
Przykłady wdrożeń: Ośrodek Brama Grodzka – Teatr NN – czynności





DInGO: dLab

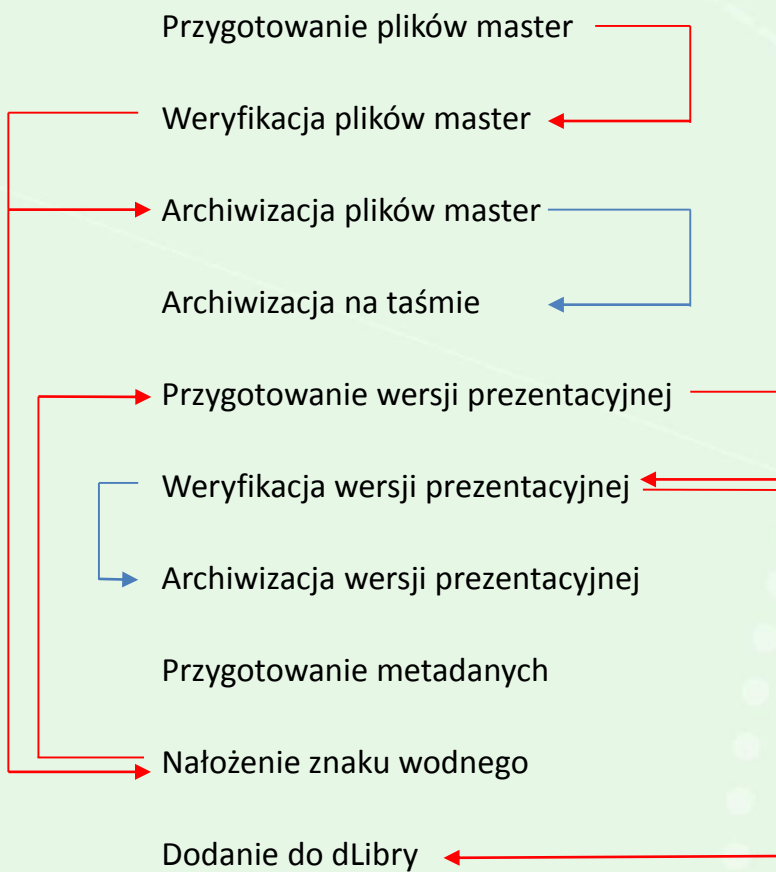
Przykłady wdrożeń: Ośrodek Brama Grodzka – Teatr NN – czynności





DInGO: dLab

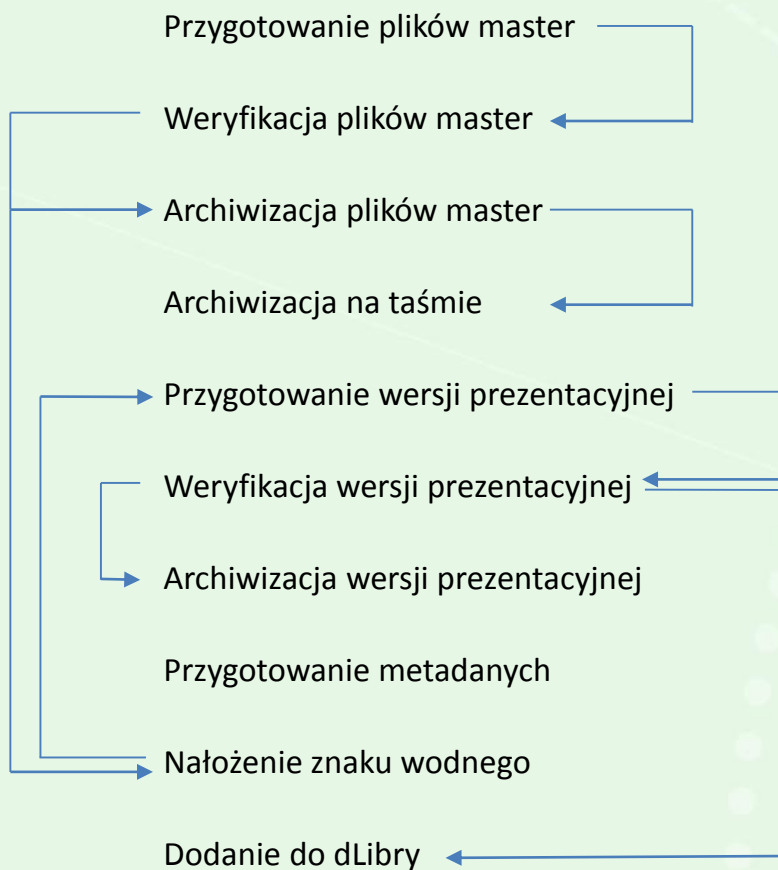
Przykłady wdrożeń: Ośrodek Brama Grodzka – Teatr NN – czynności





DInGO: dLab

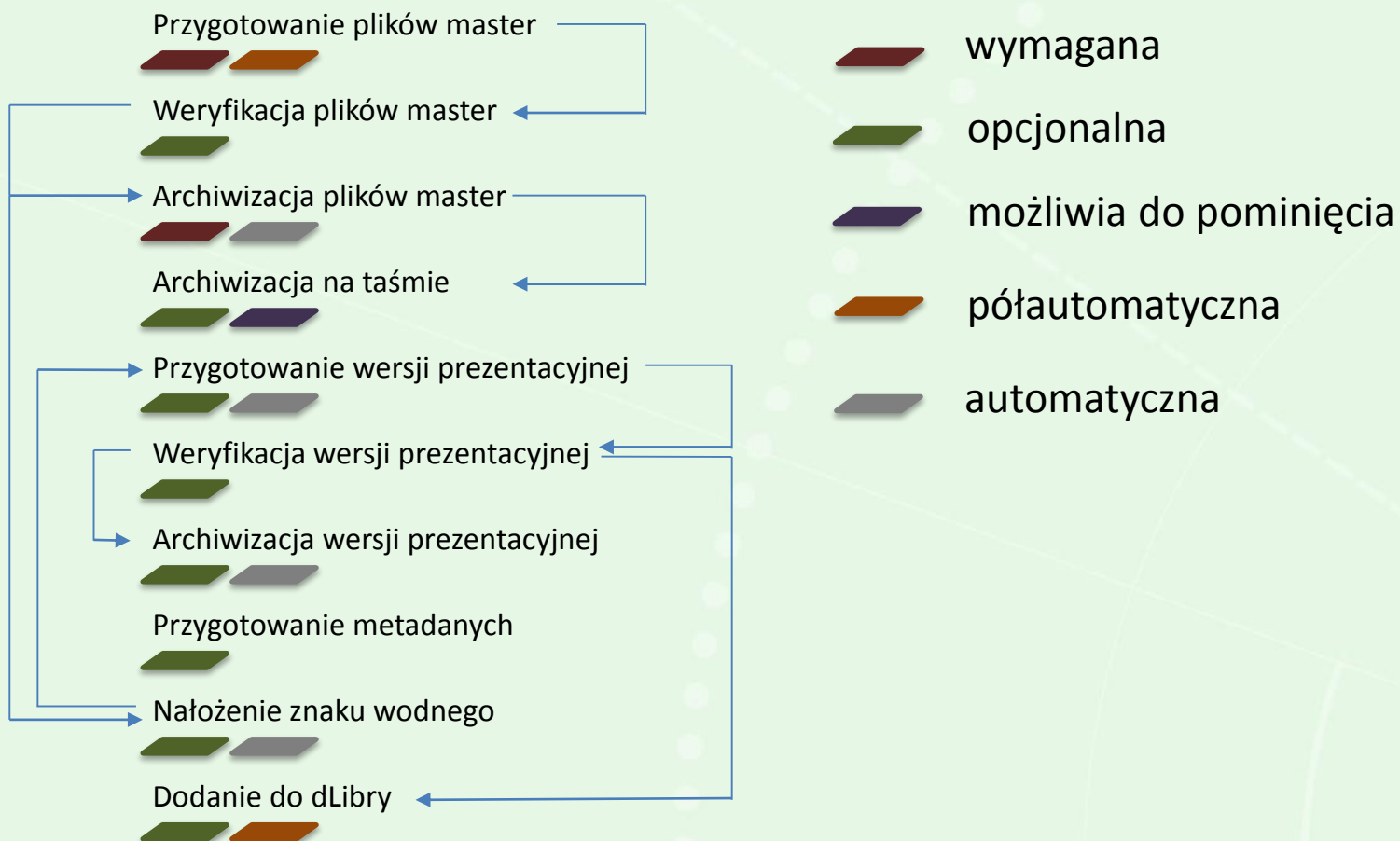
Przykłady wdrożeń: Ośrodek Brama Grodzka – Teatr NN – czynności





DInGO: dLab

Przykłady wdrożeń: Ośrodek Brama Grodzka – Teatr NN – czynności





DInGO: dLab

Przykłady wdrożeń: Ośrodek Brama Grodzka – Teatr NN – narzędzia automatyczne

Automat publikujący wersje prezentacyjne obiektów w systemie dLibra

Automat archiwizujący pliki master oraz pliki prezentacyjne w systemie dArceo

Automat generujący z plików master (TIFF, JPG) wersję prezentacyjną w postaci galerii obrazów (JPG).

Automat generujący z plików master (TIFF, JPG) wersję prezentacyjną w postaci dokumentu PDF wraz z warstwą tekstową (OCR).

Automat generujący z plików master (plik PDF o dużym rozmiarze) wersję prezentacyjną w postaci dokumentu PDF do udostępnienia w Internecie (mniejszy rozmiar pliku).

Automat generujący z plików master (plik w formacie ODT) wersję prezentacyjną w postaci dokumentu PDF.

Automat nanoszący znak wodny na pliki graficzne, które w dalszej kolejności są przetwarzane do wersji prezentacyjnych.

DInGO: dLab

Przykłady wdrożeń: Ośrodek Brama Grodzka – Teatr NN – nowe funkcje w DInGO



Dwa dedykowane raporty
dotyczące procesu digitalizacji



Tworzenie zadania bezpośrednio
w systemie dLab



Możliwość utworzenia publikacji
grupowej z publikacji planowanej

DInGO: dLab

Demonstracja

A stylized graphic of a folder or document. It has a white body with a blue rectangular label in the center. The label contains the word "DEMO" in large, white, sans-serif capital letters. Below the label, the text "DIGITISATION WORKFLOW MANAGEMENT" is written in smaller, black, sans-serif capital letters. The folder has a grey outline and a small orange and blue tab on top.

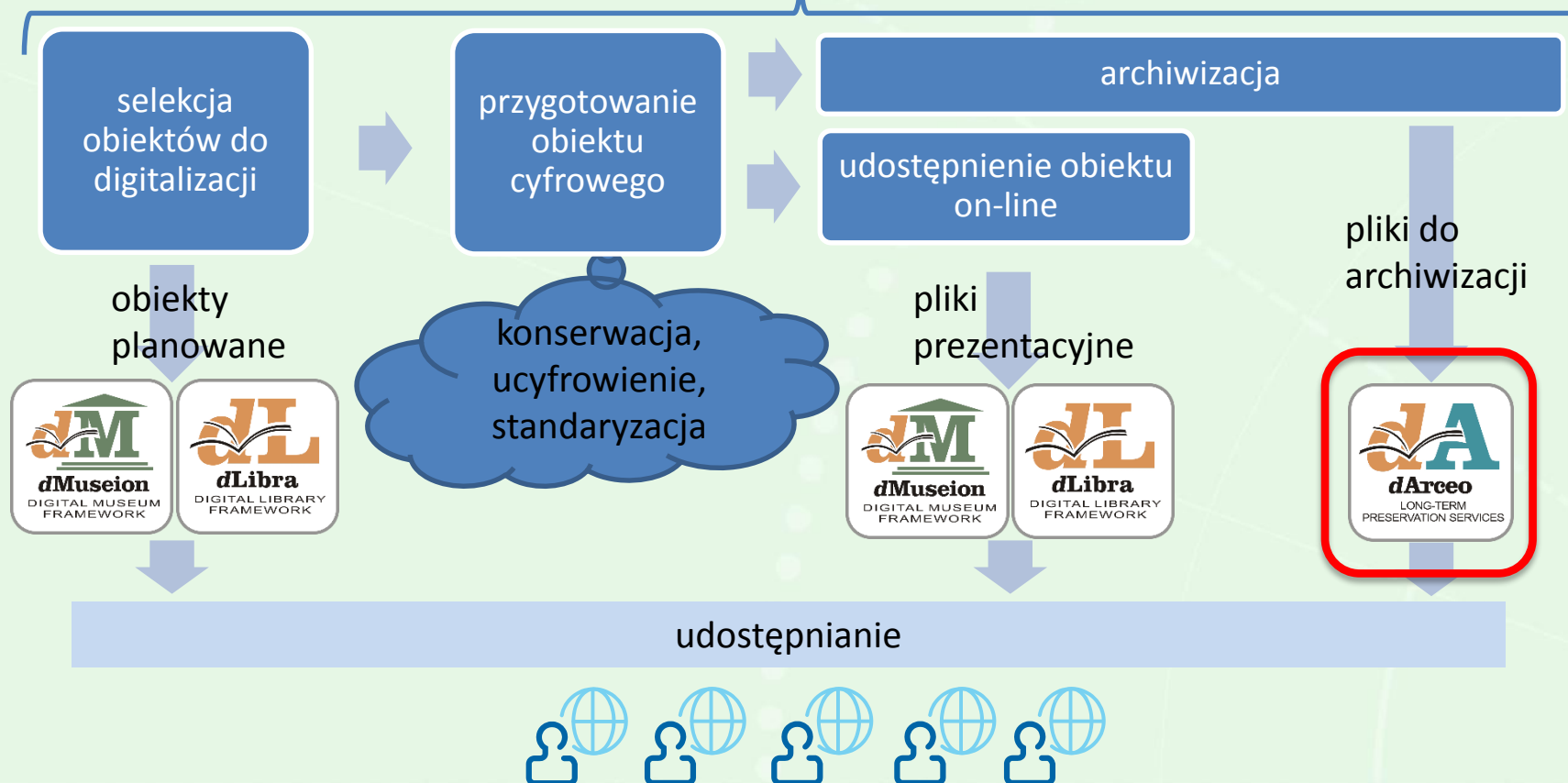
DEMO

DIGITISATION WORKFLOW
MANAGEMENT

DInGO: dArceo



DInGO: proces digitalizacji





DInGO: dArceo

Informacje podstawowe

- Rozwijany od 2011 roku
- System pozwalający na długoterminowe przechowywanie danych tekstowych, audiowizualnych i graficznych
- Założenia systemu pochodzą z modelu OAIS
- W pełni automatyczny tryb pracy
- Możliwość rozbudowy funkcji, np. konwersja, migracja danych
- Wdrażany wraz z systemem dLab, który wykorzystuje dArceo do składowania danych





DInGO: dArceo

Typowe wykorzystanie systemu

Aplikacja
użytkownika 1

Aplikacja
użytkownika 2

...

dArceo

Zasób
dyskowy A

Zasób
dyskowy B

...



DInGO: dArceo

Zarządzanie danymi w systemie

Wprowadzanie danych

- API programistyczne
- SIP: treść i metadane
- SCAPE Data Connector API

Metadane techniczne

- Bazuje na FITS oraz ffmpeg
- Identyfikacja formatu zgodnie z PRONOM
- Formaty: TextMD, AES57, MIX

Zawartość pakietu archiwalnego (AIP)

- SIP + wyekstrahowane metadane
- METS dla metadanych strukturalnych
- PREMIS dla metadanych dot. przechowywania

Dostęp do danych

- API programistyczne
- Bezpośredni dostęp do pliku
- Dostęp do całego pakietu archiwalnego AIP (w postaci ZIP)



DInGO: dArceo

Zarządzanie danymi w systemie

Adaptery zasobów dyskowych

- System plików (macierze, dyski, ...)
- **PLATON U4** - usługa powszechnej archiwizacji (SFTP)
- ...

Bezpieczeństwo danych

- Okresowe próby odczytu danych
- Sumy kontrolne (SHA-512)
- Mechanizm replikacji danych
- Weryfikacja formatów w odniesieniu do bazy UDFR



DInGO: dArceo

Przetwarzanie danych

Migracja

Na potrzeby długoterminowego przechowywania danych

Nowa wersja obiektu cyfrowego jest utworzona podczas migracji

Konwersja

Przetwarzanie danych ogólnego przeznaczenia

Przykładowo: utworzenie wersji prezentacyjnych

Zaawansowane dostarczanie

Funkcja prototypowa

Dostarczanie danych z uwzględnieniem możliwości użytkownika, np. urządzenia mobilne, czytniki e-book,...

DInGO: dArceo

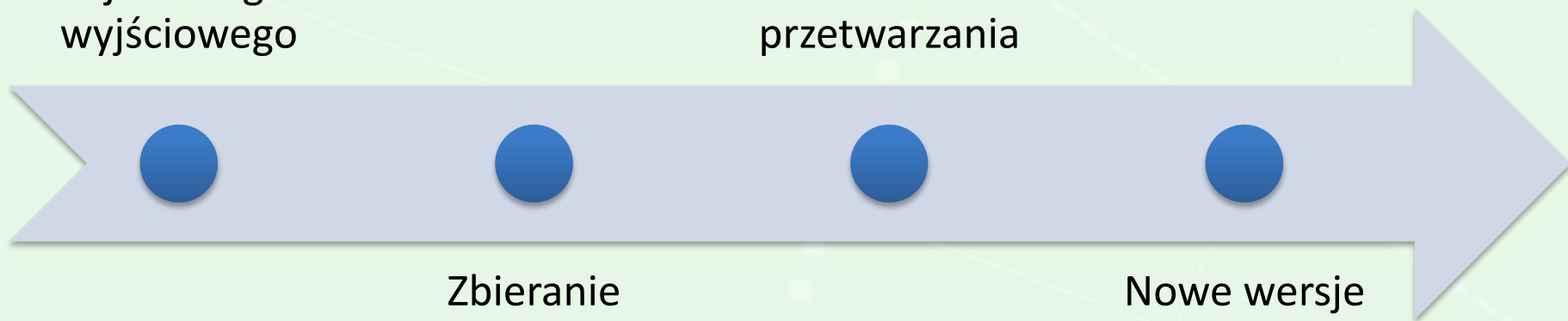
Przetwarzanie danych: przykład

Definicja
formatu
wejściowego i
wyjściowego

Wykonywanie
wybranej
ścieżki
przetwarzania

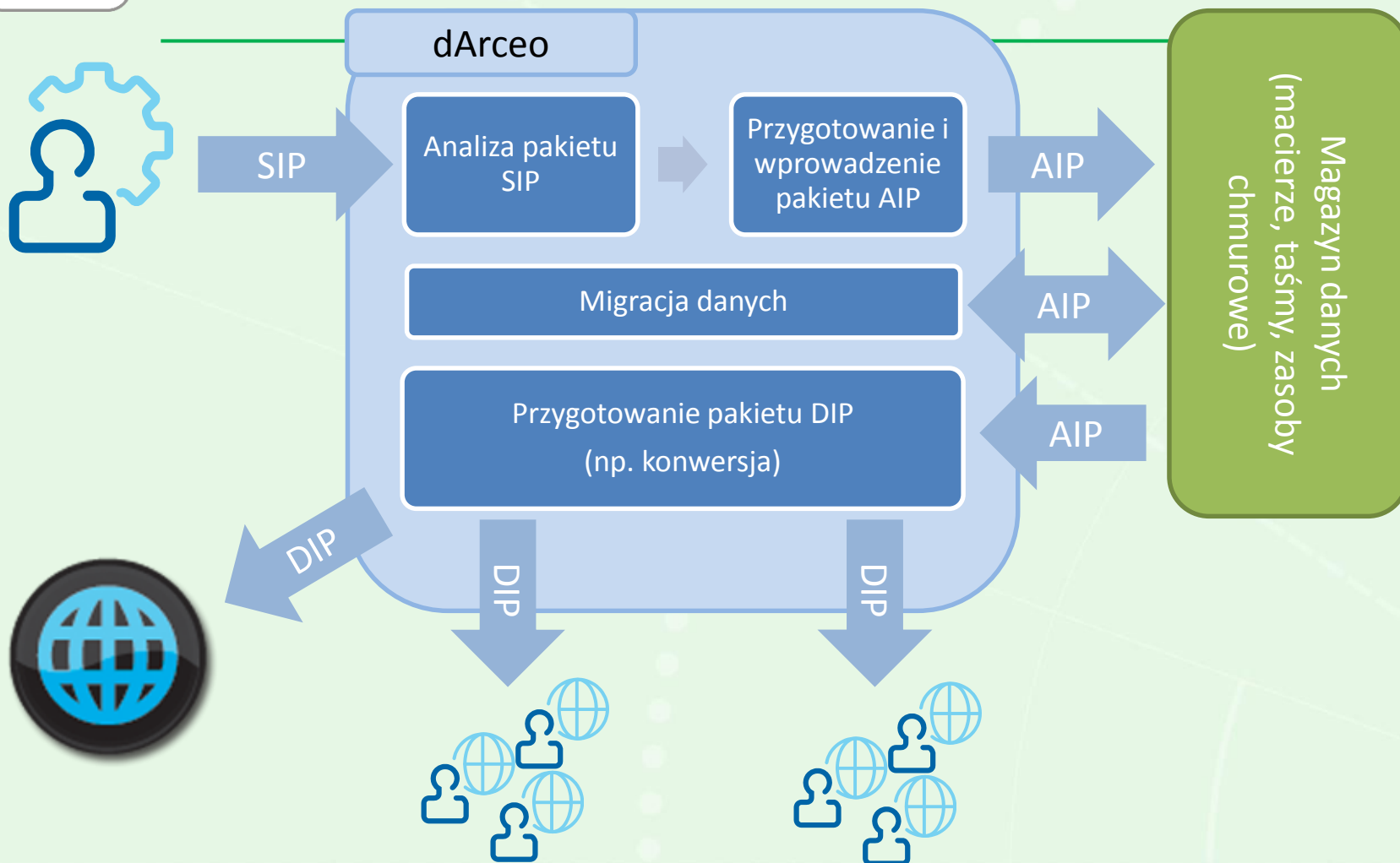
Zbieranie
informacji o
możliwych
ścieżkach
przetwarzania

Nowe wersje
obiektów w
dArceo





DInGO: dArceo i OAIS





DInGO: dArceo

Kluczowe możliwości/funkcje systemu: podsumowanie

Analiza danych

- Identyfikacja formatów plików danych zgodnie z bazą formatów PRONOM
- Ekstrakcja metadanych technicznych w zgodzie ze standardami TextMD, AES57, MIX

Składowanie danych

- Macierze dyskowe, taśmy, zewnętrzne usługi, np. chmurowe - PLATON U4
- Wbudowane mechanizmy replikacji danych
- Zapis pakietu archiwalnego zgodnie ze standardami METS oraz PREMIS

Monitorowanie i kontrolowanie danych

- Identyfikacja zagrożeń w odniesieniu do bazy UDFR (GDFR+PRONOM)
- Weryfikacja danych poprzez okresowy odczyt danych
- Zapewnienie wiarygodności poprzez weryfikację sum kontrolnych
- Generowanie raportów i podsumowań

Przetwarzanie danych

- Przetwarzanie danych zgodnie z założeniami modelu OAIS (migracja formatów)
- Rozszerzalny mechanizm usług migracji i konwertowania danych
- Planowanie migracji/konwersji danych

DInGO: dArceo

Panel administratora

General User-specific Messages Logout

Basic statistics

Metadata formats

Data file formats

Data manipulation services

MDZ configuration

Description	Value
Objects:	31,362
Files:	14,422,580
- data files:	4,779,776
- extracted metadata files:	9,555,540
- provided metadata files:	
Size:	
- data size:	
- extracted metadata size:	
- provided metadata size:	
Files per object:	
- data files per object:	
- extracted metadata files per object:	
- provided metadata files per object:	
Size per object:	
- data size per object:	
- extracted metadata size per object:	
- provided metadata size per object:	

General User-specific Messages Logout

Basic statistics

Metadata formats

Data file formats

Data manipulation services

MDZ configuration

Format	Version	Objects	Data files	Data files size
Digital Reproductions of Monographs and Manuscripts Image File Format	6.0	125	4705	153691110752
	1.2	420	420	38584687
	1.7	7	7	51831197
	1.1	1	1	246601

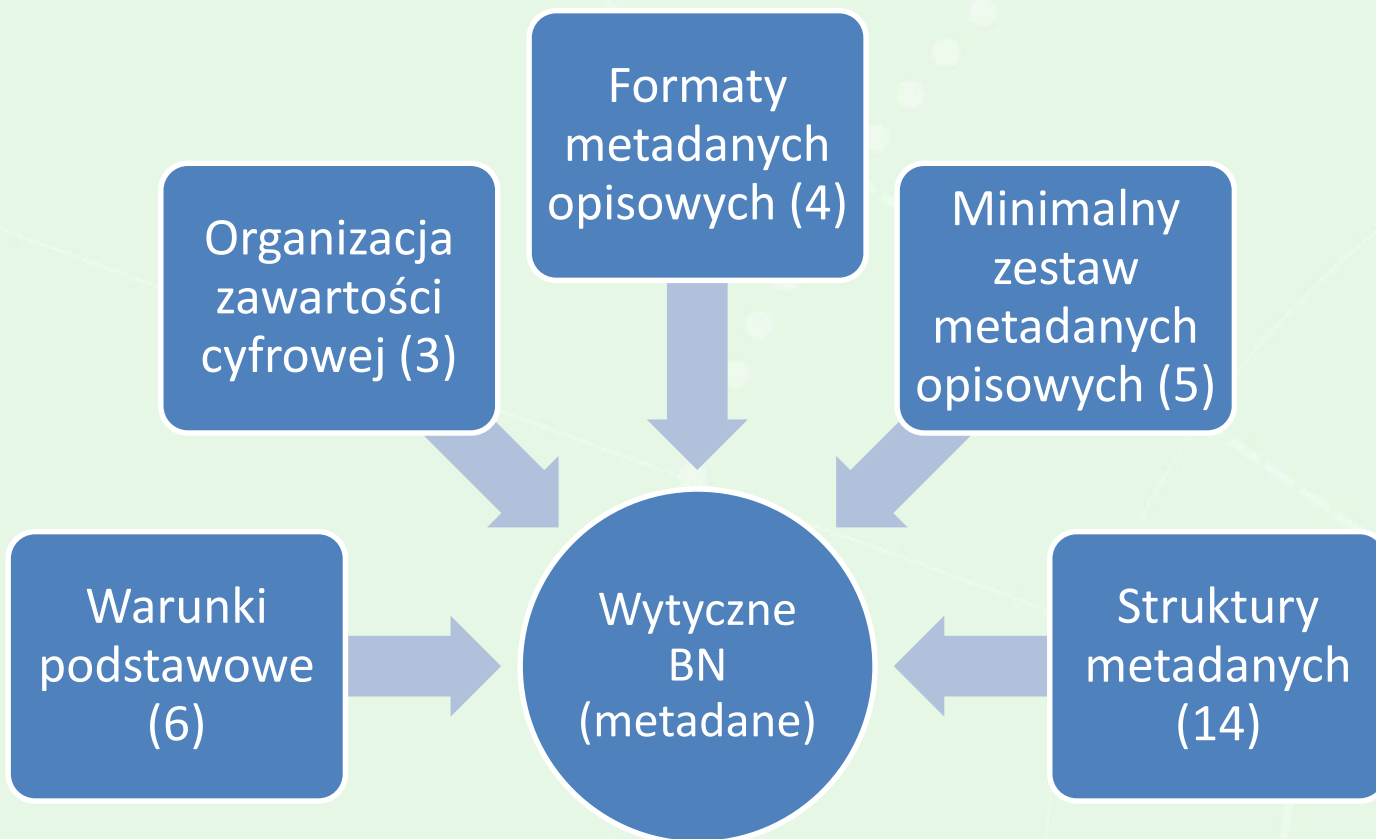
Format	Objects
DC	86850
METS	31459
PREMIS	31459
RDF	87258
UNKNOWN	87258
Format	Data files
AES57	236
DOCUMENTMD	1908
MIX	4742156
PREMIS	4779776
TEXTMD	3
VIDEO MD	2

2014-12-04 08:29

fmt/19	Portable Document Format, Tagged Image File Format	1.3	144	149	874731930
x-fmt/263	ZIP Format, raw H.263, Hypertext Markup Language, Tagged Image File Format	1.5	212	219	1599491748
fmt/40	Microsoft Word	2.0	1365	1365	358053304333
			2	2	275968

DInGO: zgodność w zakresie wymagań BN

Wymagania Biblioteki Narodowej (dotyczy metadanych)



DInGO: zgodność w zakresie wymagań BN

Poziomy zgodności

Pełny

- System DInGO spełnia dany punkt wytycznych w ramach wewnętrznego modelu składowania danych

Pełny po eksporcie

- System DInGO posiada niezbędne dane aby spełnić dany punkt
- Dostosowanie do konkretnej wytycznej realizowane byłoby na etapie eksportu danych

Potencjalny

- Spełnienie wytycznej wymaga uwzględnienia specyfiki pracy danej instytucji
- Może być zapewnione przez odpowiednie skonfigurowanie pakietu DInGO podczas wdrożenia

DInGO: zgodność w zakresie wymagań BN

Wytyczne BN – warunki podstawowe

2. Cała zawartość cyfrowa każdego obiektu musi być umieszczona w jednym osobnym folderze nazywanym dalej folderem obiektu.

3. W każdym folderze obiektu musi być jeden i tylko jeden plik METS opisujący ten obiekt nazywany dalej plikiem METS obiektu.

- Plik mets.xml znajduje się w katalogu /metadata/extracted/ - a nie w głównym katalogu obiektu

4. Plik METS obiektu musi zawierać wyłącznie lokalne odniesienia do innych plików w postaci ścieżek typu URN definiowanych względem położenia pliku METS.

5. Wszystkie dostarczone pliki zawartości cyfrowej obiektu muszą być wskazane w pliku METS obiektu (pozostała zawartość będzie traktowana jako niedostarczona).

6. Wszystkie metadane obiektu muszą być wskazane w pliku METS obiektu (pozostałe będą traktowane jako niedostarczone).

7. Akceptowane są wyłącznie poprawnie sformułowane pliki formacie METS w XML, które poprawnie przechodzą walidację na podstawie schematu XSD.

DInGO: zgodność w zakresie wymagań BN

Wytyczne BN – organizacja zawartości cyfrowej

8. Jeśli przekazywana jest jednocześnie zawartość cyfrowa w kilku postaciach (np. pliki źródłowe TIFF, pochodne JPG, PDF, ALTO) to pliki każdej z postaci można umieścić w osobnym podfolderze folderu obiektu.

- dArceo przechowuje powiązane ze sobą obiekty, tzn. wersja prezentacyjna czy produkcyjna (zoptymalizowana) jest odrębnym (ale powiązanim) obiektem w stosunku do obiektu master. Zależnie od modelu pracy danej instytucji, można połączyć poszczególne postacie w trakcie procesu eksportu.

9. Jeśli obiekt ma hierarchiczną strukturę elementów składowych, można ją odwzorować umieszczając elementy w osobnych podfolderach folderu obiektu.

10. Jeśli obiekty zorganizowane są w hierarchiczną strukturę (nadrzędną w stosunku do obiektów), można ją odwzorować umieszczając podfoldery obiektów w folderach grupujących.

- dArceo przechowuje informacje o konkretnych obiektach. Jeśli w dLab/dLibra mamy informacje o grupowaniu obiektów to można pogrupować obiekty dArceo, korzystając z informacji zawartych w dLab/dLibra.

DInGO: zgodność w zakresie wymagań BN

Wytyczne BN – formaty metadanych opisowych (1)

11. Akceptowanymi przez BN (z zastrzeżeniem w podpunkcie 13) formatami metadanych opisowych są obecnie:

- a. MARC21 (<http://www.bn.org.pl/dla-bibliotekarzy/nfs/marc/format-marc-21>)
- b. Marc XML (<http://www.loc.gov/standards/marcxml/>)
- c. MODS (<http://www.loc.gov/standards/mods/>)
- d. DCMI Metadata Terms (<http://dublincore.org/documents/dcmi-terms/>)
- e. Dlibra_AVS (<http://www.wbc.poznan.pl/dlibra/attribute-schema.xsd>)

•dLab wprowadza do systemu dArceo metadane w formacie Dublin Core (podzbiór DCMI Terms) oraz w formacie dlibra_avs. **UWAGA:** Wskazany w punkcie dlibra_avs adres URL jest poprawny tylko dla metadanych pochodzących z Wielkopolskiej Biblioteki Cyfrowej. Każda z bibliotek cyfrowych może posiadać swój własny schemat i skojarzony z nim plik XSD. Tak więc obecny zapis z odwołaniem do WBC oznacza, że tylko schemat WBC akceptowalny. Jest to prawdopodobnie pomyłka po stronie opracowującego wytyczne.

12. Jeśli biblioteka (lub inna instytucja przekazująca obiekty do BN) kataloguje obiekty źródłowe w formatach opisanych w punkcie 11, to wraz z przekazywaniem cyfrowych reprodukcji tych obiektów do BN należy przekazać wszystkie metadane obiektów źródłowych w formacie, w którym te metadane zostały oryginalnie wytworzone. Dostarczone pliki METS mają zawierać informacje pozwalające na identyfikację plików reprodukcji cyfrowych i odpowiadających im metadanych obiektów źródłowych.

•Konieczny moduł integrujący dLab z systemem katalogowania lub odpowiednia konfiguracja dLab, by przyjmował metadane w obiekcie źródłowego i przekazywał je dalej do dArceo.

DInGO: zgodność w zakresie wymagań BN

Wytyczne BN – formaty metadanych opisowych (2)

13. Jeśli biblioteka (lub inna instytucja przekazująca obiekty do BN) kataloguje obiekty źródłowe w innych formatach niż opisanych w punkcie 11, to należy przekazać wszystkie metadane obiektów źródłowych po skonwertowaniu na jeden z formatów opisanych w punkcie 11 w sposób powodujący możliwie najmniejszą utratę informacji.

- Zobacz komentarz do punktu 12. Dodatkowo należy opracować mechanizm konwersji formatu.

14. W szczególnych przypadkach, pod warunkiem wcześniejszego pisemnego uzgodnienia między instytucją przekazującą a BN kwestii technicznych i finansowych, w okolicznościach stosowania innego standardu metadanych opisowych istnieje możliwość przygotowania we współpracy z BN odpowiedniego konwertera.

- Patrz komentarz do punktu 13.

DInGO: zgodność w zakresie wymagań BN

Wytyczne BN – minimalny zestaw metadanych opisowych

15-19. Wytyczne opisują minimalny zestaw metadanych dla poszczególnych typów obiektów.

- Metadane muszą być wypełnione przez redaktorów w systemie dLibra - kwestia ustalenia schematu metadanych w dLibrze i widoków dla poszczególnych typów obiektów, w tym określenia które pola są obowiązkowe w poszczególnych widokach.

DInGO: zgodność w zakresie wymagań BN

Wytyczne BN – struktury metadanych (1)

20. W pliku METS obiektu powinna wystąpić przynajmniej jedna sekcja <METS:dmdSec> z metadanymi opisowymi dotyczącymi całego obiektu. Sekcja ta powinna być użyta (dowiązana) przy użyciu identyfikatora w najmniej zagłębionym elemencie <METS:div> elementu <METS:structMap>.

- W dArceo jest sekcja dmdSec z metadanymi ogólnymi całego obiektu. dArceo domyślnie nie posiada informacji strukturalnych (np. o podziale fizycznego dokumentu na strony), można natomiast zrobić podstawowe structMap w oparciu o strukturę obiektu cyfrowego (pliki, katalogi) tak jak to jest w tej chwili w przypadku eksportu metadanych w METS przez OAI-PMH z dLibry.

21. Zestawy metadanych obiektu muszą mieć postać XML i być osadzone w pliku METS obiektu lub wskazywane przez URN w zewnętrznych plikach XML umieszczonych w folderze obiektu.

22. Metadane deskryptywne (opisowe) odnoszące się do obiektu źródłowego mają być wskazane w sekcji <dmdSec> METS, opisane elementem <mdWrap> i zawarte w postaci XML w pliku METS lub opisane elementem <mdRef> wskazującym na zewnętrzny plik XML. W przypadku przekazywania metadanych w formacie binarnym MARC21, mają one być opisane elementem <mdRef> wskazującym na zewnętrzny plik, w którym mają być umieszczone formacie MARC ISO 2709.

- Patrz komentarz do punktu 12 i 13.

DInGO: zgodność w zakresie wymagań BN

Wytyczne BN – struktury metadanych (2)

23. Jeśli istnieje potrzeba wyróżnienia metadanych opisowych odnoszących się do obiektu źródłowego w stosunku do opisywanego obiektu cyfrowego (zawartości cyfrowej) metadane takie mogą zostać zawarte w sekcji <Mets:SourceMD>, jednak niezależnie od tego wszystkie metadane opisowe prawdziwe dla obiektu cyfrowego mają znaleźć się w podstawowej sekcji <METS:dmdSec> opisującej zawartość cyfrową.

- Patrz komentarz do punktu 12 i 13.

24. Metadane opisowe mogą zostać użyte do opisu pojedynczego fragmentu zawartości cyfrowej (np. pojedynczego skanu). W takim przypadku należy użyć identyfikatora metadanych w elemencie <METS:file>.

Przykład

```
<METS:file ID="FILE001" MIMETYPE="image/jpeg" CREATED="2003-07-04T15:00:00"  
SEQ="1" GROUPID="1" ADMID="ADM1" DMDID="dmd1">
```

- Należałoby odpowiednio skonfigurować system dLab, by umożliwiał wprowadzanie tego typu informacji i przekazywał je na etapie archiwizacji do dArceo.

25. Metadane opisowe mogą zostać użyte na bardziej zagłębionych poziomach <METS:div> będą wówczas dotyczyły tylko części struktury dokumentu (np. rozdziału, ilustracji, mapy, strony).

- Należałoby odpowiednio skonfigurować system dLab, by umożliwiał wprowadzanie tego typu informacji i przekazywał je na etapie archiwizacji do dArceo.

DInGO: zgodność w zakresie wymagań BN

Wytyczne BN – struktury metadanych (3)

26. Dla obiektów, które powstały ze zdigitalizowanych numerów czasopism (w szczególności folderów zawierających zbiorcze pliki z wielu zeszytów), należy odpowiednio skonstruować sekcje METS fileSec oraz structMap tak, aby pozwalały zidentyfikować każdy numer i jeżeli występuje - dodatek do tego numeru.

- W systemie dLab poszczególne numery czasopisma to typowo poszczególne obiekty. W przypadku digitalizacji "klocków" operator dLab ma możliwość rozbijania jednej grupy skanów z takiego klocka na poszczególne obiekty reprezentujące odrębne numery.

27. Mocno zalecane jest używanie sumy kontrolnej MD5 dla elementów opisujących zawartość cyfrową. W przypadku użycia sumy kontrolnej pliki zawartości cyfrowej będą sprawdzane na wypadek błędów w oparciu o tę sumę.

Przykład

```
<METS:file ID="FID1" MIMETYPE="image/jpeg" SEQ="1" GROUPID="GID1"
CHECKSUMTYPE="MD5" CHECKSUM="A6C66A9F3832DECBACA2D0C7496F57FB">
<METS:FLocat LOCTYPE="URN" xlink:href="example2.jpg"/>
</METS:file>
```

- System dArceo wylicza, przechowuje i weryfikuje sumy kontrolne w ramach swoich wewnętrznych mechanizmów, wykorzystując standard SHA-512. Można wprowadzić do systemu dArceo drobną zmianę aby te sumy kontrolne oraz sumy MD5 były zapisywane dodatkowo w plikach METS. UWAGA: Standard MD5 nie jest już zalecany do używania w dużych zbiorach danych ze względu na dużą powtarzalność sum kontrolnych - czyli mniejszą skuteczność w wykrywaniu błędów.

DInGO: zgodność w zakresie wymagań BN

Wytyczne BN – struktury metadanych (4)

28. Metadane techniczne dotyczące pojedynczego pliku obrazu powinny być osadzone w plikach binarnych z użyciem standardu EXIF (i) lub umieszczone w odpowiednich sekcjach pliku METS.

- Umieszczone w sekcjach pliku METS, w postaci dowiązań do zewnętrznych plików XML.

29. Doprecyzowania metadanych technicznych powinny być zamieszczane w pliku METS w sekcji <Mets:admSec> <Mets:TechMD> na przykład przy użyciu standardu NISOIMG (inaczej MIX).

Przykłady dostępne są pod adresem:

<http://www.loc.gov/standards/mets/METSPRimerRevised.pdf>

- Dla obrazów dArceo wykorzystuje MIX.

30. Podstawowe metadane dotyczące praw obiektu powinny być umieszczone w sekcji <Mets:admSec> w podsekcji <mets:rightsMD>

- Należałoby skonfigurować system dLab tak, by pozwalał wprowadzać tego rodzaju informacje lub wykorzystywał wartość określonych pól opisowych z systemu dLibra, a następnie by przekazywał je do dArceo.

DInGO: zgodność w zakresie wymagań BN

Wytyczne BN – struktury metadanych (5)

31. Metadane dotyczące proveniencji oraz cyklu życia obiektu cyfrowego (jeśli istnieją) powinny zostać zawarte w sekcji <Mets:admSec> <Mets:Digiprov:MD> zaleca się użycie w tym celu standardu PREMIS. Opis standardu PREMIS dostępny jest pod adresem:

<http://www.loc.gov/standards/premis/>

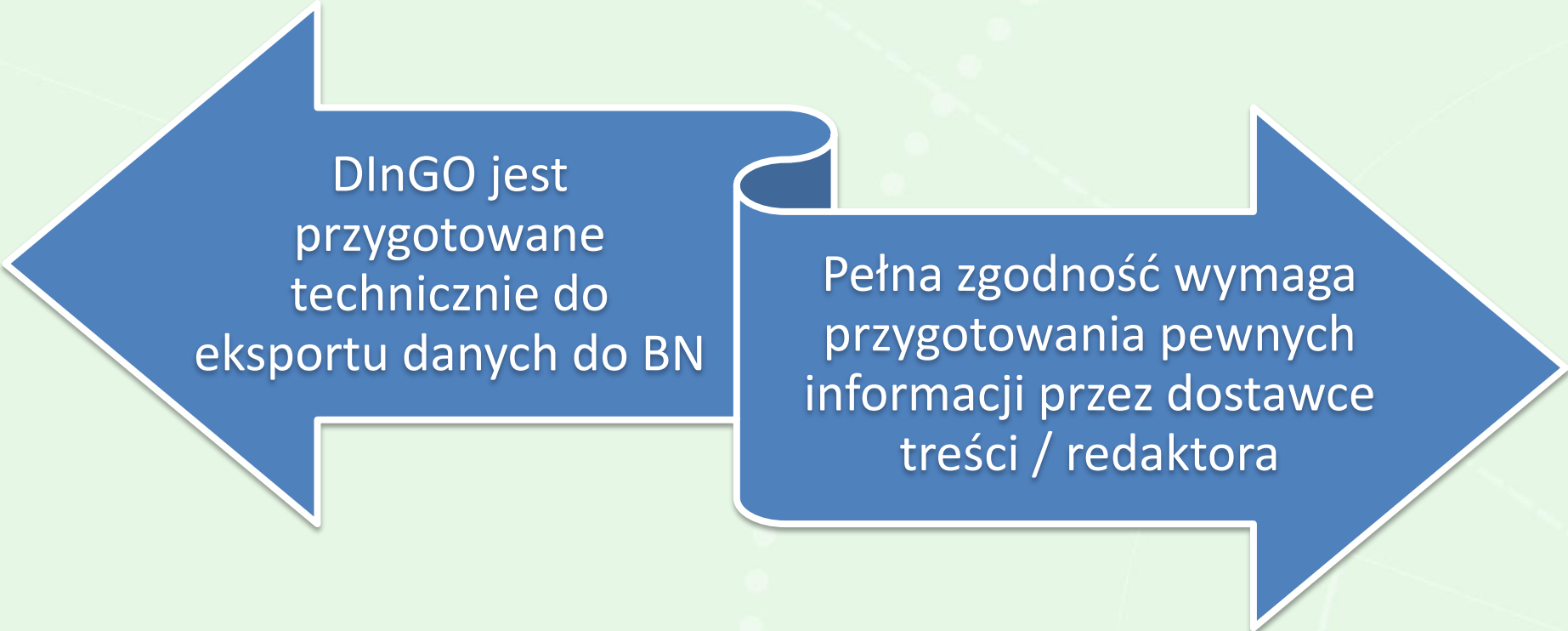
- dArceo wykorzystuje PREMIS do zapisywania automatycznego opisu zdarzeń z cyklu życia obiektu, które miały miejsce od momentu archiwizacji w dArceo. Ewentualnie można skonfigurować system dLab tak, by pozwalał wprowadzać dodatkowych tego rodzaju informacji lub wykorzystywał wartość określonych pól opisowych z systemu dLibra, a następnie by przekazywał je do dArceo. To pozwoli na zachowanie informacji o proveniencji i cyklu życia z okresu przed archiwizacją obiektu.

33. Metadane techniczne na temat struktury stron wytworzone w procesie OCR (o ile istnieją) powinny być zapisane zgodnie ze standardem ALTO. Jeśli zostało wykonane OCR w celu uzyskania tekstowej zawartości obiektu, to jego wyniki zawierające tekstową postać zawartości cyfrowej w postaci XML w formacie ALTO mają zostać umieszczone w oddzielnych plikach (względem pliku METS obiektu) wskazanych w pliku METS obiektu w taki sposób, jak zawartość cyfrowa obiektu.

- Należałoby odpowiednio skonfigurować system dLab, by ekstrahował tego typu informacje z wyników procesu OCR, kowentrował je do ALTO i przekazywał je na etapie archiwizacji do dArceo.

DInGO: zgodność w zakresie wymagań BN

Wytyczne BN – podsumowanie

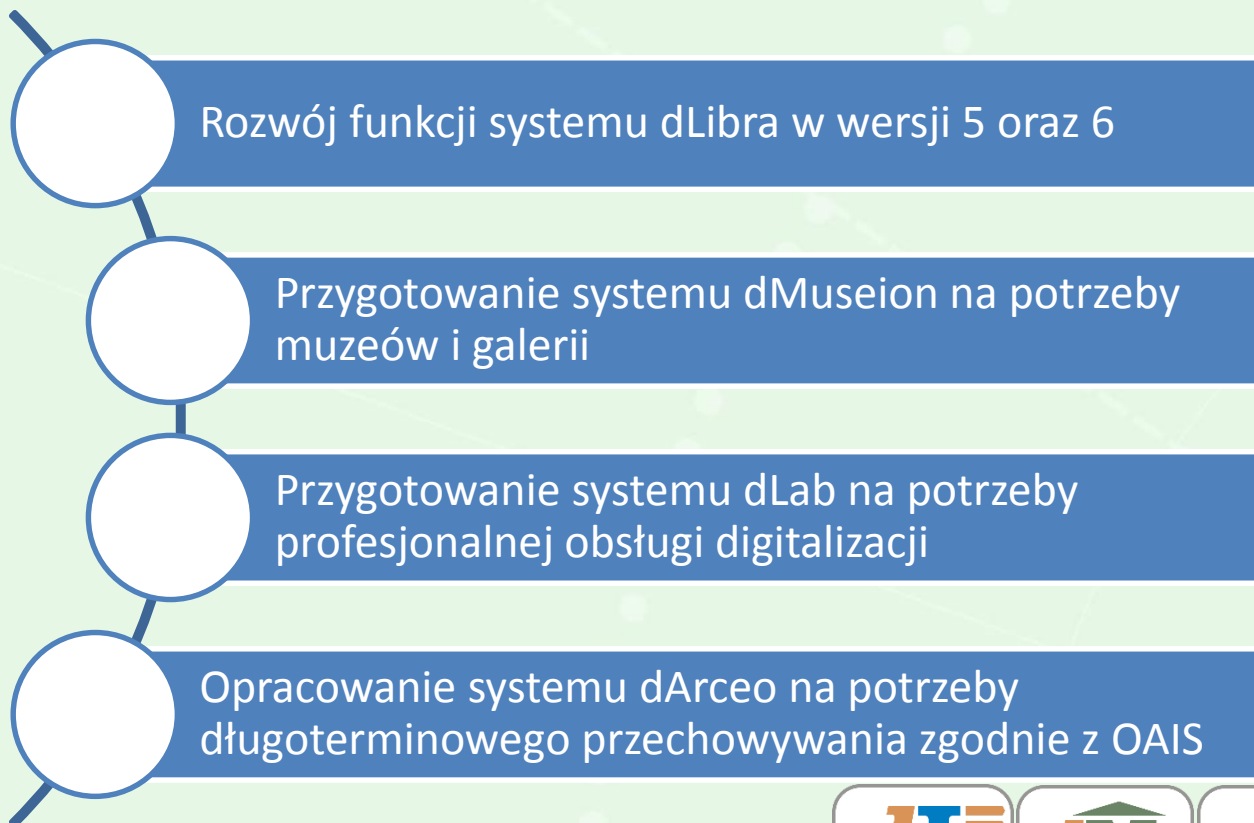
The diagram consists of two large, blue, stylized arrows pointing in opposite directions. The left arrow points towards the left, and the right arrow points towards the right. They are connected at their bases, forming a shape similar to a double-headed arrow or a bridge. The text is placed inside each arrow.

DInGO jest
przygotowane
technicznie do
eksportu danych do BN

Pełna zgodność wymaga
przygotowania pewnych
informacji przez dostawcę
treści / redaktora

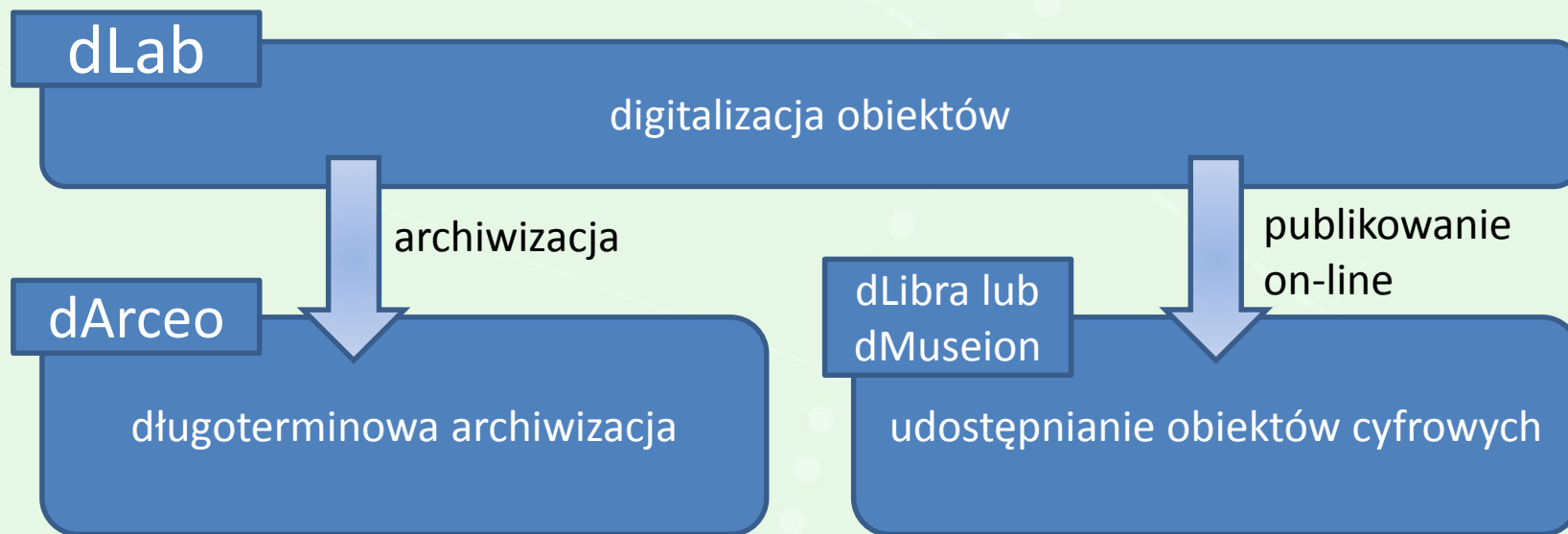
Podsumowanie

Prace w zakresie pakietu DInGO 2011-2015



Podsumowanie

Wzorcowe wykorzystanie pakietu DInGO (dMuseion/dLibra, dArceo, dLab)





Poznańskie Centrum Superkomputerowo - Sieciowe

afiliowane przy Instytucie Chemii Bioorganicznej PAN,

ul. Noskowskiego 12/14, 61-704 Poznań,
tel : (+48 61) 858-20-00, fax: (+48 61) 852-59-54,
e-mail: office@man.poznan.pl, <http://www.pcass.pl>



Dalszy rozwój pakietu DInGO

Tomasz Parkoła

Plany na najbliższe wydania DInGO

Otwarcie kodu źródłowego dLibra repozytorium i dArceo

Model rozwoju DInGO

Wieloformatowość w systemie dMuseion i dLibra

Funkcje dla repozytoriów cyfrowych w systemie dLibra

Plany na najbliższe wydania DInGO

Otwarcie kodu źródłowego systemu dLibra repozytorium i dArceo

Kod źródłowy zostanie udostępniony na jednej z licencji FLOSS (GNU GPL lub podobnej)

PCSS będzie właścicielem projektu i będzie nadal utrzymywał/rozwijał poszczególne komponenty w nowym modelu rozwoju

Zachęcamy wszystkich do współpracy i udziale w rozwoju wspólnego oprogramowania

Plany na najbliższe wydania DInGO

Model rozwoju systemu DInGO

Rozwój skokowy

Bezpośrednie finansowanie potrzebnych funkcji (jedna instytucja lub kilka instytucji wspierają finansowo wprowadzenie funkcji do systemu), np. poprzez mechanizmy projektowe

- Rola PCSS: partner lub podwykonawca
- Preferowana rola: partner (bardziej elastyczne możliwości w kontekście zakresu i efektów prac)
- Niezależnie od roli: chętnie wesprzemy przy definiowaniu technicznych założeń projektu (oprogramowanie i infrastruktura sprzętowa i chmurowa)

Rozwój ciągły

Finansowanie utrzymania i rozwoju systemu przez instytucje korzystające z oprogramowania w ramach opłacania usługi pomocy technicznej

- Bezpośrednia e-mailowa pomoc techniczna tylko dla instytucji z wykupioną usługą
- FBC Q&A dla instytucji bez wykupionej usługi
- Ułatwienia na poziomie umów (automatyczne przedłużenie umowy)
- Specjalne oferty dla instytucji z wykupioną usługą

Plany na najbliższe wydania DInGO

WielofORMATOWOŚĆ obiektów w systemie dMuseion (i dLibra)

Funkcja pojawi się w dMuseion dzięki współpracy z Muzeum Narodowym w Warszawie



Możliwość podłączenia kilku różnych formatów do tworzonego obiektu, np. JPG, MP3, 3D



Możliwość podłączenia nowych formatów do istniejącego obiektu, np. dodanie reprezentacji PDF oraz ePub do obiektu DjVu



??? Automatyczne generowanie nowych formatów i podłączanie ich pod obiekt, np. utworzenie z DjVu plików PDF ???

Plany na najbliższe wydania DInGO

Funkcje dla repozytoriów cyfrowych w systemie dLibra

Połączenie z bazą wiedzy na temat uczelni (naukowcy, projekty, wydarzenia, struktura organizacyjna, artykuły, publikacje)

- Przeglądanie publikacji wybranego autora
- Przegląd dziedzin i ekspertyzy grupy badawczej
- Wizualizacje powiązań

Dedykowane funkcje dla repozytoriów cyfrowych

- Współpraca z systemami sprawozdawczymi w ramach POL-on
- Propozycja uzupełnienia schematu metadanych o nowe pola

Plany na najbliższe wydania DInGO

Formaty obiektów cyfrowych

Będziemy uwzględniać
wnioski z dzisiejszej sesji
panelowej dot. formatów

Plany na najbliższe wydania DInGO

Co dalej?





Poznańskie Centrum Superkomputerowo - Sieciowe

afiliowane przy Instytucie Chemii Bioorganicznej PAN,

ul. Noskowskiego 12/14, 61-704 Poznań,
tel : (+48 61) 858-20-00, fax: (+48 61) 852-59-54,
e-mail: office@man.poznan.pl, <http://www.pcass.pl>