

# Metadane w procesie ciągłej digitalizacji dokumentów papierowych



Grzegorz Bednarek  
Agnieszka Wróbel

grzegorz@djvu.pl  
wrobel.a@uw.edu.pl

# Metadane w procesie digitalizacji

Katalog NUKAT



Gromadzenie i  
wprowadzanie  
informacji o obiektach



Katalog OPAC



Gromadzenie i  
wprowadzanie  
informacji o obiektach

Baza danych  
zawierająca metadane  
opisujące obiekty

Wykorzystanie metadanych

1. Przeszukiwanie zasobu, lokalizowanie obiektu cyfrowego (np. za pośrednictwem dLibry),
2. Udostępnianie informacji o oryginale obiektu i ew. jego kopii cyfrowej,
3. Zarządzanie archiwizowanymi plikami,
4. Intuicyjny i naturalny sposób zapoznawania się plikami obiektów,
5. Wsparcie dla kolejnych projektów digitalizacji,
6. Weryfikowanie spójności i kompletności repozytorium cyfrowego,
7. Dysponowanie środkami umożliwiającymi jednoznaczne potwierdzenie autentyczności i wiarygodności obiektów,
8. Dla czytelnika - udokumentowana postać obiektu cyfrowego.

# Schemat digitalizacji obiektów bibliotecznych







# Budowa tekstowego pliku „metadanych METS”

## Plik metadanych METS

- Nagłówek pliku, część opisowa

- Metadane deskryptywne  
zgodne ze standardem MODS

- Metadane techniczne zgodne  
ze standardem MIX

- Metadane administracyjne  
zgodne ze standardem PREMIS

- Sekcja dotycząca plików obiektu

- Struktura fizyczna obiektu

- Struktura logiczna obiektu

```
<mets:div ORDER="15" LABEL="Strona nr 15" TYPE="page" ID="DIV0015">
  <mets:fptr FILEID="FID000F_MF"/>
  <mets:fptr FILEID="FID000F_RC"/>
  <mets:fptr FILEID="FID000F_RT"/>
</mets:div>
<mets:div ORDER="16" LABEL="Strona nr 16" TYPE="page" ID="DIV0016">
  <mets:fptr FILEID="FID0010_MF"/>
  <mets:fptr FILEID="FID0010_RC"/>
  <mets:fptr FILEID="FID0010_RT"/>
</mets:div>
</mets:div>
/mets:structMap>
mets:structMap ID="STRUCTURE_02" LABEL="Struktura logiczna publikacji." TYPE="logical
<mets:div LABEL="Ilustrowany Kurjer Codzienny. 1925, nr 45 (14 II)" DMDID="DM_MODS
DM_DC" TYPE="text" ID="METS_02">
  <mets:div ID="Article_0001" TYPE="article" ADMID="AdM0001_RT">
    <mets:div ID="TextBlock_01" TYPE="article-part" ORDER="1">
      <mets:fptr>
        <mets:area FILEID="FID0001_MF" SHAPE="RECT" COORDS="155,166,2788,4624"/>
      </mets:fptr>
      <mets:fptr>
        <mets:area FILEID="FID0001_RT" BETYPE="IDREF" BEGIN="BLOK_01"/>
      </mets:fptr>
    </mets:div>
  </mets:div>
```

Rozpoznany tekst zgodnie  
ze standardem ALTO<sub>1</sub>

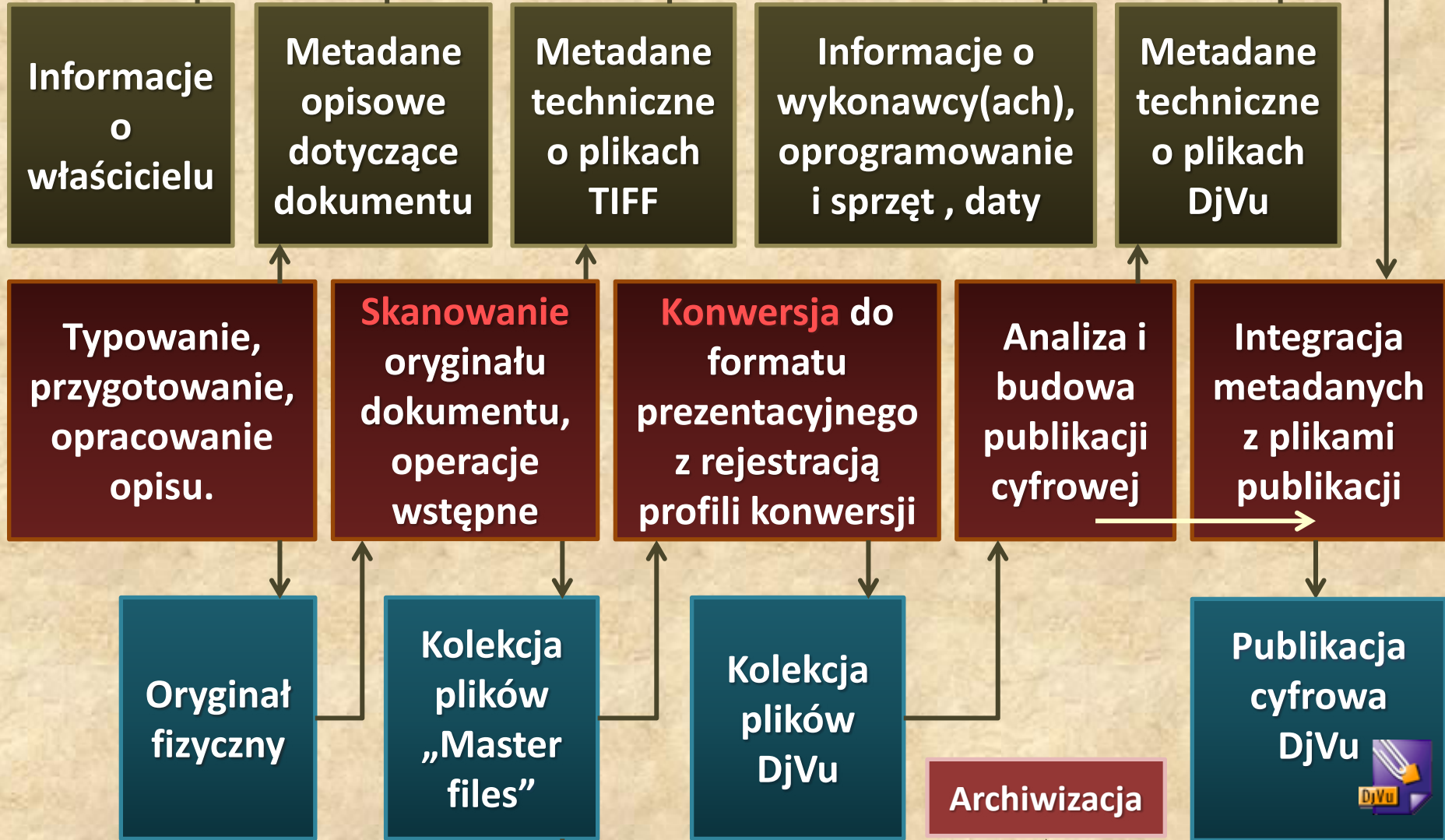
Rozpoznany tekst zgodnie  
ze standardem ALTO<sub>2</sub>

...

Rozpoznany tekst zgodnie  
ze standardem ALTO<sub>last</sub>

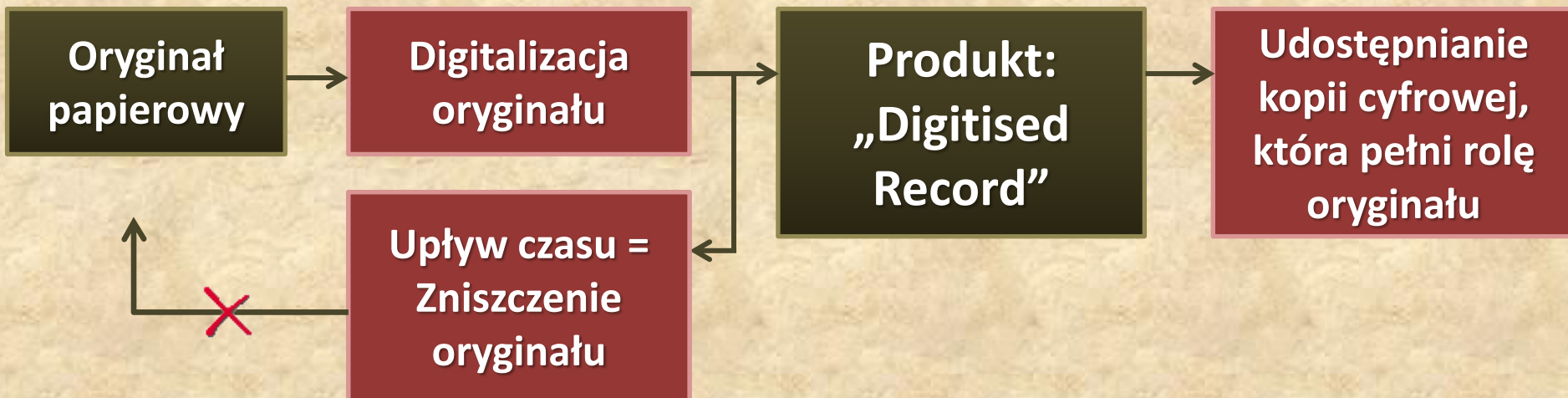
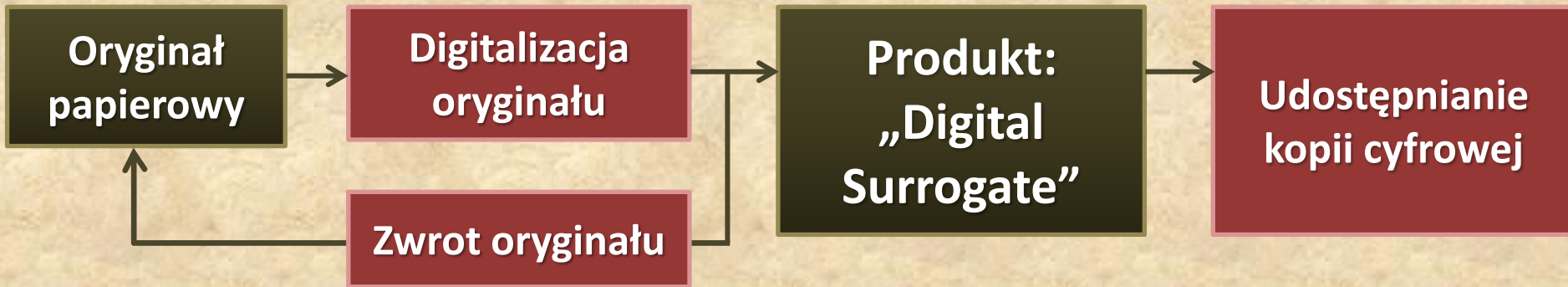
# Schemat digitalizacji obiektów bibliotecznych

METS



wariant 3

# Produkty digitalizacji „Digital surrogate” oraz „Digitised record”





# Wewnętrzne metadane uniwersalne

Plik Edycja Widok Narzędzia Pomoc

Organizuj Otwórz Pokaż slajdów Drukuj Poczta e-mail Nagraj Nowy folder

Source

- DjVu
  - Dziennik\_dla\_Wszystkich\_An
  - Goniec\_Poranny\_i\_Wieczorn
  - Kurjer\_Swiateczny\_tygodnik
  - Varsavska\_Guberskia\_Vedc
- DjVu - Indirect
- DocAnnotations
  - Dziennik\_dla\_Wszystkich\_An
  - Goniec\_Poranny\_i\_Wieczorn
  - Kurjer\_Swiateczny\_tygodnik
  - Varsavska\_Guberskia\_Vedc
- Indirect
  - Dziennik\_dla\_Wszystkich\_An
  - Goniec\_Poranny\_i\_Wieczorn
  - Kurjer\_Swiateczny\_tygodnik
  - Varsavska\_Guberskia\_Vedc
- JPG
  - Dziennik\_dla\_Wszystkich\_An
- PageMetaInfo
- pdf
  - Dziennik\_dla\_Wszystkich\_An
  - Goniec\_Poranny\_i\_Wieczorn
  - Kurjer\_Swiateczny\_tygodnik
  - Varsavska\_Guberskia\_Vedc
- Tiff
  - Dziennik\_dla\_Wszystkich\_An
  - Goniec\_Poranny\_i\_Wieczorn
  - Kurjer\_Swiateczny\_tygodnik
  - Varsavska\_Guberskia\_Vedc

Nazwa	Rozmiar	Typ	Tytuł	Tagi	Temat	Komentarze
Bundled 1887.djvu	2 322 KB	DjVu Docum...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki prezentacyjne; ...		ublifikacja poc...
Bundled 1904.djvu	1 932 KB	DjVu Docum...	Varšavskii Guberskii Vedomosti; Варшавскія Губернскія В...	biblioteka cyfrowa; ebuw; pliki prezentacyjne; ...		ublifikacja poc...
Bundled 1909.djvu	3 124 KB	DjVu Docum...	Goniec Poranny i Goniec Wieczorny - pismo codzienne / [re...	biblioteka cyfrowa; ebuw; pliki prezentacyjne; ...		ublifikacja poc...
Bundled 1919.djvu	794 KB	DjVu Docum...	Kurjer Świąteczny : tygodnik polityczny, literacki i humorysty...	biblioteka cyfrowa; ebuw; pliki prezentacyjne; ...		ublifikacja poc...
2863973.jpg	5 193 KB	IrfanView JPG...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...		plik skanu poch...
2863973a.jpg	5 271 KB	IrfanView JPG...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...		plik skanu poch...
2863973b.jpg	5 370 KB	IrfanView JPG...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...		plik skanu poch...
2863974.jpg	5 629 KB	IrfanView JPG...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...		plik skanu poch...
2863975.jpg	5 306 KB	IrfanView JPG...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...		plik skanu poch...
2863976.jpg	5 354 KB	IrfanView JPG...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...		plik skanu poch...
2863977.jpg	4 995 KB	IrfanView JPG...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...		plik skanu poch...
2863978.jpg	5 111 KB	IrfanView JPG...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...		plik skanu poch...
2863973.tif	42 485 KB	IrfanView TIF ...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...		plik skanu poch...
2863973a.tif	40 832 KB	IrfanView TIF ...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...	19 w.; 20 w.; dzienniki polskie; ogłoszenia...	Plik skanu poch...
2863973b.tif	42 557 KB	IrfanView TIF ...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...	19 w.; 20 w.; dzienniki polskie; ogłoszenia...	Plik skanu poch...
2863974.tif	42 752 KB	IrfanView TIF ...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...	19 w.; 20 w.; dzienniki polskie; ogłoszenia...	Plik skanu poch...
2863975.tif	41 966 KB	IrfanView TIF ...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...	19 w.; 20 w.; dzienniki polskie; ogłoszenia...	Plik skanu poch...
2863976.tif	42 714 KB	IrfanView TIF ...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...	19 w.; 20 w.; dzienniki polskie; ogłoszenia...	Plik skanu poch...
2863977.tif	42 583 KB	IrfanView TIF ...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...	19 w.; 20 w.; dzienniki polskie; ogłoszenia...	Plik skanu poch...
2863978.tif	41 624 KB	IrfanView TIF ...	Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd...	biblioteka cyfrowa; ebuw; pliki macierzyste; di...	19 w.; 20 w.; dzienniki polskie; ogłoszenia...	Plik skanu poch...

2863978.tif  
IrfanView TIF File

Data wykonania: 2012-04-19 01:32

Tagi: biblioteka cyfrowa; ebuw; pliki macierzyste; digitisation; digital library

Klasyfikacja: ☆☆☆☆

Wymiary: 3156 x 4501

Rozmiar: 40,6 MB

Tytuł: Dziennik dla Wszystkich i Anonsowy / [red. H. Perzyński; wyd. J. Śliwowski].

Autorzy: Gama s.c, Katowice

Komentarze: Plik skanu pochodzi z zasobów Biblioteki Uniwersyteckiej w Warszawie


Producent aparatu... Dodaj tekst

Model aparatu fot... Dodaj nazwę

Temat: 19 w.; 20 w.; dzienniki polskie; ogłoszenia drobne; Polska; Warszawa

Data utworzenia: 2012-10-26 22:44

Data modyfikacji: 2013-01-08 21:00





# Metadane w procesie ciągłej digitalizacji dokumentów papierowych



*Dziękujemy za uwagę*



Grzegorz Bednarek  
grzegorz@djvu.pl  
GB Soft sp.j.,  
Zabrze

Agnieszka Wróbel  
wrobel.a@uw.edu.pl  
Biblioteka Uniwersytecka w  
Warszawie