

# Organizacja informacji o obiektach cyfrowych z wykorzystaniem standardu metadanych METS.

Pliki metadanych zewnętrznych

---

Grzegorz Bednarek, GB Soft sp.j.  
21/28 listopada 2017

# Cechy charakterystyczne schematu METS



## Schemat metadanych METS :

- Jest standardem otwartym,
- Nie jest zastrzeżony,
- Został opracowany, jest rozwijany oraz dostarczony przez społeczność środowiska bibliotekarzy,
- Relatywnie prosty,
- Rozszerzalny,
- Posiada budowę modułową,
- Jest schematem wyrażonym za pomocą języka XML.

## Podstawowe właściwości plików METS

METS został zaprojektowany w celu tworzenia dokumentów XML, które określają :

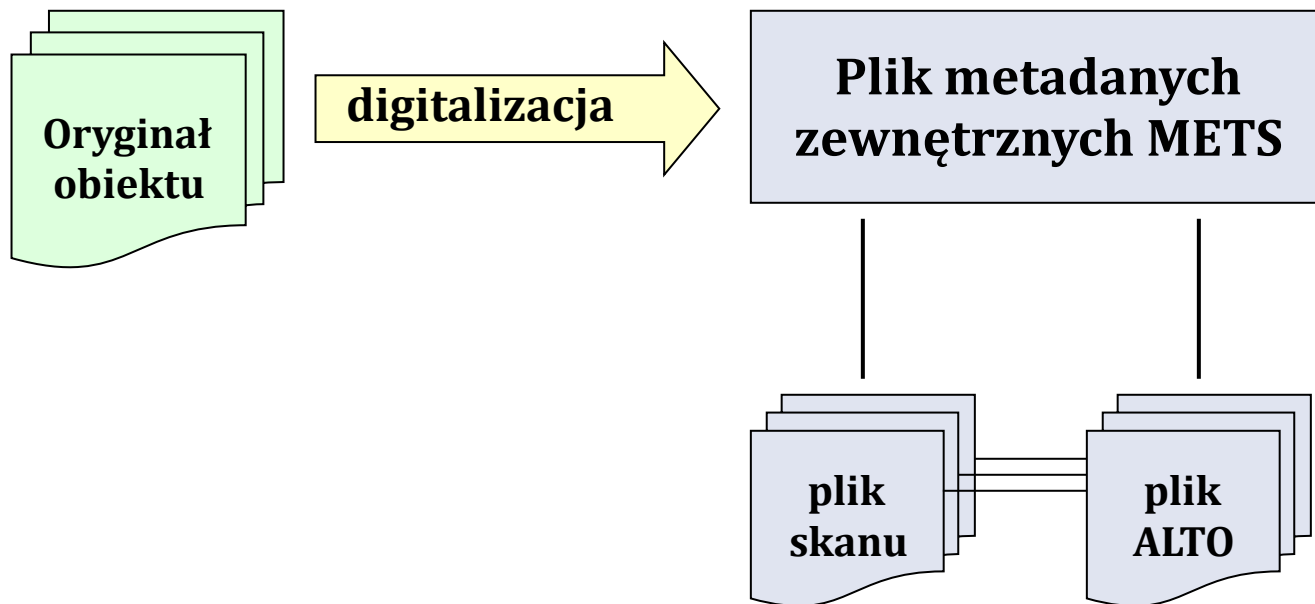
- Hierarchiczną strukturę obiektów cyfrowych biblioteki, muzeum lub archiwum,
- Nazwy i lokalizacje plików, które dotyczą tych obiektów oraz powiązanych z nimi metadanych,
- Zawartość i sposób wykonania obiektów cyfrowych,
- Sposób przekazywania do archiwów „long-term” obiektów cyfrowych pod postacią pakietów transferowych.

## Obszary zastosowań plików METS

Dzięki plikom metadanych zewnętrznych „METS”, możliwym jest :

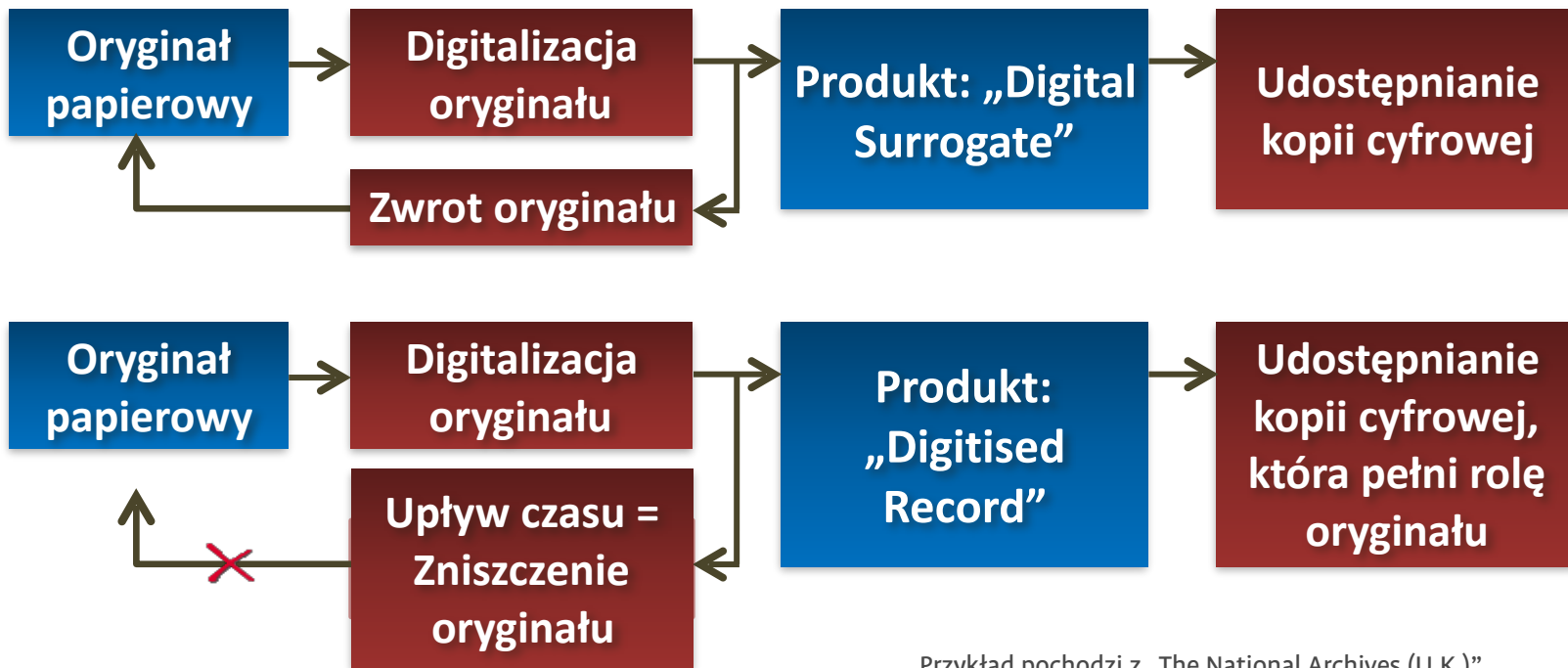
- wspólne „opakowanie” kolejnych plików obiektów cyfrowych i powiązanych z nimi metadanych,
- przekazywanie zasobów cyfrowych pomiędzy repozytoriami o zróżnicowanej wzajemnie organizacji,
- potwierdzenia przeprowadzenia procesu digitalizacji w taki sposób, który gwarantuje, że obiekt cyfrowy może pełnić rolę kopii - bądź wprost - oryginału obiektu bibliotecznego.

# Postać produktu digitalizacji obiektów bibliotecznych i muzealnych



wspólne „opakowanie” kolejnych plików obiektu cyfrowego

## Możliwe scenariusze realizacji projektu digitalizacyjnego



Przykład pochodzi z „The National Archives (U.K.)”

przeprowadzenie procesu digitalizacji w taki sposób, który gwarantuje, że powstały obiekt cyfrowy może pełnić rolę kopii - bądź wprost - oryginału obiektu bibliotecznego

## Dane i informacje umożliwiające wykonywanie kolejnych plików METS podczas realizacji projektu digitalizacyjnego

- Metadane deskryptywne dotyczące zarówno oryginału obiektu jak i jego kopii cyfrowej,
- inne informacje opisowe dotyczące oryginału obiektu,
- Informacje dotyczące projektu digitalizacji takie jak nazwy instytucji i/lub dane osób odpowiedzialnych za koordynację działań, wykonanie plików cyfrowych, ...
- Zestawienie operacji dotyczących okresu życia obiektu cyfrowego, które zostaną umieszczone w plikach METS (opis obiektów danych, operacji którym je poddano od momentu ich powstania oraz biorących w nich udział agentów).

# Obiekt cyfrowy

## Obiekt cyfrowy

### Grupowy obiekt danych - (1)

Obiekt danych  
001.tif

Obiekt danych  
001.jpg

...

Obiekt danych  
001.xml

### Grupowy obiekt danych - (2)

Obiekt danych  
002.tif

Obiekt danych  
002.jpg

...

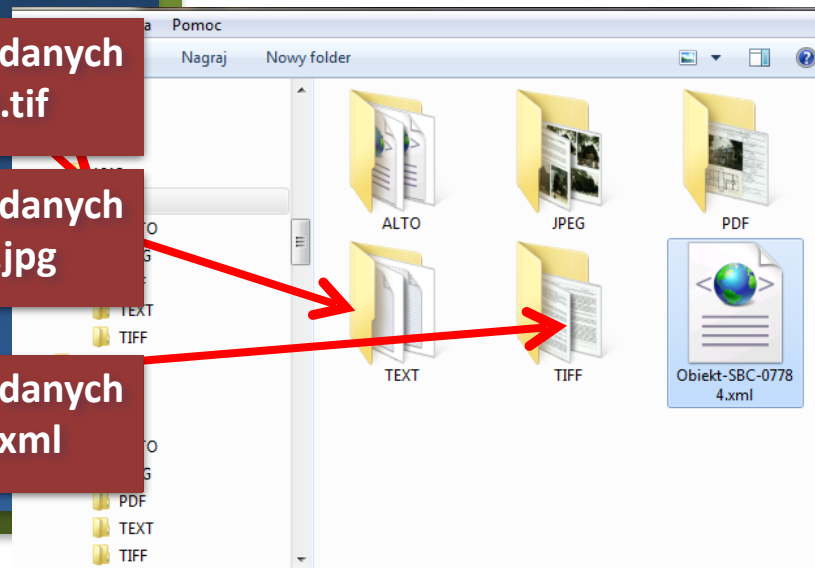
Obiekt danych  
002.xml

### Grupowy obiekt danych - (n)

Obiekt danych  
024.tif

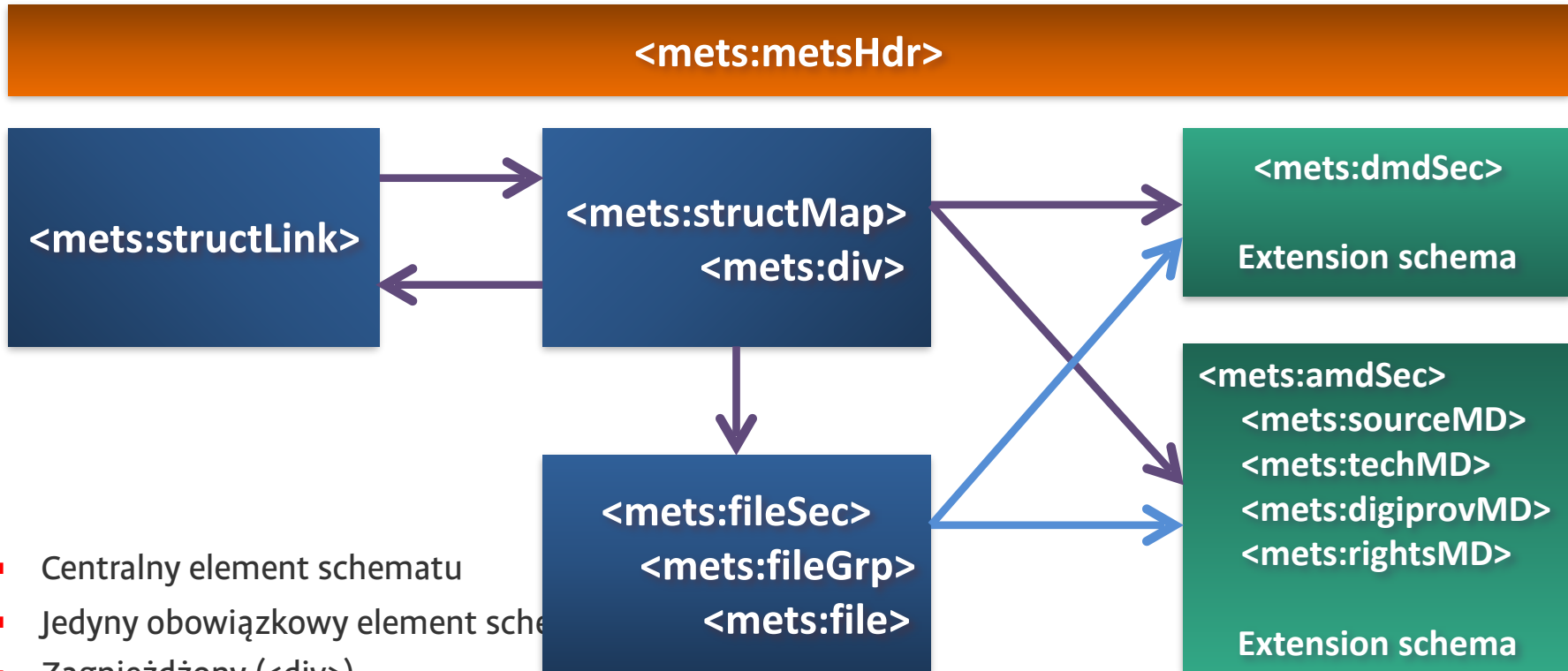
Obiekt danych  
024.jpg

Obiekt danych  
024.xml





# Elementy wysokiego poziomu plików METS



- Centralny element schematu
- Jedyny obowiązkowy element schematu
- Zagnieżdżony (<div>)
- Różne odmiany, wskazywane atrybutem TYPE="..."

# Kategorie metadanych

- Metadane strukturalne,
- Metadane deskryptywne (opisowe),
- Metadane administracyjno-techniczne :
  - Metadane administracyjne,
  - Metadane techniczne,
  - Metadane konserwatorskie,
  - Metadane praw i uprawnień udostępniania obiektu cyfrowego,
- Metadane rozpoznanego tekstu obiektu cyfrowego.

## Zalecane standardy metadanych (schematy metadanych).

- METS - Metadata Encoding and Transmission Standard
- MODS - Metadata Object Description Schema
- MARCxml - Machine-Readable Cataloging
- DCMES - Dublin Core Metadata Element Set
- PREMIS - Preservation Metadata: Implementation Strategies
- MIX - NISO Metadata for Images in XML (NISO MIX)
- textMD - Technical Metadata for Text
- PAIS - Producer-Archive Interface Specification
- ALTO - The Analyzed Layout and Text Object XML Schema

# Schematy rozszerzeń (extension schemas) w plikach METS

Dla standardu METS, zatwierdzonymi schematami rozszerzeń są :

- Metadane deskryptywne

MARCxml, MODS, DCMES (Dublin Core),

- Metadane techniczne

MIX, textMD,

- Metadane konserwatorskie

PREMIS Data Dictionary.

Innymi, a często wykorzystywanymi w plikach METS schematami metadanych są :

METSRights, PAIS, ProcessMD, audioMD i videoMD.

# Budowa pliku METS

## Plik METS

Deklaracje lokalizacji definicji użytych schematów xml (Name Space)

Nagłówek informacyjny (Header, `<mets:metsHdr>`)

Metadane deskryptywne (Descriptive Metadata Section, `<mets:dmdSec>`)

Metadane administracyjno-techniczne (Administrative Metadata Section, `<mets:amdSec>` :  
`<mets:sourceMD>`, `<mets:techMD>`, `<mets:digiprovMD>`, `<mets:rightsMD>`)

Zestawienie plików współtworzących obiekt cyfrowy (File Section, `<mets:fileSec>`)

Struktura(y) plików – fizyczna, logiczna, obiektu transferowego (Structural Map, `<mets:structMap>`)

Rzadko wykorzystywane sekcje schematu (`<mets:structLink>` oraz `<mets:behaviorSec>`)

Metadane rozpoznanego tekstu

# Wartości elementów pliku xml, kontenery danych

## Wartość prosta, element schematu

```
<accessCondition lang="pol">domena publiczna</accessCondition>  
<element atrybut="wartość atrybutu">wartość elementu</element>
```

## Wartość złożona, kontener danych

Przykład: „Kramorz, Stefan. Red.”

```
<name type="personal">  
  <namePart type="family">Kramorz</namePart>  
  <namePart type="given">Stefan</namePart>  
  <role>  
    <roleTerm type="code" lang="pol">Red.</roleTerm>  
    <roleTerm type="code" authority="marcrelator">red</roleTerm>  
    <roleTerm type="text" authority="marcrelator">Redaktor</roleTerm>  
  </role>  
  <displayForm>Stefan Kramorz. Red.</displayForm>  
</name>
```

# Elementy wysokiego poziomu schematu METS

Element główny <mets:mets> tworzy kontener dla przechowywanych metadanych

~~<mets:mets xmlns:mets="...~~

<mets:metsHdr> ... </mets:metsHdr>

<mets:dmdSec> ... </mets:dmdSec>

<mets:amdSec>

<mets:sourceMD> ... </mets:sourceMD>

<mets:techMD> ... </mets:techMD>

<mets:digiprovMD> ... </mets:digiprovMD>

<mets:rightsMD> ... </mets:rightsMD>

</mets:amdSec>

<mets:fileSec> ... </mets:fileSec>

<mets:structMap> ... </mets:structMap>

<mets:structLink> ... </mets:structLink>

</mets:mets>

# Określenie warunków i sposobu wykonywania plików METS

Postać elementów współtworzących plik METS, określana jest przed rozpoczęciem digitalizacji.

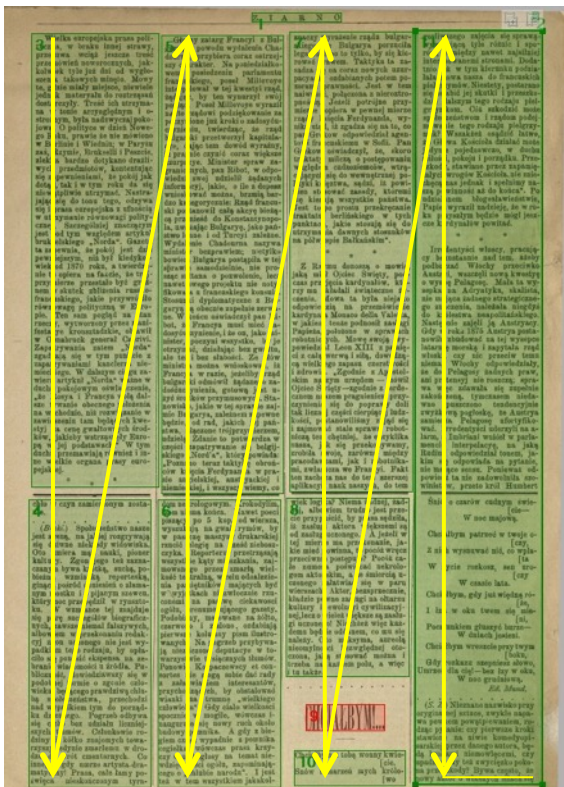
Podaje się m.in. :

- Typy struktur, za pomocą których przedstawione będą obiekty w plikach METS,
- Formuły w oparciu o które wykonane będą nazwy identyfikatorów obiektów danych i kolejnych sekcji w plikach METS,
- Informacje o użytych słownikach kontrolowanych,
- Optymalizacja sekcji metadanych technicznych,
- Sposób wykonania operacji rozpoznania tekstu – podstawowy, z uwzględnieniem układu graficznego, z uwzględnieniem układu graficzno–logicznego treści obiektu,
- Relacje pomiędzy metadanych opisowymi obiektów a nazwami folderów jego obiektów danych.

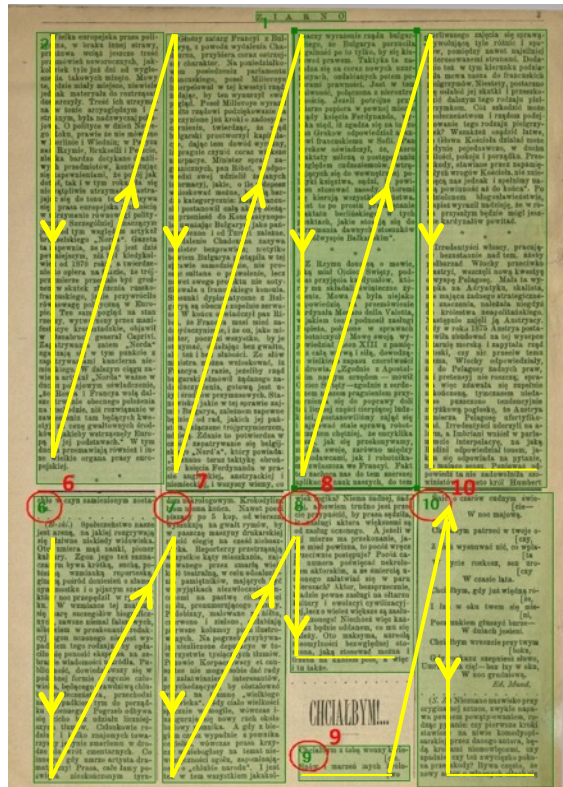


# Sposób wykonania operacji rozpoznania tekstu na stronie obiektu

## Rozpoznanie uwzględniające graficzny układ stron



## Rozpoznanie uwzględniające układ graficzno-logiczny stron



# Identyfikacja i odwoływanie do elementów schematu METS

Aby możliwym było odwołanie do pewnego elementu schematu, element taki musi posiadać unikalny w obrębie całego pliku METS identyfikator zdefiniowany jako **ID="..."**.

```
<mets:file ID="FID.054189_MF">  
<mets:dmdSec ID="DMD.DC">  
<mets:dmdSec ID="DMD.MODS.Cho1">  
<mets:digiprovMD ID="PMD.0064538_MF">
```

Odwołanie do sekcji metadanych wyrażonych za pomocą schematów rozszerzeń realizuje się za pomocą atrybutów **DMDID="..."** lub **ADMID="..."**.

```
... DMDID="DMD.MODS DMD.DC"  
... ADMID="CMD.001257_MF TMD.0054189_MF AMD.0054189_MF"
```

Odwołanie do pliku realizuje się za pomocą wskaźnika „file pointer” **FILEID="..."**.

```
<mets:fptr FILEID="FID.0054189_MF"/>
```

## Typowe struktury dokumentu cyfrowego

- Plakat, widokówka, zdjęcie, plakat teatralny, afisz :

physical \ stillImage

- Książka :

physical \ book \ page

physical \ book (\ chapter) \ page (\ paragraph)

- Inkunabuł współoprawny :

physical \ volume (\ book) \ page

physical \ volume \ book (\ chapter) (\ page)

- Czasopismo :

physical \ issue \ page

logical (\ yearbook) \ issue \ article (\ paragraph)

# Przykładowe struktury dokumentu cyfrowego

<!-- Fotografia, afisz -->

```
<mets:structMap ID="METS01" TYPE="physical">
  <mets:div ID="photography" TYPE="stillImage">
    <mets:fptr FILEID="FID.007231_MF"/>
    <mets:fptr FILEID="FID.007231_RC"/>
  </mets:div>
</mets:structMap>
```

<!-- Czasopismo -->

```
<mets:structMap ID="METS02" TYPE="logical">
  <mets:div ID="Periodical" TYPE="issue">
    <mets:div ID="TBlock1" TYPE="article">
      <mets:div ID="Pg_01" TYPE="paragraph">
        <mets:fptr><mets:area FILEID="FID.007231_MF" SHAPE="RECT" COORDS="219,2050,323,2105"/></mets:fptr>
        <mets:fptr><mets:area FILEID="FID.007231_RT" BETYPE="IDREF" BEGIN="BLOK_003"/></mets:fptr>
      </mets:div>
      ...
    </mets:div>
    ...
  </mets:div>
</mets:structMap>
```

<!-- Książka -->

```
<mets:structMap ID="METS01" TYPE="physical">
  <mets:div ID="BookPages" TYPE="book">
    <mets:div ID="Page_01" TYPE="page">
      <mets:fptr FILEID="FID.007231_MF"/>
      <mets:fptr FILEID="FID.007231_RC"/>
      <mets:fptr FILEID="FID.007231_RT"/>
    </mets:div>
    ...
  </mets:div>
</mets:structMap>
```

# Zachowanie informacji o użytych słownikach kontrolowanych

## Jakość metadanych

- **authority** nazwa/identyfikator użytego słownika kontrolowanego.
  - **authorityURI** adres użytego słownika kontrolowanego.
  - **valueURI** identyfikator rekordu z wybraną wartością z tego słownika.
- Dotyczą metadanych deskryptywnych wyrażonych za pomocą schematu MODS.
  - Są podstawą oceny jakości metadanych zawartych w tej części pliku METS.

```
<languageTerm type="text" authority="iso639-2b">Polish</languageTerm>
```

```
<languageTerm type="code" authority="iso639-2b">pol</languageTerm>
```

```
<frequency lang="eng" authority="marcfrequency">Weekly</frequency>
```

```
<frequency lang="pol" authority="local">Tygodnik</frequency>
```

```
<physicalLocation type="current" authority="marcorg">PIWaBN</physicalLocation>
```

```
<physicalLocation type="current" authority="lcsn" valueURI="https://lcn.loc.gov/n80014571" lang="pol">
```

```
  Biblioteka Narodowa w Warszawie</physicalLocation>
```

## Sposoby dołączania metadanych deskryptywnych

### Opakowanie dołączonego schematu (wrapping)

```
<mets:dmdSec ID="DMD.MODS">  
  <mets:mdWrap MDTYPE = "MODS"  
    MIMETYPE = "text/xml">  
    <mets:xmlData>  
      <mods:mods version="3.6">  
        ...  
      </mods:mods>  
    </mets:xmlData>  
  </mets:mdWrap>  
</mets:dmdSec>
```

### Odnosząc się do pliku schematu metadanych poza METS (reference)

```
<mets:dmdSec ID="DMD.MODS">  
  <mets:mdRef MDTYPE = "MODS"  
    LOCTYPE = "URL"  
    xlink:href =  
      "file://Books/MData/MODS-000663.xml"  
  />  
</mets:dmdSec>
```

## Sposoby dołączania metadanych administracyjnych

### Opakowanie dołączonego schematu (wrapping)

```
<mets:amdSec>
  <mets:techMD ID="TMD.053DFB_MF">
    <mets:mdWrap MDTYPE = "NISOIMG"
      MIMETYPE = "text/xml">
      <mets:xmlData>
        <mix:mix>
          ...
        </mix:mix>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:techMD>
</mets:amdSec>
```

### Odnosząc się do pliku schematu metadanych poza METS (reference)

```
<mets:amdSec>
  <mets:techMD ID="TMD.053DFB_MF">
    < mets:mdRef MDTYPE = "NISOIMG"
      LOCTYPE = "URL"
      xlink:href =
        "file://Books/MData/TechMD-000663.xml"
    />
  </mets:techMD>
</mets:amdSec>
```

Metadane techniczne opisujące plik graficzny (np. \*.tif, \*.jpg, \*.pdf, \*.djvu).

## Sposoby dołączania metadanych administracyjnych

### Opakowanie dołączonego schematu (wrapping)

```
<mets:amdSec>
  <mets:techMD ID="AMD.053DFD_RT">
    <mets:mdWrap MIMETYPE="text/xml"
      MDTYPE="PREMIS:OBJECT">
      <mets:xmlData>
        <premis:premis version="3.0">
          ...
        </premis:premis>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:techMD>
</mets:amdSec>
```

### Odnosząc się do pliku schematu metadanych poza METS (reference)

```
<mets:amdSec>
  <mets:techMD ID="AMD.053DFD_RT">
    <mets:mdRef MDTYPE="PREMIS:OBJECT"
      LOCTYPE="URL"
      xlink:href =
        "file://Books/MData/AltoMD-000663.xml"
    />
  </mets:techMD>
</mets:amdSec>
```

Metadane opisujące obiekt plikowy.





**BIBLIOTEKA  
NARODOWA**