

TBX

format danych terminologicznych

Jacek Tomaszczyk
Instytut Bibliotekoznawstwa
i Informacji Naukowej
Uniwersytet Śląski

Terminologia

1. **Terminologia jako dyscyplina naukowa.**
2. **Terminologia jako zbiór terminów (słownictwo specjalistyczne).**

Dane terminologiczne

Terminologię tworzą terminy, będące reprezentacją pojęć. Pojęcia są opisywane za pomocą **danych terminologicznych** (terminy, definicje pojęć, synonimy, ekwiwalenty w językach obcych, etc).

Dane terminologiczne są:

- gromadzone, przechowywane, przetwarzane w różnych systemach, wykorzystywane w wielu aplikacjach (np. bazy danych), zarządzane przez osoby prywatne, firmy lub instytucje.
- gromadzone w terminologicznych bazy danych, które zawierają różne zestawy danych terminologicznych, przechowując je w różnorodnych strukturach i formatach.

Metamodel

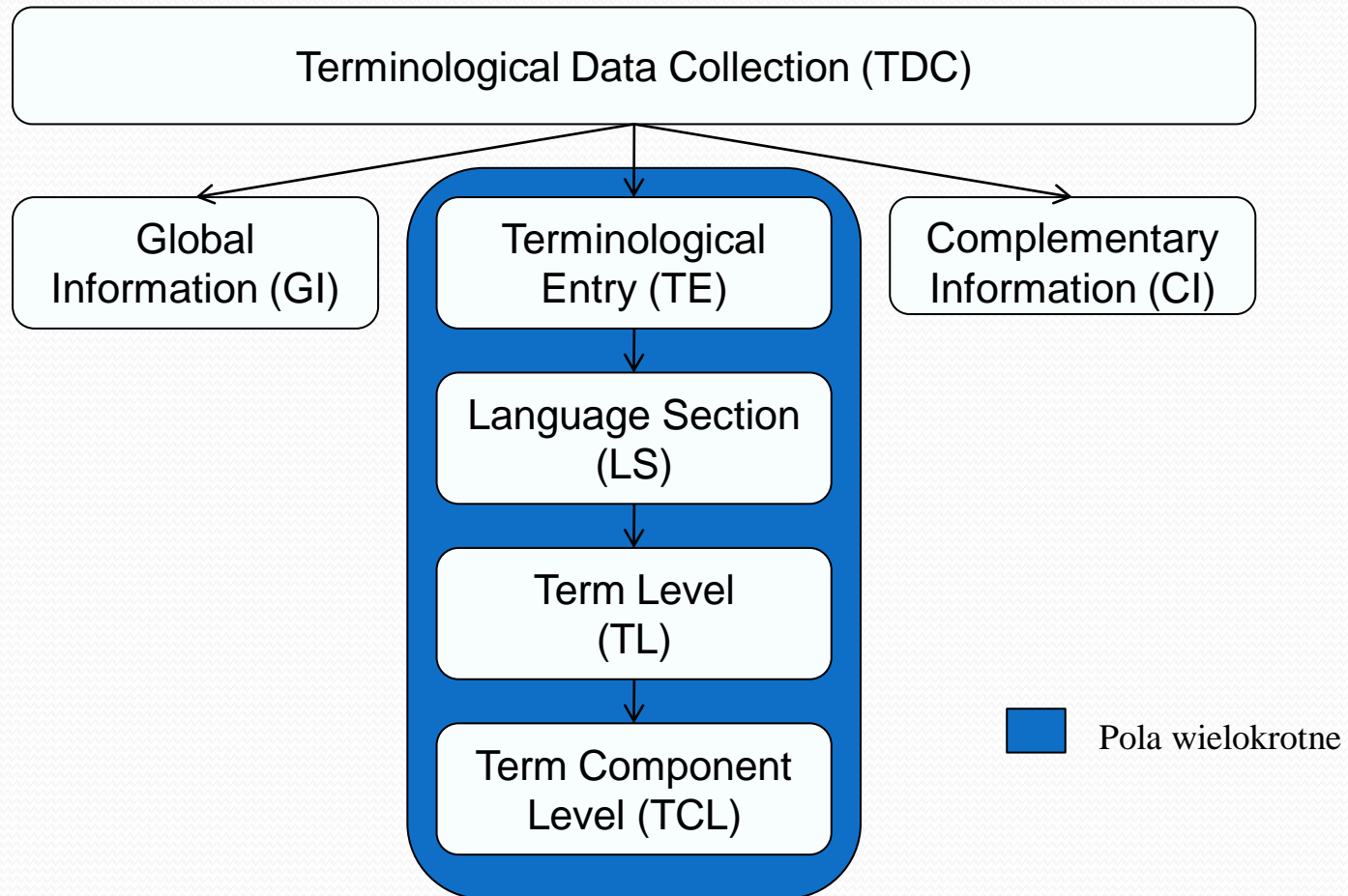
Dane terminologiczne są często współużytkowane przez różne systemy i przetwarzane przez aplikacje różnego typu za pomocą formatów pośrednich.

Aby ułatwić reprezentację, współużytkowanie i wymianę danych terminologicznych w zróżnicowanym środowisku komputerowym potrzebny jest **metamodel**, który nie opisuje konkretnego formatu, ale **jest nadrzędną strukturą wspólną dla wszystkich formatów.**

Terminological Markup Framework (TMF)

Metamodel (nadrzędna struktura danych terminologicznych)

źródło ISO 16642:2003



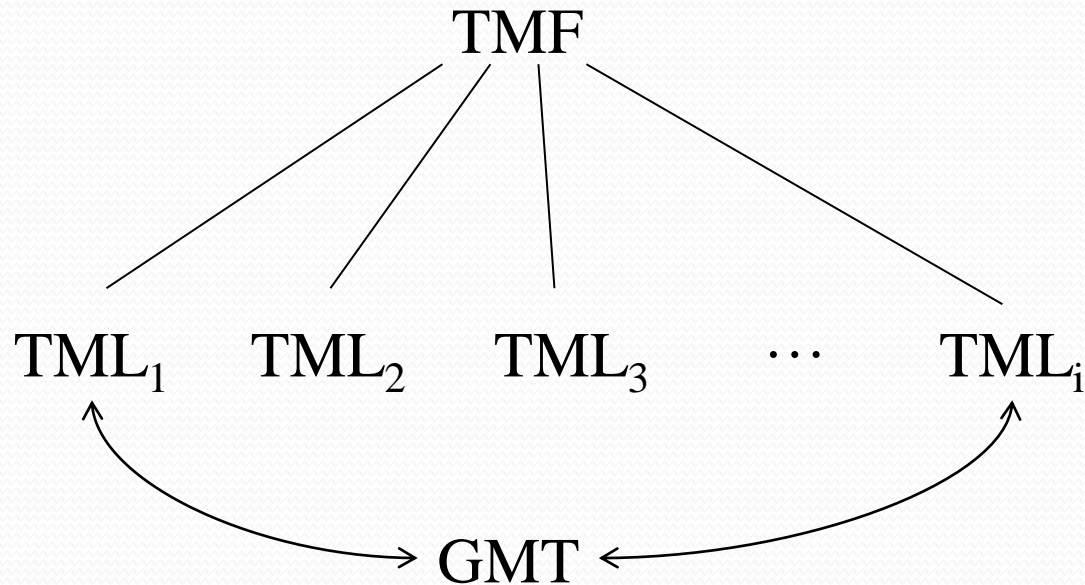
Terminological Markup Language (TML)

(język znaczników terminologicznych)

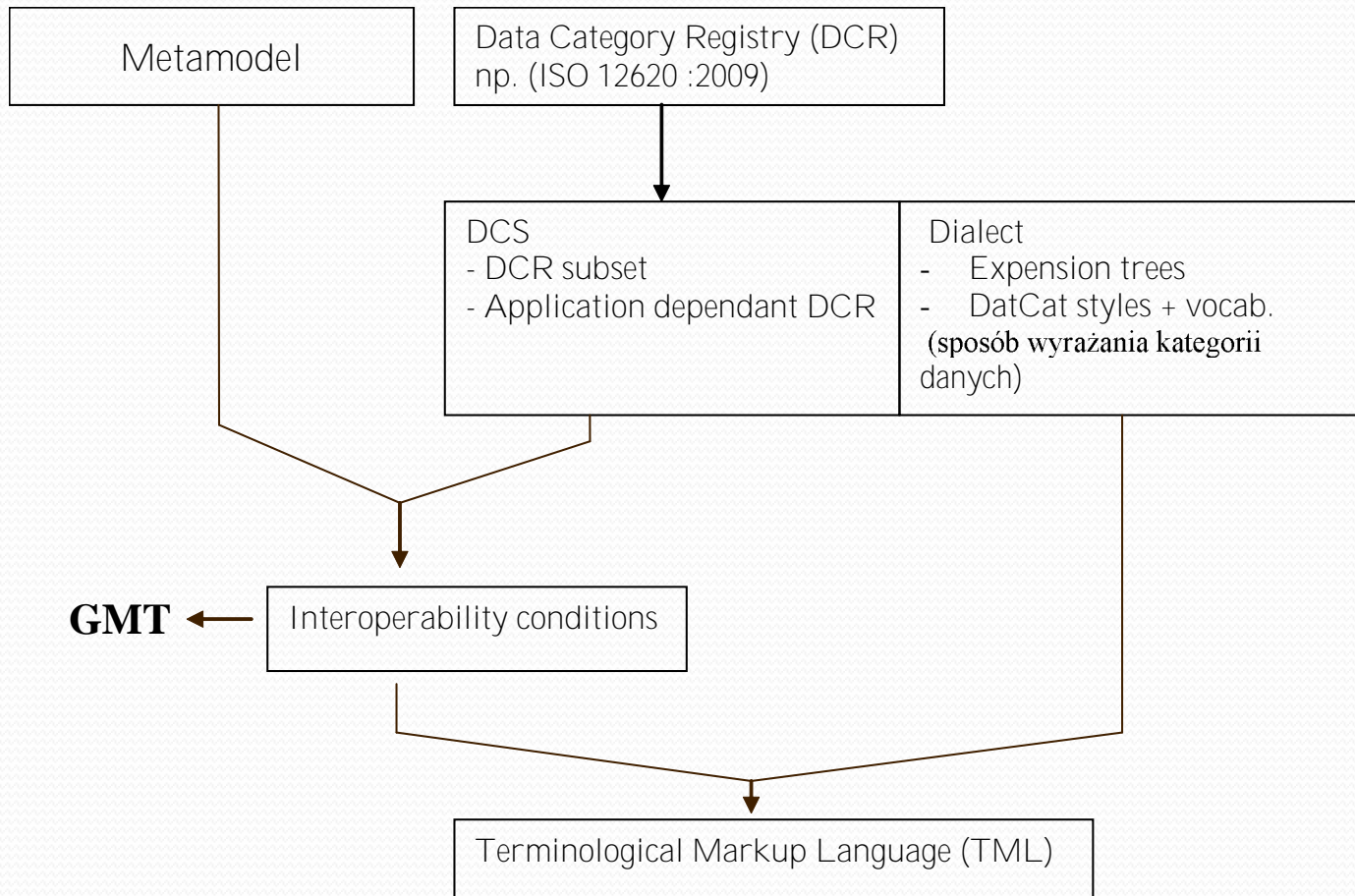
Implementacja (realizacja) metamodelu danych terminologicznych za pomocą języka XML (możliwe jest również użycie języków RDF i UML) nosi nazwę **języka znaczników terminologicznych** (terminological markup language, TML).

Generic Mapping Tool (GMT)

Metamodel może być realizowany przez wiele języków, dlatego musi istnieć narzędzie, które zapewni możliwość ich współdziałania. Tym narzędziem jest **GMT** (*ogólne narzędzie mapowania*), będące schematem XML, zdefiniowane w normie ISO 16642:2003



Definiowanie specyfikacji TML



Model TMF a TBX

Metamodel jest realizowany przez języki TML. Jednym z takich języków jest, zdefiniowany przez ISO 16642, MSC (MARTIF with Specified Constraints), a ten z kolei został użyty do zdefiniowania formatu **TBX** (TermBase eXchange Format), czyli **formatu wymiany danych terminologicznych** (między bazami terminologicznymi).

MARTIF = *Machine-Readable Terminology Interchange Format*, zorientowany na pojęcia, zawiera 150 kategorii danych

Zastosowanie TBX

- Format TBX ułatwia wymianę danych terminologicznych wśród użytkowników (np. tłumaczy, autorów dokumentacji) oraz między aplikacjami (np. programami wspomagającymi tłumaczenia: CAT) i systemami (np. systemami baz danych, w tym bankami terminologicznymi).
- Format służy więc wymianie informacji terminologicznych zarówno między ludźmi, jak i maszynami, dzięki czemu umożliwia przepływ danych terminologicznych w całym cyklu tworzenia informacji w obrębie firmy/instytucji i z zewnętrznymi dostawcami usług.
- Ułatwia integrację istniejących zasobów terminologicznych.

TBX a normy ISO

TBX został zdefiniowany w normie **ISO 30042:2008 Systems to manage terminology, knowledge and content -- TermBase eXchange (TBX)**. Norma została opracowana na podstawie 3 innych norm:

1. ISO 12200:1999 – Computer applications in terminology -- Machine-readable terminology interchange format (MARTIF) -- Negotiated interchange.
2. ISO 12620:1999 Computer applications in terminology -- Data categories.
3. ISO 16642:2003 Computer applications in terminology -- Terminological markup framework.

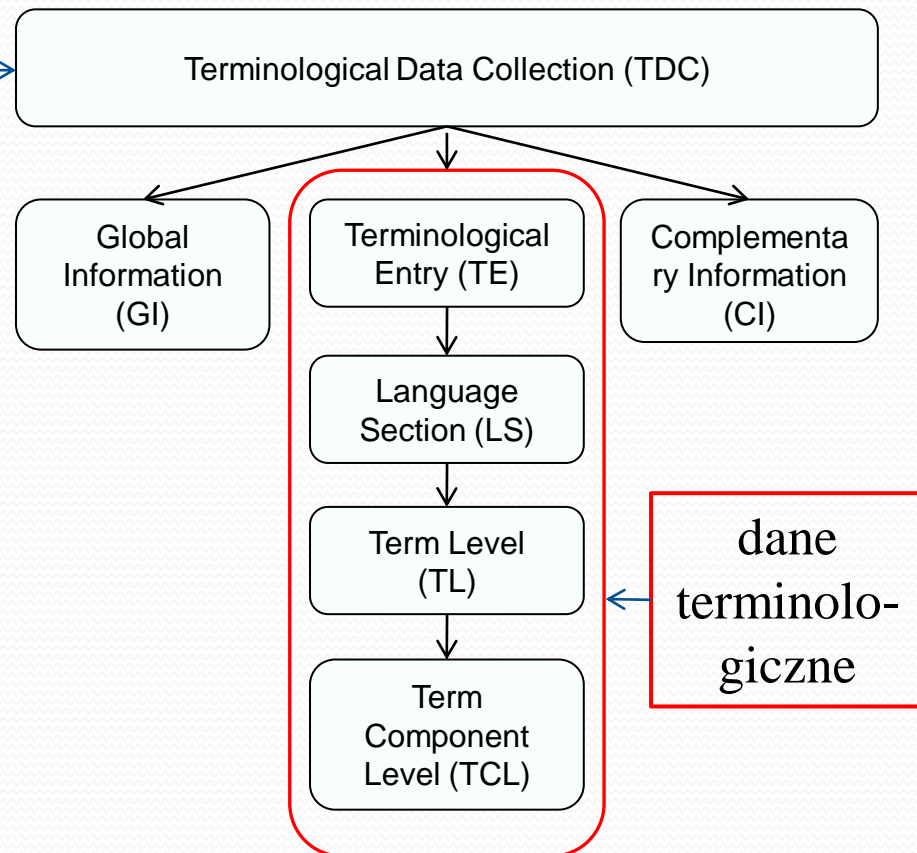
Struktura dokumentu w TBX

TBX odzwierciedla metamodel TMF.

Plik TBX jest kolekcją danych terminologicznych (**TDC**)

Struktura dokumentu

```
<?xml version="1.0"?> [<!DOCTYPE martif
SYSTEM "definicje.dtd"> ]
<martif type="TBX" xml:lang="pl">
<martifHeader> [zawiera GI]</martifHeader>
<text>
<body> [dane terminolog.] </body>
<back> [CI, np. bibliogr., pliki graf.] </back>
</text>
</martif>
```



// dtd - definicja typu dokumentu

MartifHeader

```
<martifHeader>  
  <fileDesc>  
    <sourceDesc>  
      <p>Konferencja Zarządzanie Informacją</p>  
    </sourceDesc>  
  </fileDesc>  
  <encodingDesc>  
    <p type="DCSName"> TBXdefaultXCS-v-1-0.XML </p>  
  </encodingDesc>  
</martifHeader>
```

DCSName określa nazwę pliku XCS, w którym znajdują się wyszczególnione kategorie danych.

Body

```
<body>  
  <termEntry id="0001">  
    [pojęcie: skontrum]  
  </termEntry>  
  <termEntry id="0002">  
    [pojęcie: słowo kluczowe]  
  </termEntry>  
</body>
```

termEntry

```
<termEntry id="0002">  
  <descrip type="subjectField">JIW</descrip>  
  <descrip type="definition">Wyraz lub wyrażenie  
    wybrane z tytułu lub tekstu dokumentu,  
    charakteryzującego jego treść. </descrip>  
  <langSet xml:lang="pl">  
    <tig> <term>słowo kluczowe</term></tig>  
  </langSet>  
  <langSet xml:lang="en">  
    <tig> <term>keyword</term></tig>  
  </langSet>  
</termEntry>
```

tig = term information group: zawiera **<term>** - nazwę (termin);
<termNote> - np. typ terminu (np. akronim), inf. gramatyczne; może
również zawierać odesłanie do źródła w sekcji **back**.

Back

```
<back>
  <refObjectList type="bibl">
    <refObject id="STIN">
      <item type="creator">Dembowska, Maria</item>
      <item type="title">Słownik terminologiczny
                          informacji naukowej.</item>
      <item type="type">book</item>
      <item type="identifier"> 83-911604-0-8 </item>
      <item type="publisher">Zakład Narodowy imienia
                          Ossolińskich</item>
      <item type="date">1979</item>
    </refObject>
  </refObjectList>
</back>
```


Relacje (1)

Relacje zapisuje się w sekcji **termEntry**:

`<descrip type="nazwa relacji" target = "Id pojęcia">`

Relation	Relacja
antonymConcept	antonimii
relatedConcept	kojarzeniowa
subordinateConceptGeneric	węższa generycznie
subordinateConceptPartitive	węższa mereologicznie
superordinateConceptGeneric	szersza generycznie
superordinateConceptPartitive	szersza mereologicznie

Relacje (2)

Relation	Relacja
broaderConceptGeneric	szersza generycznie o co najmniej 2 poz.
broaderConceptPartitive	szersza mereologicznie o co najmniej 2 poz.
conceptPosition	wskazująca miejsce w systemie (np. nr klasy w klasyfikacji)
coordinateConceptGeneric	współrzędności generyczna
coordinateConceptPartitive	współrzędności mereologiczna
relatedConceptBroader	kojarzeniowa szersza
relatedConceptNarrower	kojarzeniowa węższa
sequentiallyRelatedConcept	sekwencyjna (czas lub przestrzeń) (np. jesień-lato)
spatiallyRelatedConcept	przestrzenna miejsca/pozycji (Katowice-Sosnowiec)
temporallyRelatedConcept	czasowa (porozumienia Okrągłego Stołu – upadek muru berlińskiego)

Relacja ekwiwalencji

- 1) Ekwiwalenty terminu w językach obcych zapisuje się w sekcji termEntry za pomocą znacznika `<langSet xml:lang="pl">`
- 2) Nie ma osobnego znacznika na relację synonimii. Jej realizacja wygląda następująco:

```
<termEntry>  
  <tig><term>funkcja impresywna</term></tig>  
  <tig><term>funkcja wokatywna</term></tig>  
  <tig><term>funkcja apelatywna</term></tig>  
</termEntry>
```

Podsumowanie

Dzięki TBX mamy możliwość:

- wykorzystywania rozproszonych źródeł terminologicznych w różnych systemach, aplikacjach, projektach;
- zdecentralizowanego tworzenia zasobów terminologicznych (np. w różnych dziedzinach wiedzy);
- łączenia różnych zasobów w całość, tworząc ogólnokrajowy/ogólnoswiatowy bank danych terminologicznych.

A large, curling blue wave crashing against a clear blue sky. The wave is the central focus, with its crest curling over and creating a tunnel-like structure. The water is a deep, vibrant blue, and the sky is a clear, bright blue. The overall scene is dynamic and powerful.

Dziękuję za uwagę i cierpliwość.