

# **Metodika pro zpřesňování vztahů mezi hesly v databázi TDKIV a zpřístupňování údajů v systému Wikidata**

**Výstup z Pilotního projektu zpřístupnění TDKIV v podobě propojených  
dat**

**Verze 1.0**

Zpracovala PhDr. Linda Jansová, Ph.D.

Praha, prosinec 2018

# OBSAH

1 Úvod.....	3
2 Propojená data.....	3
3 Systém Wikidata.....	5
4 Propojená data a TDKIV.....	7
5 Doporučené pracovní postupy.....	8
5.1 Získání dat a jejich příprava pro další zpracování.....	9
5.2 Zpracování dat.....	10
5.3 Hlavní použité vlastnosti.....	15
5.3.1 Symetrické vlastnosti.....	16
5.3.2 Vlastnosti bez obrácených vlastností.....	16
5.3.3 Vlastnosti s obrácenými vlastnostmi.....	17
5.4 Přehled propojených položek.....	20
6 Možnosti dalšího využití zpracovaných dat.....	20
7 Shrnutí.....	21

# 1 Úvod

Metodika vznikla v rámci *Pilotního projektu zpřístupnění TDKIV v podobě propojených dat* (dále též jako projekt nebo pilotní projekt), který realizovala Národní knihovna ČR v roce 2018, a to s využitím financí z Ministerstva kultury ČR (konkrétně z dotačního programu VISK 1). Doplnuje druhou verzi Metodického materiálu pro zpracování hesel České terminologické databáze knihovnictví a informační vědy (TDKIV)<sup>1</sup>. Dokument schválila redakční rada TDKIV na svém zasedání konaném 8. ledna 2019 v Národní knihovně ČR.

Návrh metodiky připomínkovali následující odborníci, kteří spolupracovali na pilotním projektu a/nebo jsou členy redakční rady TDKIV:

- Mgr. Marie Balíková (Národní knihovna ČR);
- Mgr. Eva Bartůňková (Národní knihovna ČR);
- PhDr. Marcela Buřilová (Vyšší odborná škola informačních studií a Střední škola elektrotechniky, multimédií a informatiky Praha);
- Mgr. Eva Cerniňáková (Knihovna Jabok, doktorandka na Ústavu informačních studií a knihovnictví Filozofické fakulty Univerzity Karlovy);
- Bc. Jaroslava Citová, DiS. (Národní knihovna ČR);
- PhDr. Helena Kučerová, Ph.D. (Vysoká škola ekonomická v Praze);
- PhDr. Vít Richter (Národní knihovna ČR);
- PhDr. Alena Richterová, CSc. (Národní knihovna ČR);
- PhDr. Klára Rösslerová, Ph.D. (Knihovna Filozofické fakulty Univerzity Karlovy).

Jazykovou korekturu provedla PhDr. Anna Machová (Národní knihovna ČR).

V metodice je nejprve uvedena stručná charakteristika propojených dat a systému Wikidata, následují základní informace o pilotním projektu, v jehož rámci vznikla metodika, a poté již podrobný popis pracovních postupů použitých v rámci projektu a doporučených i k využití v budoucnu. Zmíněny jsou i některé z možností dalšího využití zpracovaných dat.

## 2 Propojená data

Propojená data (angl. *linked data*) představují nejvhodnější způsob zveřejňování, sdílení a propojování dat (informací a poznatků) v prostředí tzv. sémantického webu. Využívají jednotné identifikátory zdrojů (angl. *uniform resource identifiers*, URIs) a rámeček pro popis zdrojů (angl. *Resource Description Framework*, RDF)<sup>2,3,4</sup>.

---

1 *Metodický materiál pro zpracování hesel České terminologické databáze knihovnictví a informační vědy (TDKIV)* [online]. Zpracoval Josef SCHWARZ, upravila Jaroslava HAVLOVÁ. 2. upr. verze, 2013 [cit. 2018-11-25]. 24 s. Dostupné z: <https://tdkiv.nkp.cz/docs/MM2013vv.docx>

2 *Linked Data - Connect Distributed Data across the Web* [online]. [cit. 2018-11-25]. Dostupné z: <http://linkeddata.org/>

3 *Linked data. Wikipedia: the free encyclopedia* [online]. San Francisco (CA, USA): Wikimedia Foundation, 2001- [cit. 2018-11-25]. Dostupné z: [https://en.wikipedia.org/wiki/Linked\\_data](https://en.wikipedia.org/wiki/Linked_data)

Často je používán také termín propojená otevřená data (angl. *linked open data*). Jedná se o propojená data, která jsou zároveň daty zpřístupněnými pod otevřenou licenci, jež nebrání jejich dalšímu využívání.<sup>5</sup>

V podobě propojených otevřených dat jsou především v posledních letech často zpřístupňovány také nejrůznější zdroje pocházející z knihoven. Svědčí o tom i jejich přítomnost v grafickém přehledu propojených otevřených dat známém jako tzv. LOD cloud<sup>6</sup>.

V tomto přehledu nechybí např. následující zdroje:

- předmětový heslář Kongresové knihovny ve Washingtonu<sup>7</sup>;
- německá národní bibliografie<sup>8</sup>;
- britská národní bibliografie<sup>9</sup>;
- virtuální mezinárodní soubor autoritních záznamů<sup>10</sup>.

V českém prostředí se problematikou propojených otevřených dat v oblasti knihovnictví zabývali především následující odborníci:

- Jindřich Mynarz, který se zasloužil i o zpřístupnění Polytematického strukturovaného hesláře (PSH) z Národní technické knihovny v podobě propojených otevřených dat a jeho zařazení do zmiňovaného grafického přehledu propojených dat<sup>11,12,13</sup>;
- Helena Kučerová a Eva Bratková, které v letech 2013 až 2015 pod hlavičkou Ústavu informačních studií a knihovnictví Filozofické fakulty Univerzity Karlovy (ÚISK FF UK) realizovaly projekt *Znalostní báze pro obor organizace informací a znalosti*<sup>14,15</sup>;

---

4 HAVLOVÁ, Jaroslava. Propojená data. *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR, 2003- [cit. 2018-11-26]. Dostupné z: [https://aleph.nkp.cz/F/?func=direct&doc\\_number=000016064&local\\_base=KTD](https://aleph.nkp.cz/F/?func=direct&doc_number=000016064&local_base=KTD)

5 Viz též HAVLOVÁ, Jaroslava. Propojená otevřená data. *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR, 2003- [cit. 2018-11-26]. Dostupné z: [https://aleph.nkp.cz/F/?func=direct&doc\\_number=000016065&local\\_base=KTD](https://aleph.nkp.cz/F/?func=direct&doc_number=000016065&local_base=KTD)

6 *The Linked Open Data Cloud* [online]. [cit. 2018-11-25]. Dostupné z: <https://lod-cloud.net/>

7 Library of Congress Subject Headings. *The Linked Open Data Cloud* [online]. [cit. 2018-11-25]. Dostupné z: <https://lod-cloud.net/dataset/lcsh>

8 Deutsche Nationalbibliografie (DNB). *The Linked Open Data Cloud* [online]. [cit. 2018-11-25]. Dostupné z: <https://lod-cloud.net/dataset/deutsche-nationalbibliografie-dnb>

9 British National Bibliography (BNB) - Linked Open Data. *The Linked Open Data Cloud* [online]. [cit. 2018-11-25]. Dostupné z: <https://lod-cloud.net/dataset/bluk-bnb>

10 VIAF: The Virtual International Authority File. *The Linked Open Data Cloud* [online]. [cit. 2018-11-25]. Dostupné z: <https://lod-cloud.net/dataset/viaf>

11 Polythematic Structured Subject Heading System. *The Linked Open Data Cloud* [online]. [cit. 2018-11-25]. Dostupné z: <https://lod-cloud.net/dataset/psh-subject-headings>

12 MYNARZ, Jindřich a Jan ZEMÁNEK. Úvod k linked data. *Knihovna plus* [online]. 2010, č. 1 [cit. 2018-11-26]. ISSN 1801-5948. Dostupné z: <http://knihovna.nkp.cz/knihovnaplus101/myna.htm>

13 MYNARZ, Jindřich, Kateřina KAMRÁDKOVÁ a Kristýna KOŽUCHOVÁ. Novinky z oblasti Polytematického strukturovaného hesláře. *Ikaros* [online]. 2009, roč. 13, č. 7 [cit. 2018-11-25]. urn:nbn:cz:ik-13200. ISSN 1212-5075. Dostupné z: <http://ikaros.cz/node/13200>

14 *Znalostní báze pro obor organizace informací a znalosti* [online]. Verze 1.0. Praha: Univerzita Karlova v Praze – Filozofická fakulta, c2015-2016 [cit. 2018-11-25]. Dostupné z: <http://ko.cuni.cz/>

15 BRATKOVÁ, Eva a Helena KUČEROVÁ. K otázkám metadatového popisu systémů organizace znalostí. *Knihovna: knihovnická revue* [online]. 2015, roč. 26, č. 1, s. 5–36 [cit. 2018-11-26]. ISSN 1801-3252. Dostupné z: <http://knihovnarevue.nkp.cz/archiv/2015-01/recenzovane-prispevky/k-otazkam-metadatoveho-popisu-systemu-organizace-znalosti>

- Klára Rösslerová, která se tématu věnovala mj. v části své disertační práce *Historie, současnost a budoucnost výměnných formátů bibliografických dat*, obhájené v roce 2017 na ÚISK FF UK<sup>16</sup>, a v další publikační a přednáškové činnosti<sup>17,18</sup>.

### 3 Systém Wikidata

Systém Wikidata<sup>19</sup> může být stručně popsán jako rozsáhlá znalostní báze, která má podobu propojených otevřených dat, je stejně jako Wikipedie vytvářena uživateli z celého světa a funguje pod hlavičkou Nadace Wikimedia (Wikimedia Foundation). Jeho cílem je poskytovat společnou otevřenou datovou základnu, ze které mohou čerpat další projekty Nadace Wikimedia i kterékoliv jiné projekty. Například v listopadu 2018 byla data z Wikidat využívána v téměř 60 % článků anglické Wikipedie<sup>20</sup>.

Základní jednotkou systému Wikidata je **položka** (angl. *item*), která může obsahovat řadu údajů od štítku a popisu v různých jazycích (viz obr. 1) přes specifikaci vztahu k jiným položkám až po propojení na konkrétní jednotky z různých externích zdrojů (např. pořádacích systémů).

#### inkunábule (Q216665)

book, pamphlet, or broadside that was printed before the year 1501 in Europe angličtina  
 prvotisk | inkunabule

 editovat

▼ V dalších jazycích Konfigurovat

Jazyk	Štítek	Popis	Také znám jako
čeština	inkunábule	Popis nebyl definován	prvotisk inkunabule
angličtina	incunable	book, pamphlet, or broadside that was printed before the year 1501 in Europe	incunabulum fifteener
slovenština	Inkunábula	Popis nebyl definován	
němčina	Inkunabel	mit beweglichen Lettern gedruckte Bücher und Einblattdrucke im 15. Jh.	Inkunabeln Frühdruck Wiegendruck

Všechny zadané jazyky

Obr. 1: Příklad položky Q216665 (inkunábule) – názvy a popisy v různých jazycích

16 RÖSSLEROVÁ, Klára. *Historie, současnost a budoucnost výměnných formátů bibliografických dat [History, Presence and Future of Exchangeable Formats of Bibliographic Data]*. Praha, 2016. 188 s., xlvii s. příl. Disertační práce. Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví. Školitel doc. PhDr. Rudolf Vlasák. Dostupné také z: <https://is.cuni.cz/webapps/zzp/detail/105313/24537045/?q=%7B%22searchform%22%3A%22historie+po%5Cu010d%5Cu00edta%5Cu010dov%5Cu00fdch+her%22%2C%22searchform%22%3A%22Vyhledat%22%2C%22PNzzpSearchListbasic%22%3A%2250%22%7D&lang=cs>

17 RÖSSLEROVÁ, Klára. European BIBFRAME Workshop 2017 aneb První evropský workshop BIBFRAME. *Bulletin SKIP* [online]. 2018, roč. 27, č. 1 [cit. 2018-11-25]. ISSN 1213-5828. Dostupné z: <https://bulletinskip.skipcr.cz/node/311>

18 RÖSSLEROVÁ, Klára. Trendy v oblasti bibliografických formátů aneb Má MARC budoucnost?. *Bulletin SKIP* [online]. 2017, roč. 26, č. 2 [cit. 2018-11-25]. ISSN 1213-5828. Dostupné z: <https://bulletinskip.skipcr.cz/node/166>

19 Wikidata [online]. This page was last edited on 20 September 2018, at 21:13 [cit. 2018-11-27]. Dostupné z: [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)

20 Wikidata. *Wikipedia: the free encyclopedia* [online]. San Francisco (CA, USA): Wikimedia Foundation, 2001- [cit. 2018-11-25]. Dostupné z: <https://en.wikipedia.org/wiki/Wikidata>

Podstatnou částí položek jsou **výroky**, což jsou tvrzení, která na základě přesných vztahů propojují jednu položku s jinou položkou. Příkladem může být výrok, který říká, že položka Q216665 (inkunábule) má jako nadtřídu položku Q1261026 (tiskovina), viz obr. 2.

#### Výroky

nadtřída	tiskovina ▼ 0 referencí	editovat + přidat referenci
	dokument ▼ 0 referencí	editovat + přidat referenci
	broadside <i>angličtina</i> ▼ 0 referencí	editovat + přidat referenci
	pamflet ▼ 0 referencí	editovat + přidat referenci
	kniha ▼ 0 referencí	editovat + přidat referenci
		+ přidat hodnotu

Obr. 2: Příklad položky Q216665 (inkunábule) – výroky

Do systému Wikidata přidávají své zdroje i paměťové instituce. V České republice se jedná např. o data z Národního památkového ústavu<sup>21</sup>. Do Wikidat jsou zařazována katalogová a rejstříková čísla odkazující na položky z Ústředního seznamu kulturních památek a Památkového katalogu, viz obr. 3.

21 NPÚ navázal užší spolupráci s Wikipedií, konal se i společný workshop. *Národní památkový ústav* [online]. Praha: Národní památkový ústav, 19. 1. 2018 [cit. 2018-11-25]. Dostupné z: <https://www.npu.cz/cs/npu-a-pamatkova-pece/npu-jako-institute/zpravy/31585-npu-navazal-uzsi-spolupraci-s-wikipedii-konal-se-i-spolecny-workshop>

## Identifikátory

identifikátor Freebase	 /m/059ty9 ▶ 1 reference
Rejstříkové číslo Ústředního seznamu kulturních památek	 38162/1-7 ▶ 1 reference
Katalogové číslo Památkového katalogu	 1000001634 ▶ 1 reference

Obr. 3: Příklad položky Q1100429 (Klementinum) – identifikátory z Národního památkového ústavu

## 4 Propojená data a TDKIV

V roce 2018 Národní knihovna ČR v rámci projektu financovaného Ministerstvem kultury ČR prostřednictvím dotačního programu VISK 1 realizovala *Pilotní projekt zpřístupnění TDKIV v podobě propojených dat*. Jeho cílem bylo zvýšit vzdělávací a informační potenciál České terminologické databáze knihovnictví a informační vědy (TDKIV) prostřednictvím propojených dat a zpřesnit vzájemné vztahy mezi hesly. Hlavním výstupem z projektu je předložená metodika, která je založena na praktických zkušenostech získaných při zpracování části dat z TDKIV.

Výstupy z projektu rozšíří možnosti využití dat z TDKIV v knihovnicko-informačním prostoru i mimo něj, a to v českém i mezinárodním kontextu. Usnadní také orientaci v bázi TDKIV jako celku, a to jak pro uživatele, tak pro odborníky, kteří se podílejí na zpracování nových hesel. Přispěje také ke snazšímu zpracování nových hesel a jejich vazeb na existující hesla.

Báze TDKIV<sup>22</sup> byla vytvořena v letech 2001–2002 v rámci projektu, který byl podpořen grantem Ministerstva kultury ČR. Je spravována Knihovnickým institutem Národní knihovny ČR.

22 TDKIV: Česká terminologická databáze knihovnictví a informační vědy [online]. Praha: Národní knihovna ČR [cit. 2018-11-25]. Dostupné z: <http://tdkiv.nkp.cz/>

V průběhu dosavadních patnácti let své existence se stala jedním ze základních pilířů formálního i neformálního vzdělávání v knihovnicko-informačních oborech. Je hojně citována v odborných publikacích (časopiseckých člancích, příspěvcích ve sbornících, monografiích atd.), ale také například ve vysokoškolských kvalifikačních pracích. V roce 2016 byla využita jako základ terminologického slovníku v knize *Otevřený přístup k vědeckým informacím: současný stav v České republice a ve světě*<sup>23</sup>.

V současné době tvoří TDKIV cca 3 200 definitivních hesel, která obsahují řadu dílčích údajů. Kromě termínů samotných jsou uváděny jejich české a anglické ekvivalenty, příbuzné termíny, výklady (včetně normativních), věcné skupiny, poznámky (viditelné pro uživatele i interní), údaje o autorech, lektorech, redaktorech a konzultantech, statusy záznamů a údaje o zpracování.

Databázi využívá často česká verze online encyklopedie Wikipedia, jejíž články odkazují na řadu hesel z báze TDKIV. V posledních letech se ukazuje, že právě přístupy do báze TDKIV prostřednictvím těchto odkazů mají nemalý podíl na celkovém počtu přístupů do báze. S ohledem na tuto skutečnost i na zkušenosti se zpřístupňováním informačních zdrojů v podobě vhodné pro vyhledávače a další webové služby je přínosné ještě výrazněji vylepšit dostupnost dat z báze TDKIV. Vzhledem ke struktuře dat a k trendům v oblasti zpřístupňování dat v prostředí internetu se jako vhodný prostředek jeví tzv. propojená data.

Jednotlivá hesla v databázi jsou mezi sebou propojena pouze vztahem příbuznosti. Je tedy možné zjistit, že konkrétní termín je příbuzný s konkrétním jiným termínem, ale nelze se orientovat podle přesnější (sémanticky bohatší) vazby. Jinými slovy součástí báze nejsou vztahy obvykle používané v pořádacích systémech (např. hierarchické vazby) ani ještě přesnější vazby, které mohou být například součástí ontologií. Obohacení dat z TDKIV o tyto vazby rozšíří možnosti jejich dalšího využití pro vzdělávací i vědecké účely, a je tedy žádoucí.

V dalším textu jsou uvedeny jak konkrétní postupy použité v rámci projektu, tak postupy, které z nich přímo vycházejí a jejichž využití se jeví jako optimální pro zpracovávání dat v budoucnu.

## 5 Doporučené pracovní postupy

Při řešení projektu byly na vzorcích dat prakticky ověřeny postupy pro zpřesňování vazeb mezi hesly TDKIV a pro zpřístupňování dat v systému Wikidata.

Postupy jsou určeny zejména pro zpracování většího množství dat zároveň. Budou tedy využity při zpracovávání dat po ukončení pilotního projektu, případně v upravené podobě v budoucnu, pokud by bylo třeba provádět změny ve větším počtu hesel zároveň. Při zpracování jednotlivých hesel nově zařazovaných do databáze TDKIV již nebude třeba provádět úvodní kroky (např. získání dat z exportu v XML apod.), neboť bude postačující pracovat s jednotlivými hesly v běžně dostupné podobě záznamu v systému Aleph<sup>24</sup>.

---

23 DĚDIČOVÁ, Petra, Pavla RYGLOVÁ, Jiří MAREK a Jan SKŮPA (eds.). *Otevřený přístup k vědeckým informacím: současný stav v České republice a ve světě* [online]. Brno: VUTIUM, 2016 [cit. 2018-11-25]. Dostupné z: <https://doi.org/10.13164/book.0a>

24 KTD: *Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR, 2003- [cit. 2018-11-27]. Dostupné z: <https://aleph.nkp.cz/cze/ktid>



## 5.1 Získání dat a jejich příprava pro další zpracování

Data je nejprve třeba vyexportovat<sup>25</sup> ze systému Aleph v upravené podobě formátu MARCXML. Jednotlivá hesla (záznamy) mají jednotnou podobu, viz příklad na obr. 4.

```
<record xmlns="http://www.loc.gov/MARC21/slim" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://www.loc.gov/MARC21/slim
http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd">
  <leader>nx a22 n 4500</leader>
  <controlfield tag="LDR">nx a22 n 4500</controlfield>
  <controlfield tag="FMT">TD</controlfield>
  <datafield tag="TER" ind1=" " ind2=" ">
    <subfield code="a">inkunábule</subfield>
  </datafield>
  <datafield tag="EKV" ind1=" " ind2=" ">
    <subfield code="a">prvotisk</subfield>
    <subfield code="b">x</subfield>
  </datafield>
  <datafield tag="PTE" ind1=" " ind2=" ">
    <subfield code="a">inkunabulistika</subfield>
  </datafield>
  <datafield tag="PTE" ind1=" " ind2=" ">
    <subfield code="a">paleotyp</subfield>
  </datafield>
  <datafield tag="PTE" ind1=" " ind2=" ">
    <subfield code="a">postinkunábule</subfield>
  </datafield>
  <datafield tag="ANG" ind1=" " ind2=" ">
    <subfield code="a">incunable</subfield>
  </datafield>
  <datafield tag="ANG" ind1=" " ind2=" ">
    <subfield code="a">incunabulum</subfield>
  </datafield>
  <datafield tag="VYK" ind1=" " ind2=" ">
    <subfield code="a">
      Tiskárna vyrobena pomocí pohyblivých liter v období od vynálezu knihtisku do roku 1500
      včetně. Zpočátku se projevovává snaha napodobit písmem, výzdobou a úpravou dobovou
      rukopisnou knihu.
    </subfield>
    <subfield code="b">HORÁK-1948:27</subfield>
    <subfield code="b">BOHATCOVÁ-1990:121-154</subfield>
  </datafield>
  <datafield tag="OSV" ind1=" " ind2=" ">
    <subfield code="a">
      Svazek tištěný v Evropě pohyblivými typy a datovaný před 1. lednem 1501. Pozn.: Jméno
      tiskaře inkunábule je možné zjistit z fontu použitého pro tisk.
    </subfield>
    <subfield code="b">ČSN ISO 5127-2003</subfield>
  </datafield>
  <datafield tag="VSK" ind1="1" ind2=" ">
    <subfield code="a">XAA</subfield>
  </datafield>
  <datafield tag="AUT" ind1=" " ind2=" ">
    <subfield code="a">Uhlíř, Zdeněk</subfield>
  </datafield>
  <datafield tag="RED" ind1=" " ind2=" ">
    <subfield code="a">Schwarz, Josef</subfield>
  </datafield>
  <datafield tag="RED" ind1=" " ind2=" ">
    <subfield code="a">Burgetová, Jarmila</subfield>
  </datafield>
  <datafield tag="LEK" ind1=" " ind2=" ">
    <subfield code="a">Žibritová, Gabriela</subfield>
  </datafield>
  <datafield tag="ZDR" ind1="1" ind2=" ">
    <subfield code="u">
      http://aleph.nkp.cz/F/?func=direct&local_base=KKL&doc_number=000028855
    </subfield>
    <subfield code="y">HORÁK-1948:27</subfield>
    <subfield code="4">N</subfield>
  </datafield>
  <datafield tag="ZDR" ind1="1" ind2=" ">
    <subfield code="u">
      http://aleph.nkp.cz/F/?func=direct&local_base=KKL&doc_number=000041774
    </subfield>
    <subfield code="y">BOHATCOVÁ-1990:121-154</subfield>
    <subfield code="4">N</subfield>
  </datafield>
  <datafield tag="ZDN" ind1="1" ind2=" ">
    <subfield code="u">
      http://aleph.nkp.cz/F/?func=direct&local_base=KKL&doc_number=000053613
    </subfield>
    <subfield code="y">ČSN ISO 5127-2003</subfield>
    <subfield code="4">N</subfield>
  </datafield>
  <datafield tag="STS" ind1=" " ind2=" ">
    <subfield code="a">D</subfield>
  </datafield>
  <controlfield tag="BAS">W</controlfield>
  <datafield tag="IST" ind1=" " ind2=" ">
    <subfield code="a">im20030210</subfield>
  </datafield>
  <datafield tag="IST" ind1=" " ind2=" ">
    <subfield code="a">op20031212</subfield>
  </datafield>
  <datafield tag="KSL" ind1=" " ind2=" ">
    <subfield code="a">cradle book</subfield>
  </datafield>
</record>
```

Obr. 4: Příklad záznamu v exportním XML formátu

<sup>25</sup> Export může provést správce databáze TDKIV.

Tato data je vhodné dále zpracovat do podoby tabulky (resp. souboru TSV či CSV) prostřednictvím XSLT, viz např. obr. 5 (ze souboru byly již odstraněny jmenné prostory).

```
<xsl:stylesheet version = '1.0' xmlns:xsl='http://www.w3.org/1999/XSL/Transform'>
<xsl:output method="text" encoding="utf-8" />
<xsl:strip-space elements="*" />

  <xsl:template match="record">
<xsl:for-each select="datafield[contains(@tag,'TER')]">
  <xsl:value-of select="subfield"/>
  <xsl:text>|</xsl:text>
</xsl:for-each>
<xsl:for-each select="datafield[contains(@tag,'STS')]">
  <xsl:value-of select="subfield"/>
  <xsl:text>|</xsl:text>
</xsl:for-each>
<xsl:if test="not(datafield[@tag='VYK'])">
  <xsl:text>bez hodnoty</xsl:text>
  <xsl:value-of select="'&#10;'" />
</xsl:if>
<xsl:for-each select="datafield[contains(@tag,'VYK')]">
  <xsl:value-of select="subfield"/>
  <xsl:if test="count(subfield) &gt; 1 and position() != last()">
  <xsl:text>|</xsl:text></xsl:if>
  <xsl:value-of select="'&#10;'" />
</xsl:for-each>
</xsl:template>
```

Obr. 5: Příklad šablony XSLT sloužící k vypsaní termínů, statusů a výkladů

K aplikaci šablony XSLT a získání potřebných dat lze využít např. program xsltproc.

Protože export dat v XML neobsahuje identifikátory záznamů, je užitečné tyto údaje doplnit z výstupního formátu, který je obsahuje. Lze využít skutečnosti, že termíny jako řetězce jsou v rámci databáze jedinečné.

K propojení údajů je možné použít např. programy cat, sed a podle potřeby další (např. dos2unix) a následující sekvenci příkazů:

```
cat sys_utf-2.txt | grep -e '* @' -e '^TER: ' > filtrace1.txt
dos2unix filtrace1.txt
cat filtrace1.txt | tr -d '\n' | sed -e "s/* @/\n/g" | sed -e "s/TER: /\t/g" > filtrace_hotova.tsv
```

Lze rovněž doporučit vybrat si pouze záznamy se statusem D nebo R<sup>26</sup> (záznamy s jinými statusy nejsou úplné a jsou součástí pracovní, nikoliv oficiální verze databáze). Toho je možné docílit více způsoby, např. výběrem (filtrací či řazením) příslušných dat v tabulkovém procesoru.

## 5.2 Zpracování dat

Další zpracování dat je možné realizovat v tabulkovém procesoru. V tabulce je vhodné přidat sloupce pro přidávání odpovídajících položek z Wikidat (URL, eventuálně názvů v angličtině

<sup>26</sup> Status D mají záznamy v definitivní podobě, status R záznamy v téměř definitivní podobě.

a/nebo češtině), resp. tematických autorit Národní knihovny ČR<sup>27</sup> (identifikátor, eventuálně preferovaný termín). Osvědčilo se ruční mapování hesel TDKIV na položky Wikidat s využitím standardních vyhledávacích nástrojů systému Wikidata (ruční mapování lze použít i v případě tematických autorit Národní knihovny ČR). Při vyhledávání údajů ve Wikidatech se osvědčuje využívání jak českých, tak zejména anglických výrazů a podle potřeby také ekvivalentů (synonym). Při mapování na tematické autority Národní knihovny ČR je nutné brát ohled na skutečnost, že zatímco v databázi TDKIV jsou termíny uváděny převážně v singuláru, v tematických autoritách je obvykle upřednostňován plurál. Mapování přináší kromě jiného užitečné podněty pro zpřesňování existujících vztahů mezi hesly a případně přidávání dalších vztahů.

K vizualizaci zpracovávaných dat je možné využít jednak pracovní online aplikaci, která vznikla v rámci projektu a je dostupná na adrese <http://tdkiv.nicmila.science/>, jednak podle potřeby další nástroje, např. editor Protégé vytvářený na Stanfordově univerzitě obohacený o doplňkové moduly, např. o moduly Cellfie<sup>28</sup> či csv-export-plugin<sup>29</sup>.

Pokud jde o přímé vkládání údajů do systému Wikidata, k propojení položek v systému Wikidata s hesly v TDKIV slouží vlastnost TDKIV term ID<sup>30</sup>, o jejíž vytvoření bylo požádáno v létě 2018<sup>31</sup>. Přestože editaci obsahu Wikidat mohou provádět i nepřihlášení/neregistrovaní uživatelé, ti, kteří počítají s výraznější aktivitou v systému, si obvykle účet zřizují<sup>32</sup> a editace provádějí po přihlášení.

Za účelem zpřístupnění dat v systému Wikidata je optimální zvolit následující postup:

Nejprve je třeba **vybrat heslo ke zpracování** a seznámit se s jeho aktuální podobou ve zdrojové databázi v systému Aleph.

Poté následuje **ověření, zda v systému Wikidata již existuje odpovídající položka**. Za tím účelem je nevhodnější použít vyhledávací okénko, viz obr. 6.

---

27 AUT – Databáze národních autorit NK ČR [online]. Praha: Národní knihovna ČR [cit. 2018-12-31]. Dostupné z: <https://aleph.nkp.cz/cze/auvt>

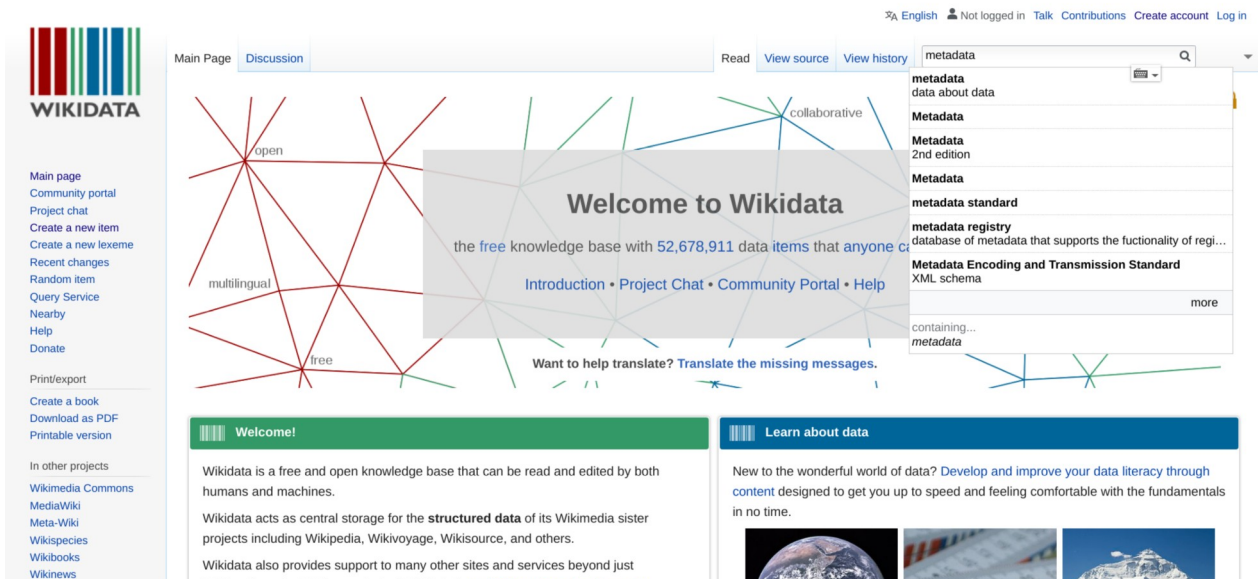
28 Cellfie [online]. *GitHub* [cit. 2018-11-27]. Dostupné z: <https://github.com/protegeproject/cellfie-plugin>

29 csv-export-plugin [online]. *GitHub* [cit. 2018-11-27]. Dostupné z: <https://github.com/protegeproject/csv-export-plugin>

30 identifikátor termínu v TDKIV. *Wikidata* [online]. Stránka byla naposledy editována 25. 10. 2018 v 17:05 [cit. 2018-11-25]. Dostupné z: <https://www.wikidata.org/wiki/Property:P5398>

31 Wikidata:Property proposal/TDKIV term ID. *Wikidata* [online]. Stránka byla naposledy editována 29. 6. 2018 v 08:08 [cit. 2018-11-25]. Dostupné z: [https://www.wikidata.org/wiki/Wikidata:Property\\_proposal/TDKIV\\_term\\_ID](https://www.wikidata.org/wiki/Wikidata:Property_proposal/TDKIV_term_ID)

32 Vytvořit účet. *Wikidata* [online]. [cit. 2018-11-25]. Dostupné z: <https://www.wikidata.org/w/index.php?title=Special:CreateAccount&returnto=Wikidata%3AMain+Page>



Obr. 6: Příklad vyhledávání položky v systému Wikidata

V něm funguje našeptávání, tj. začneme-li psát hledaný výraz, objevují se nám postupně se zpřesňující navrhované výsledky. Jak již bylo naznačeno, je vhodné zkusit jak české, tak především anglické termíny (protože jsou Wikidata rozvíjena uživateli z celého světa, je vysoká pravděpodobnost, že odpovídající položka ve Wikidatech už bude, ale nemusí mít šútek v češtině).

Při formulaci dotazu je třeba mít na paměti skutečnost, že v TDKIV jsou nyní jako ekvivalenty v některých případech uváděny i termíny, které jsou ve vztahu k hlavnímu termínu v heslu podřazené (ale protože nebyly samostatně zpracovány, jsou uvedeny právě jako ekvivalenty) – v takových případech je vhodnější za odpovídající položku na Wikidatech pokládat tu, která odpovídá hlavnímu termínu.

Při našeptávání jsou k dispozici i hodnoty pole popis (angl. *description*), které přinášejí stručnou charakteristiku položky a mohou být využity k rozlišení položek se stejným nebo podobným názvem.

Je třeba dát pozor na velká počáteční písmena u názvů položek. Wikidata sice uživatelům doporučují, aby nepsali názvy položek s velkými počátečními písmeny, pokud se nejedná o vlastní jména, to však není vždy možné dodržet (často se jedná o případy automatických importů dat z různých zdrojů, kdy velikost písmen není vždy možné strojově upravit, protože v řadě případů stroj bez lidského zásahu správnou velikost písmene nemůže umět rozpoznat). Jinými slovy relevantní mohou být i výrazy, které nejsou vlastními jmény, ale jsou napsány s počátečními velkými písmeny.

Řada položek s odpovídajícím názvem nemá vyplněno pole popis, popř. obsahuje pouze minimum údajů. Může být tedy relativně složité zjistit, zda je to skutečně správná položka, či nikoliv. V případě pochybností lze doporučit průzkum odkázaných článků z Wikipedie (pochází-li článek z jinojazyčné verze, je možné pro základní orientaci použít např. překladáč provozovaný společností Google<sup>33</sup>). V případě nepřesvědčivých výsledků je lepší založit novou položku. Pokud se později ukáže, že jsou položky totožné, je k jejich propojení možné použít vlastnost *údajně totéž co* (angl. *said to be the same as*)<sup>34</sup> či *trvalá kopie* (angl. *permanent duplicate item*)<sup>35</sup>.

33 Google Translate [online]. Mountain View (CA, USA): Google [cit. 2018-11-25]. Dostupné z: <https://translate.google.com/>

Pokud je položka vyhodnocena jako odpovídající, je třeba **přidat výrok propojující ji s konkrétním heslem z TDKIV**. Nejdříve je třeba vybrat vlastnost (stačí začít psát TDKIV, v políčku funguje našeptávání) a poté jako hodnotu uvést devítimístný identifikátor hesla z TDKIV, tj. např. 000000543 pro heslo *metadata*.

Jako základ vztahů k dalším položkám mohou být využity příbuzné termíny, které jsou součástí většiny hesel v databázi TDKIV. Ty (je-li to v konkrétním případě možné a žádoucí) mohou být zpřesněny výběrem vlastností z Wikidat (nejčastěji používané vlastnosti jsou uvedeny dále v textu) a podle potřeby doplněny o další vztahy.

Pro ilustraci jako konkrétní příklad můžeme uvést heslo *impaktovaný časopis*, viz obr. 7.

#### KTD - Úplné zobrazení záznamu

Zvolte formát: [Standardní](#) -- [Termín & výklad](#) -- [Termín & angl.ekvivalent](#) -- [Heslo s historií](#)

Záznam 1 z 1

Termín	<b>impaktovaný časopis</b>
Termín anglicky	impact journal high-impact journal
Výklad termínu	Odborný <b>časopis</b> , jehož články procházejí před publikováním náročným recenzním řízením, a proto je uznáván odbornou veřejností, o čemž svědčí jeho vysoký impakt faktor, který je mu přiznáván v hodnocení časopisů. To znamená, že na něj, resp. na jeho články odkazují často autoři článků v mnoha jiných uznávaných časopisech. [KATUŠČÁK, MATTHAEIDESOVÁ, NOVÁKOVÁ-1998:173] [DOBEŠ-2009] [IMPAKTČAS-2009]
Zdroj výkladu	<a href="#">KATUŠČÁK, MATTHAEIDESOVÁ, NOVÁKOVÁ-1998:173</a> <a href="#">DOBEŠ-2009</a> <a href="#">IMPAKTČAS-2009</a>
Příbuzný termín	<b>impakt faktor</b> <b>časopis</b>
Autor hesla	Havlová, Jaroslava
Konzultováno s	Nová, Hana
Systém. číslo	000013935

[ [https://aleph.nkp.cz/F/?func=direct&doc\\_number=000013935&local\\_base=KTD](https://aleph.nkp.cz/F/?func=direct&doc_number=000013935&local_base=KTD) ]

Obr. 7: Záznam hesla *impaktovaný časopis* v databázi TDKIV

V databázi TDKIV má heslo uvedeno dva příbuzné termíny, a to *impakt faktor* a *časopis*.

Pokud jde o vztah mezi *impaktovaným časopisem* a *impakt faktorem*, lze jej popsat např. využitím vlastnosti *příčina*, tj. výrok může mít podobu *impaktovaný časopis – příčina* (angl. *has cause*) – *impakt faktor*.

U druhého příbuzného hesla, tedy hesla obsahujícího termín *časopis*, jako vhodnější se jeví propojení *impaktovaného časopisu* s významově užším vědeckým časopisem, neboť *impaktovanými časopisy* nemohou být jiné časopisy než časopisy vědecké. Výrok pak může mít podobu *impaktovaný časopis – nadtřída*<sup>36</sup> (angl. *subclass of*) – *vědecký časopis*.<sup>37</sup>

Jsou-li navrženy vztahy k dalším heslům a mají-li tato hesla již své položky v systému Wikidata, je možné je vzájemně propojit. Pokud další hesla ještě nemají své položky na Wikidatech, vyplatí se je nejprve doplnit a až pak položky mezi sebou propojit.

34 údajně totéž co. *Wikidata* [online]. Stránka byla naposledy editována 3. 10. 2018 v 08:27 [cit. 2018-11-25].

Dostupné z: <https://www.wikidata.org/wiki/Property:P460>

35 trvalá kopie. *Wikidata* [online]. Stránka byla naposledy editována 7. 10. 2018 v 09:43 [cit. 2018-11-25]. Dostupné z: <https://www.wikidata.org/wiki/Property:P2959>

36 V češtině by byl pravděpodobně srozumitelnější překlad „má nadtřidu“ nebo „je podtřídou“, uvedený překlad je převzat z Wikidat, kde je upřednostněn stručný výraz „nadtřída“.

37 Tento příklad názorně ukazuje, že tvorba vztahů pro Wikidata může vést k revizi stávajících vztahů v databázi TDKIV.



O navrhovaných a přidávaných vztazích mezi hesly bude **průběžně informována redakční rada TDKIV**, která díky tomu bude mít možnost navrhnout případné úpravy vztahů.

V případě, že **položka ve Wikidatech neexistuje**, je třeba nejprve **vytvořit novou položku**<sup>38</sup>. V prvním kroku stačí uvést základní údaje, především název položky odpovídající vybranému jazyku, případně další základní údaje vždy podle výchozího zvoleného jazyka (údaje v dalším jazyce lze doplnit při editaci již uložené položky).

Podle potřeby je možné doplnit další údaje, především stručný popis v českém a anglickém jazyce. Jako hlavní zdroj údajů může být využit právě výklad hesla v TDKIV, užitečný je i průzkum dalších zdrojů. Stručný popis by měl mít délku od dvou do max. dvanácti slov<sup>39</sup>.

Pokud je k dispozici **odpovídající tematická autorita Národní knihovny ČR** a u položky ještě není uvedena, je možné ji doplnit. Jedná se o vlastnost NKCR AUT ID<sup>40</sup>.

Při zadávání údajů se vyskytují i některé **dílčí problémy**, které lze rozdělit do čtyř hlavních okruhů:

1. V některých případech jsou ve Wikidatech v ekvivalentech (*aliasch*) uváděny podřazené termíny, které jsou do databáze TDKIV začleněny v samostatných heslech (např. položka *plagiarism*<sup>41</sup> ve Wikidatech vs. hesla *plagiát*<sup>42</sup> a *plagiátorství*<sup>43</sup> v databázi TDKIV).
2. Především u jiných než nejčastěji používaných vlastností nemusí být zřejmé, jak by měly být správně použity. Konkrétním příkladem je vlastnost *product or material produced*, která počítá i s možností zařazení procesu, kdežto popis opačné vlastnosti *manufacturer* proces výslovně neuvádí<sup>44</sup>.
3. V databázi TDKIV se také nachází **řada hesel se dvěma či více výklady**, přičemž v některých případech se každý výklad týká jiného pojmu<sup>45</sup> (společný je jim tedy pouze termín).
4. V systému Wikidata pak může dojít k tomu, že **pro jeden pojem existuje více položek**.

38 Vytvořit novou položku. Wikidata [online]. [cit. 2018-11-25]. Dostupné z:

<https://www.wikidata.org/wiki/Special:NewItem>

39 Help:Description. Wikidata [online]. Stránka byla naposledy editována 17. 11. 2018 v 22:33 [cit. 2018-11-25].

Dostupné z: <https://www.wikidata.org/wiki/Help:Description>

40 NKCR AUT ID. Wikidata [online]. This page was last edited on 12 November 2018, at 17:50 [cit. 2018-11-27].

Dostupné z: <https://www.wikidata.org/wiki/Property:P691>

41 plagiarism. Wikidata [online]. This page was last edited on 14 October 2018, at 20:06 [cit. 2018-11-27]. Dostupné z:

<https://www.wikidata.org/wiki/Q164666>

42 CELBOVÁ, Iva. Plagiát. KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV) [online].

Praha: Národní knihovna ČR, 2003- [cit. 2018-11-27]. Dostupné z: [http://aleph.nkp.cz/F/?func=direct&doc\\_number=000002675&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000002675&local_base=KTD)

43 HAVLOVÁ, Jaroslava. Plagiátorství. KTD: Česká terminologická databáze knihovnictví a informační vědy

(TDKIV) [online]. Praha: Národní knihovna ČR, 2003- [cit. 2018-11-27]. Dostupné z: [http://aleph.nkp.cz/F/?func=direct&doc\\_number=000014609&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000014609&local_base=KTD)

[http://aleph.nkp.cz/F/?func=direct&doc\\_number=000014609&local\\_base=KTD](http://aleph.nkp.cz/F/?func=direct&doc_number=000014609&local_base=KTD)

44 Viz blíže kap. Hlavní použité vlastnosti.

45 Za pojem je pokládána množina pravdivých a podstatných tvrzení, která se týkají referentu. Referentem je pak cokoli, o čem je možné něco tvrdit. Viz blíže FUGMANN, Robert. *Subject analysis and indexing: theoretical foundations and practical advice*. Frankfurt/Main: Indeks Verlag, 1993, s. IX a XIV. ISBN 3-88672-500-6.

Grafické znázornění tzv. sémiotického trojúhelníku (s vrcholy *věc*, *pojem* a *označení*, kdy označení by v našem případě ovšem bylo třeba zúžit pouze na termín) je k dispozici v:

KUČEROVÁ, Helena. Pojem modelu a pojmový model v informační vědě. *Knihovna: knihovnická revue*. 2018, roč. 29, č. 2, s. 5–32, část 2.2. ISSN 1801-3252. Dostupné také z:

<https://knihovnaverevue.nkp.cz/archiv/2018-2/recenzovane-prispevky/pojem-modelu-a-pojmovy-model-v-informacni-vede>

Řešení prvních dvou okruhů problémů je vhodné konzultovat se spolkem Wikimedia Česká republika, popř. s dalšími aktivními členy komunity spojené s Wikidaty v ČR (nejlépe prostřednictvím facebookové skupiny Wikidata CS<sup>46</sup>), případně v zahraničí.

Řešení třetího okruhu problémů je plně v kompetenci Národní knihovny ČR, resp. redakční rady TDKIV. V tomto případě k řešení problému výrazně přispěla realizace pilotního projektu, neboť byla vybrána všechna hesla s očíslovanými výklady (tedy většina hesel s více než jedním výkladem) a k vhodnosti či nevhodnosti spojení výkladů se vyjadřovali členové projektového týmu. Jejich doporučení následně projedná redakční rada TDKIV a příslušné změny budou postupně prováděny přímo v databázi TDKIV (např. s využitím významových závorkových kvalifikátorů). Ta se díky tomu bude do budoucna více orientovat na pojmy než na termíny jako dosud, a tak termíny představující více pojmů s rozdílným výkladem budou zpracovány jako více samostatných hesel, nikoli v rámci jednoho hesla s více dílčími výklady, což byla dosavadní praxe. Pro databázi TDKIV by tedy v budoucnu měla být výstižnější charakteristika „co pojem, to heslo“ než „co termín, to heslo“.

V případě čtvrtého okruhu lze doporučit výběr položky, která je nejpřesnější.

Pokud se týká **aktualizace termínů v systému Wikidata**, je možné využít běžných nástrojů pro editaci položek. Až budou do Wikidat vloženy všechny termíny (a další údaje) z databáze TDKIV a položky odpovídající jednotlivým termínům budou podle potřeby propojeny prostřednictvím vlastností, bude třeba provádět úpravy v podstatě pouze v návaznosti na běžné úpravy (zadávaní nových hesel či aktualizace hesel stávajících) v databázi TDKIV. Vzhledem k počtu těchto úprav, které se obvykle ročně týkají nejvýše desítek hesel, bude tento proces možné realizovat ručně. Užitečné však bude i sledovat vývoj nástrojů pro manipulaci s daty v systému Wikidata – mohou vzniknout nástroje, které tuto aktualizaci umožní provádět automaticky<sup>47</sup>.

Na adrese [https://www.wikidata.org/wiki/Wikidata:Database\\_reports/Constraint\\_violations/P5398](https://www.wikidata.org/wiki/Wikidata:Database_reports/Constraint_violations/P5398) je vhodné průběžně sledovat, zda nejsou hlášena propojení neodpovídající stanoveným omezením vlastností *identifikátor termínu* v TDKIV; tyto sporné případy je žádoucí vyřešit.

## 5.3 Hlavní použité vlastnosti

Ve Wikidatech je k dispozici řada vlastností (viz Wikidata:List of properties<sup>48</sup>), v rámci projektu se však ukázalo, že v případě databáze TDKIV vystačíme s jejich relativně nízkým počtem.

Níže jsou uvedeny nejdůležitější vlastnosti, s nimiž se při řešení projektu pracovalo. Pro přehlednost jsou rozčleněny do tří základních skupin, a to na vlastnosti:

- symetrické (jestliže má položka A vztah X k položce B, pak má položka B také vztah X k položce A);
- bez obrácených vlastností (pouze jednosměrné vlastnosti; opačná vlastnost není v systému Wikidata zavedena);
- s obrácenými vlastnostmi (jestliže má položka A vztah X k položce B, pak má položka B vztah Y k položce A, přičemž vztahy X a Y jsou vzájemně inverzní).

46 Wikidata CS. *Facebook* [online]. [cit. 2018-11-25]. Dostupné z:

<https://www.facebook.com/groups/151013568681003/>

47 Grants:Project/MFFUK/Wikidata & ETL. *Wikimedia Meta Wiki* [online]. This page was last edited on 24 November 2018, at 14:24 [cit. 2018-11-29]. Dostupné z: [https://meta.wikimedia.org/wiki/Grants:Project/MFFUK/Wikidata\\_%26\\_ETL#Endorsements](https://meta.wikimedia.org/wiki/Grants:Project/MFFUK/Wikidata_%26_ETL#Endorsements)

48 Wikidata:List of properties. *Wikidata* [online]. This page was last edited on 3 September 2018, at 18:04 [cit. 2018-11-25]. Dostupné z: [https://www.wikidata.org/wiki/Wikidata:List\\_of\\_properties](https://www.wikidata.org/wiki/Wikidata:List_of_properties)

Součástí přehledu je **celkem šestnáct vlastností**, konkrétně dvě symetrické vlastnosti, pět vlastností bez obrácených vlastností a devět vlastností s vlastnostmi obrácenými.

Popisy v češtině, které jsou kurzívou, jsou pracovními překlady vytvořenými pro účely této metodiky; ostatní popisy v češtině (stejně jako popisy v angličtině) jsou převzaty přímo z Wikidat.

### 5.3.1 Symetrické vlastnosti

Vlastnost	<b>different from – rozdílné od</b>
Popis v angličtině	item that is different from another item, but they are often confused
Popis v češtině	položka, která je rozdílná od jiné položky, ale jsou často zaměňovány
URL	<a href="https://www.wikidata.org/wiki/Property:P1889">https://www.wikidata.org/wiki/Property:P1889</a>

Vlastnost	<b>opposite of – protiklad</b>
Popis v angličtině	item that is the opposite of this item
Popis v češtině	položka, která je protikladem této položky
URL	<a href="https://www.wikidata.org/wiki/Property:P461">https://www.wikidata.org/wiki/Property:P461</a>

### 5.3.2 Vlastnosti bez obrácených vlastností

Vlastnost	<b>facet of – aspekt (čeho)</b>
Popis v angličtině	topic of which this item is an aspect, item that offers a broader perspective on the same topic
Popis v češtině	<i>téma, jehož je tato položka aspektem, položka, která nabízí širší pohled na shodné téma</i>
URL	<a href="https://www.wikidata.org/wiki/Property:P1269">https://www.wikidata.org/wiki/Property:P1269</a>

Vlastnost	<b>field of work – zaměření</b>
Popis v angličtině	specialization of a person or organization; see P106 for the occupation
Popis v češtině	oblast hlavního zaměření této položky (např. fyzika, historie); pro zaměstnání (např. fyzik, historik) použijte P106; lze používat i u osob
URL	<a href="https://www.wikidata.org/wiki/Property:P101">https://www.wikidata.org/wiki/Property:P101</a>



Vlastnost	<b>instance of – instance (čeho)</b>
Popis v angličtině	that class of which this subject is a particular example and member. (Subject typically an individual member with Proper Name label.) Different from P279 (subclass of)
Popis v češtině	tato položka je jedna konkrétní věc (exemplář, příklad) patřící do této třídy, kategorie nebo skupiny předmětů
URL	<a href="https://www.wikidata.org/wiki/Property:P31">https://www.wikidata.org/wiki/Property:P31</a>

Vlastnost	<b>subclass of – nadtřída<sup>49</sup></b>
Popis v angličtině	all instances of these items are instances of those items; this item is a class (subset) of that item. Not to be confused with P31 (instance of)
Popis v češtině	třída, jejíž podtřídou/podskupinou tato třída je (všechny prvky této množiny jsou také prvky oné množiny)
URL	<a href="https://www.wikidata.org/wiki/Property:P279">https://www.wikidata.org/wiki/Property:P279</a>

Vlastnost	<b>use – použití</b>
Popis v angličtině	main use of the subject (includes current and former usage)
Popis v češtině	hlavní použití předmětu
URL	<a href="https://www.wikidata.org/wiki/Property:P366">https://www.wikidata.org/wiki/Property:P366</a>

### 5.3.3 Vlastnosti s obrácenými vlastnostmi

Vlastnost	<b>field of this occupation – obor tohoto povolání</b>
Popis v angličtině	activity corresponding to this occupation (use only for occupations - for people use Property:P101, for companies use P452)
Popis v češtině	<i>činnost odpovídající tomuto povolání (používejte pouze pro povolání – pro osoby používejte vlastnost P101 a pro organizace P452)</i>
Obrácený vztah	practised by – praktikováno (kým)
URL	<a href="https://www.wikidata.org/wiki/Property:P425">https://www.wikidata.org/wiki/Property:P425</a>

<sup>49</sup> Viz též pozn. 36.

Vlastnost	<b>follows – předchozí</b>
Popis v angličtině	immediately prior item in a series of which the subject is a part [if the subject has replaced the preceding item, e.g. political offices, use "replaces" (P1365)]
Popis v češtině	bezprostředně předcházející položka v řadě, jíž je předmět součástí; pokud předchozí položku nahradila, použijte P1365
Obrácený vztah	followed by – následující
URL	<a href="https://www.wikidata.org/wiki/Property:P155">https://www.wikidata.org/wiki/Property:P155</a>

Vlastnost	<b>has cause – příčina</b>
Popis v angličtině	underlying cause, thing that ultimately resulted in this effect
Popis v češtině	hlavní důvod, který vedl k tomuto následku
Obrácený vztah	has effect – příčina (čeho)
URL	<a href="https://www.wikidata.org/wiki/Property:P828">https://www.wikidata.org/wiki/Property:P828</a>

Vlastnost	<b>part of - část (čeho)</b>
Popis v angličtině	object of which the subject is a part. Inverse property of "has part" (P527). See also "has parts of the class" (P2670)
Popis v češtině	subjekt je součástí tohoto objektu; opak vlastnosti "skládá se z"
Obrácený vztah	has part – skládá se z
URL	<a href="https://www.wikidata.org/wiki/Property:P361">https://www.wikidata.org/wiki/Property:P361</a>

Vlastnost	<b>product or material produced – produkuje</b>
Popis v angličtině	material or product produced by a government agency, business, industry, facility, or process
Popis v češtině	materiál, který produkuje důl, továrna nebo jiný produkční proces
Obrácený vztah	manufacturer – výrobce
URL	<a href="https://www.wikidata.org/wiki/Property:P1056">https://www.wikidata.org/wiki/Property:P1056</a>

Vlastnost	<b>studies – studuje (co)</b>
Popis v angličtině	subject item is the academic field studying the object item of this property
Popis v češtině	předmět studia příslušné vědy či nauky
Obrácený vztah	studied by – studováno (čím)
URL	<a href="https://www.wikidata.org/wiki/Property:P2578">https://www.wikidata.org/wiki/Property:P2578</a>

Vlastnost	<b>uses – používá</b>
Popis v angličtině	item or concept used by the subject or in the operation
Popis v češtině	položka používaná subjektem
Obrácený vztah	used by – uživatel
URL	<a href="https://www.wikidata.org/wiki/Property:P2283">https://www.wikidata.org/wiki/Property:P2283</a>

Vlastnost	<b>replaces – předchůdce</b>
Popis v angličtině	person or item replaced. Use P1398 (structure replaces) for structures. Use P155 (follows) if the previous item was not replaced or if predecessor and successor are identical.
Popis v češtině	osoba nebo položka, kterou tato položka nahradila; pro stavby použijte P1398; pokud nešlo o nahrazení, použijte P155
Obrácený vztah	replaced by – nástupce
URL	<a href="https://www.wikidata.org/wiki/Property:P1365">https://www.wikidata.org/wiki/Property:P1365</a>

Vlastnost	<b>based on – podle</b>
Popis v angličtině	the work(s) used as the basis for subject item
Popis v češtině	díla, která byla základem
Obrácený vztah	derivative work – odvozené dílo
URL	<a href="https://www.wikidata.org/wiki/Property:P144">https://www.wikidata.org/wiki/Property:P144</a>

Při zpracování návrhů vztahů se jako účelné ukázalo využití obrácených vztahů i v případech, kdy ve Wikidatech takové vztahy neexistují. Do Wikidat tyto obrácené vztahy pochopitelně nepatří, ale při zpracování a úvahách o vhodných vztazích mezi hesly mohou být užitečné.

Bude-li v budoucnu třeba využít dalších vlastností, je možné je (nejlépe v angličtině) vyhledat na výše zmíněné stránce Wikidata:List of properties.

V případě nejasností apod. lze kontaktovat spolek Wikimedia Česká republika<sup>50</sup>, popř. využít komunikaci prostřednictvím již zmíněné facebookové skupiny Wikidata CS.

## 5.4 Přehled propojených položek

Aktuální přehled propojených položek je možné získat pomocí dotazu zadaného na adrese <https://query.wikidata.org/>.

Dotaz, který vypíše identifikátory a názvy všech položek, k nimž je přidáno propojení na heslo v databázi TDKIV, má následující podobu:

```
select ?item ?itemLabel ?tdkiv where {
?item wdt:P5398 ?tdkiv .
SERVICE wikibase:label { bd:serviceParam wikibase:language "cs,en". }
}
```

Je také možné použít odkaz <http://tinyurl.com/ybzmuykd>.

## 6 Možnosti dalšího využití zpracovaných dat

Realizace projektu připravila půdu pro další aktivity, které – budou-li uskutečněny – přispějí k dalšímu vylepšení a obohacení databáze:

- po přidání všech identifikátorů z databáze TDKIV do systému Wikidata bude možné přes identifikátor z databáze TDKIV využít propojení na položky Wikidat prostřednictvím tzv. Hubu<sup>51</sup> (příklad: <https://tools.wmflabs.org/hub/P5398:00000543?s=wd> – zvýrazněná část je identifikátor z databáze TDKIV);
- po zpracování všech dat (nejlépe s využitím hromadného importu) bude možné uvést zpřesněné vztahy přímo v databázi TDKIV;
- bude možné dále rozvíjet vytvořenou pracovní aplikaci sloužící k vizualizaci návrhů zpřesněných vztahů;
- výstupy umožní provádění analýz dat včetně jejich vizualizace s využitím dotazovacího prostředí Wikidata Query Service a nadstavbových vizualizačních nástrojů, např. Wikidata Visualization<sup>52</sup>.

50 Kontakt. *Wikimedia Česká republika* [online]. Praha: Wikimedia ČR, Stránka byla naposledy editována 14. 8. 2018 v 10:56 [cit. 2018-11-25]. Dostupné z: <https://www.wikimedia.cz/web/Kontakt>

51 Hub. *Wikimedia Toolforge* [online]. [cit. 2018-11-25]. Dostupné z: <https://tools.wmflabs.org/hub/>

52 Wikidata Visualization. *Wikimedia Toolforge* [online]. [cit. 2018-11-25]. Dostupné z: <https://tools.wmflabs.org/dataviz/>

Na obr. 8 (aktuálním k 24. 11. 2018) je pro ilustraci uveden příklad vizualizace s využitím nástroje Wikidata Visualization. Jedná se o třicet položek z Wikidat, které jsou propojeny na hesla v databázi TDKIV a zároveň mají přiřazeno nejvíce nadtříd.

Vizualizace byla vytvořena s využitím dat získaných následujícím dotazem:

```
SELECT ?polozkaLabel (COUNT(?nadtrida) AS ?pocet_nadtrid) WHERE {
  ?polozka wdt:P5398 ?item.
  ?polozka wdt:P279 ?nadtrida.
  FILTER EXISTS {?polozka wdt:P5398 ?item}
  SERVICE wikibase:label { bd:serviceParam wikibase:language "cs,en". }
}
GROUP BY ?polozkaLabel
ORDER BY desc(?pocet_nadtrid)
LIMIT 30
```



Obr. 8: Třicet položek propojených na hesla TDKIV s nejvíce nadtřídami

## 7 Shrnutí

Předložená metodika vychází z praktických zkušeností získaných při realizaci *Pilotního projektu zpřístupnění TDKIV v podobě propojených dat*. Obsahuje tedy prozatímní poznatky, které budou v následujícím období dále zpřesňovány.

Realizace projektu ukázala několik zásadních skutečností:

- s ohledem na smysluplnou propojitelnost dat je vhodné přejít z pojetí hesel orientovaných na termíny na pojetí hesel orientovaných na pojmy;
- zpracování dat z různých úhlů pohledu a v konfrontaci s vnějšími informačními zdroji přineslo podle očekávání řadu dílčích podnětů týkajících se možného vylepšení znění jednotlivých hesel v databázi TDKIV;
- ač je systém Wikidata naplňován daty z heterogenních zdrojů a na jeho tvorbě se podílejí tisíce dobrovolníků, jsou jednotlivé položky (pokud obsahují více údajů) většinou velmi dobře zpracovány;
- databáze TDKIV má díky zpřesňování vazeb mezi hesly a vstupu do prostředí sémantického webu potenciál výrazně zvýšit svou vzdělávací hodnotu a přispět ke kultivaci odborné terminologie v oblasti knihovnictví a informační vědy.

Se získanými poznatky a také již zpracovanými daty se bude ještě dále pracovat, aby se databáze TDKIV rozvíjela v souladu s očekávanými a potřebami svých uživatelů a zároveň s využitím moderních nástrojů sémantického webu.