



WEBARCHIV PRÁVNÍ ANALÝZA

© Národní knihovna, Lucie Straková | v. 11 | červen 2019

OBSAH

1.	POPIS SITUACE A IDENTIFIKOVANÉ PROBLÉMY	3
2.	ZÁKLADNÍ VÝCHODISKA.....	4
2.1	Fungování Webarchivu: Faktická činnost.....	5
2.2	Fungování Webarchivu: Právní základ.....	5
	Knihovní fond	5
	Výjimky a omezení autorského práva: Knihovní licence	7
	Knihovní licence – Aplikace případ webarchivu.....	7
3.	Sdílení dat v rámci projektu	9
3.1	Creative Commons.....	10
3.2	Užití licence Creative Commons na webu Webarchivu	10
3.3	Zpřístupňování fondů „na místě samém“	12
3.4	Sdílení v režimu veřejných licencí	13
3.5	Možnost pořízení tiskové rozmnoženiny pro osobní potřebu fyzické osoby či vlastní vnitřní potřebu právnické osoby.....	13
3.6	Text and Data Mining	14
	Možnost datové analýzy fondů Národní knihovny	14
	Text and data mining ze strany třetí osoby	15
	Nová evropská legislativa v oblasti TDM	16
4.	ZÁVĚRY A ODPOVĚDI NA ZADANÉ OTÁZKY.....	18
	DOPORUČENÍ:.....	21
5.	Zdroje a literatura	22
	Příloha č. 1 - DŮLEŽITÉ POJMY.....	23

1. POPIS SITUACE A IDENTIFIKOVANÉ PROBLÉMY

Právní otázky, na které narazil realizační tým projektu „**Vývoj centralizovaného rozhraní pro vytěžování velkých dat z webových archivů**“, lze rozdělit do tří oblastí.

- V první oblasti jde o **sdílení archivních dat** s partnery projektu.
- Druhá oblast je zaměřena na **možnosti zpřístupnění metadat uživatelům**.
- V poslední oblasti jde o **zpřístupnění archivních dat pro výzkumné účely**.

Odpovědi na otázky v první oblasti jsou esenciální otázkou pro další směřování projektu. Partneři (Sociologický ústav AV ČR a Západočeská univerzita v Plzni) projektu potřebují mít přístup k datům NK ČR, aby mohli s daty pracovat přísně na výzkumné a vědecké bázi (tj. bez jakéhokoli komerčního využití či šíření třetím osobám) **na vlastní nebo sdílené infrastruktuře**, kterou NK ČR nedisponuje.

Hlavním problémem je, jakým způsobem může NK ČR sdílet data z webového archivu spolu s partnery projektu, tak aby s nimi mohl být prováděn výzkum. Jedná se o **plný přístup k datům webového archivu**, který je nyní, pro běžné uživatele, tj. ne pro vědeckou komunitu, možný jen v zabezpečeném režimu v NK ČR (na místě samém).

Další oblast se již týká výstupů projektu, které mají za cíl **sdílet a zpřístupnit uživatelům metadata**. Vycházíme z toho, že nám autorský zákon neumožňuje poskytnout veřejnosti plný přístup do archivu a tak chceme uživatelům nabídnout metadata. Metadata v tomto případě mohou být informace, které jsou vytvářeny v rámci procesu archivace, metadata extrahovaná z webových stránek nebo metadata vytvořená analýzou archivních dat. Příklady metadat mohou být například:

- názvy stránek, titulky
- krátké textové výňatky z obsahu
- hypertextové odkazy mezi stránkami
- frekvence slov v textovém obsahu a další obsahové analýzy
- strojově určená témata stránek
- meta tagy
- technické parametry stránky (např. znaková sada, verze HTML...)
- vlastník stránky
- odpovědi serveru
- seznamy archivovaných URL adres
- odhadovaný čas publikace/stažení stránek

Poslední oblast se týká zveřejnění archivních dat potenciálním uživatelům webového archivu. V této oblasti již máme právní analýzy z minulosti a dle nich máme nastaven současný přístup k datům webového archivu, tj. pouze z prostor NK ČR v zabezpečeném režimu. Zajímalo by nás, jestli by nebylo možné za určitých podmínek, např. pro vědecké účely, zpřístupnit archiv nebo jeho část badatelům, např. univerzitám, vědcům atd.,

například na smluvní bázi, která by na základě sankcí zamezovala komerčnímu použití či dalšími sdílení třetím osobám.

2. ZÁKLADNÍ VÝCHODISKA

Archivovaný typ dat:

Příklad je dostupný např. zde:

<https://wayback.Webarchiv.cz/wayback/20110628122156/http://www.nkp.cz/>

Pokud je to technicky možné, tak se prostřednictvím Webarchivu archivuje veškerý obsah, který je veřejně přístupný (tzn. není potřeba přihlášení, registrace nebo lidská interakce – to platí třeba pro vyhledávače nebo různé databáze) a to nezávisle na datovém formátu, tj. včetně videí, obrázků, dokumentů a ostatních formátů.

Smlouva s nositeli práv je dostupná na webarchiv.cz/cs/smlouva

Kolektivní správa: S kolektivními správci Webarchiv v současné době nekomunikuje

Metadata: Obsahem metadat může být prakticky cokoliv, co se dá extrahovat z archivních dat, záleží na potřebě potenciálních badatelů. Ve Webarchivu je momentálně přes devět miliard digitálních objektů na více než čtyřech milionech unikátních domén. V případě, že extrahujeme seznam všech domén, předpokládáme, že tento seznam můžeme zpřístupnit. Pokud extrahujeme propojení mezi devíti miliardami digitálních objektů, obsahovou analýzou vytvoříme katalog témat nebo vytvoříme seznam autorů či data publikování dokumentů, tak to bude podobné knihovnímu katalogu. Když z dat extrahujeme všechny texty z webových stránek a uděláme z nich jazykový korpus, můžeme ho zpřístupnit nebo to už je za hranicí? Můžeme vyextrahovat všechny obrázky, provést jejich analýzu a vytvořit nad nimi vyhledávač, nebo převést audio na text atd.?

2.1 FUNGOVÁNÍ WEBARCHIVU: FAKTICKÁ ČINNOST

Webarchiv sklízí do svého archivu na základě bohemistického charakteru URL adres s doménou .cz a dle specifických kritérií¹. Mezi kritéria archivace patří následující:

- místem vydání je Česká republika
- stránky jsou v českém jazyce
- autor stránek je české národnosti
- obsah se týká České republiky nebo českého národa

Důvodem je mj., že množství dokumentů, které vzniká v českém prostoru, má výhradně digitální formu.² Vzhledem k přirozené funkci internetu a digitálního prostoru neustále se měnit a reagovat na aktuální podněty, kdy se mnoho dokumentů může stát velmi obtížně dohledatelnými, realizuje projekt Webarchiv službu, která zajišťuje archivaci stránek s vazbou k českému internetu (viz kritéria archivace výše).

2.2 FUNGOVÁNÍ WEBARCHIVU: PRÁVNÍ ZÁKLAD

KNIHOVNÍ FOND

Dle zřizovací listiny NK³ buduje knihovna mj. specializované knihovní fondy, které se sestávají z knihovních dokumentů určitého zaměření. Jako takové zaměření lze chápat i fond on-line dokumentů ve formě webových stránek. V rámci zřizovací listiny NK je jako forma doplňování fondů zahrnuto i „sklizení volně dostupného internetu“, to lze však považovat pouze za doplňkový zdroj. (Pro tuto konkrétní činnost neexistuje opora v zákoně, a to i přes unikátní postavení NK vzhledem ke zřizovateli).

V rámci prvotní analýzy bylo řešeno, zda se systém archivovaných dokumentů v rámci Webarchivu dá podřadit pod definici knihovního fondu. Ačkoliv je situace v českém právním prostředí bezprecedentní, po dalších konzultacích se příkláním (v souladu s informacemi poskytnutými Mgr. Kvasnicou a Mgr. Vozárem) k následujícímu výkladu:

Dle § 2 písm. c) zákon č. 257/2001 Sb., je *knihovním fondem organizovaný, soustavně doplňovaný, zpracováváný, ochraňovaný a uchovávaný soubor knihovních dokumentů, přičemž knihovním dokumentem je informační pramen evidovaný jako samostatná jednotka knihovního fondu knihovny.*

Webarchiv tedy musí splňovat následující kritéria:

- organizovaný

¹ <https://www.webarchiv.cz/cs/nominovat/vyber-zdroju>

² <https://Webarchiv.cz/static/www/download/collection-policy-2017.pdf>

³ <https://text.nkp.cz/soubory/ostatni/zrizovaci-listina-nk.pdf>

- soustavně doplňovaný
- ochraňovaný
- uchovávaný
- soubor knihovních dokumentů

A jednotlivé knihovní dokumenty, tedy jednotlivé uchovávané webové stránky musí být:

- evidované jako samostatná jednotka knihovního fondu.

Ke zjištění, zda jsou patřičná kritéria naplněna:

V rámci Webarchivu je **veden index** o všech elektronických záznamech, které jsou v něm obsaženy, a to pomocí cdx záznamů. Tyto záznamy jsou vedeny jak ke každému balíku dat (jednotlivé webové stránce, která tvoří jeden archivovaný celek), tak ke všem balíkům dat. V tomto případě se jedná o tříděný (abecedně) **Hlavní index**. **Integritu** jednotlivých balíků dat **je možno ověřit** pomocí MD5 kontrolních součtů.

Veškerá data, shromažďovaná ve Webarchivu, jsou přepisována na LTO archivní pásky, které slouží k **dlouhodobé archivaci** (20-30 let). Tyto pásky fungují v režimu standardního zálohování v rámci NK, budou pravidelně přetáčeny a zálohovány. Do budoucna bude Webarchiv nahrán na úložiště dlouhodobé ochrany NK ČR, které slouží k uchování a ochraňování digitalizovaných dokumentů.

Webarchiv je **soustavně doplňován**, a to každý měsíc, případně i častěji. Zároveň je minimálně jednou ročně organizována celoplošná sklizeň českého internetu.

Co se týče jednotlivých záznamů ve Webarchivu, jedná se o **samostatné položky** webových stránek (jejich jednotlivé zdrojové soubory), které jsou **samostatně ukládané** ve formátu warc, **opatřené hlavičkou a zaindexované** do indexu warc souboru a v Hlavním indexu. Současně se eviduje datum pořízení záznamu, odkaz na zdrojový dokument a odkaz místa uložení včetně pozice v místě uložení, kontrolního součtu a typologického zařazení, které vychází z mezinárodního standardu MIME Type.

Co se týče webových stránek, k nimž jsou práva Webarchivu **poskytována prostřednictvím licenční smlouvy**, jsou katalogizované a jsou součástí národní bibliografie, jako jiné typy dokumentů v knihovně.

U dalších zdrojů, tedy stránek, které jsou získávány na základě výjimky pro archivní a konzervační potřeby, není evidence tak podrobná, nicméně splňuje kritéria uvedená výše. Řešený projekt evidenci ještě prohloubí.

Vzhledem k výše provedenému rozboru tak lze Webarchiv považovat za specializovaný knihovní fond. Na základě tohoto zjištění budeme zkoumat další nakládání s archivovanými dokumenty.

VÝJIMKY A OMEZENÍ AUTORSKÉHO PRÁVA: KNIHOVNÍ LICENCE

Vzhledem k povaze Webarchivu jakožto specializovaného knihovního fondu, uvažujeme možnost aplikace tzv. knihovní licence dle Autorského zákona.

Knihovní licence je institutem řazeným mezi tzv. výjimky a omezení autorského práva.⁴ Výjimky a omezení autorského práva je nutno vykládat v mezích tzv. Bernského třístupňového testu,⁵ který v českém autorském najedeme v § 29. Tento test slouží k tomu, že výjimky a omezení autorského práva, jakožto zásah do výsostných práv autora, lze vykládat restriktivně, pouze v jeho mezích, aby nedošlo k nepřiměřenému omezení majetkových práv autorských. Třístupňový test říká, že „*Výjimky a omezení práva autorského lze uplatnit pouze ve zvláštních případech stanovených zákonem a pouze tehdy, pokud takové užití díla není v rozporu s běžným způsobem užití díla a ani jím nejsou nepřiměřeně dotčeny oprávněné zájmy autora.*“⁶

KNIHOVNÍ LICENCE – APLIKACE PŘÍPAD WEBARCHIVU

Webarchiv v současné době (tj. bez zohlednění této Analýzy) odvozuje⁷ svůj právní důvod pro získávání rozmnoženin webových stránek dostupných na českém internetu, od § 37 odst. 1 písm. a) resp. c) Autorského zákona⁸, tedy tzv. knihovní licence.

„(1) Do práva autorského nezasahuje knihovna, archiv, muzeum, galerie, škola, vysoká škola a jiné nevýdělečné školské a vzdělávací zařízení

a) zhotoví-li rozmnoženinu díla, která neslouží k přímému nebo nepřímému hospodářskému nebo obchodnímu účelu, pro své archivní a

⁴ V autorském zákoně je jim věnován zejména Díl 4, mimo knihovní licenci se jedná např. o rozmnoženinu pro osobní potřebu, citaci, úřední a zpravodajskou licenci a mnohé další.

⁵ Vychází z Bernské úmluvy, která stanovuje minimální standard ochrany v signatářských zemích.

⁶ § 29 odst. 1 AutZ.

⁷ Východiska, na kterých tato analýza staví, byla získána buď od tvůrců a vývojářů Webarchivu nebo přímo ze stránek Webarchivu. Nejedná se tedy o subsumpci, ale o podkladovou informaci pro další závěry. Veškerá omezení autorského práva je třeba vykládat restriktivně, s ohledem k oprávněným zájmům autora. Ač se tedy mohou objevit různé názory, jak přistupovat k užití autorskopravní výjimek (spojování výjimek a omezení autorského práva), tak jak je zmíněno dále v textu, omezení autorského práva dle § 37 odst. 1 písm. c) nelze na předmětný způsob užití aplikovat.

⁸ Zákon č. 121/2000 Sb., zákon o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů.

konzervační potřeby, a to v počtech a formátech nezbytných pro trvalé uchování díla,

- b) (...)
- c) *zpřístupňuje-li dílo, včetně zhotovení jeho rozmnoženiny nezbytné pro takové zpřístupnění, které je součástí jeho sbírek a jehož užití není předmětem prodejních nebo licenčních podmínek, s výjimkou sdělování díla způsobem uvedeným v § 18 odst. 2 (zpřístupňování díla veřejnosti způsobem, že kdokoli může mít k němu přístup na místě a v čase podle své vlastní volby zejména počítačovou nebo obdobnou síť, pozn. aut.,) jednotlivcům ze strany veřejnosti prostřednictvím k tomu určených technických zařízení umístěných v jeho objektech, a to výhradně pro účely výzkumu nebo soukromého studia takových osob, a zamezí-li takovým osobám zhotovit rozmnoženinu díla; ustanovení § 30a odst. 1 písm. c) a d) tím nejsou dotčena,*

S tímto titulem nastávají následující problémy (vychází z textace § 37 odst. 1 písm. a) AutZ):

1) Rozmnoženina díla pro archivní a konzervační potřeby

Rozmnoženinu pro archivní a konzervační potřebu lze vytvořit pouze a výhradně z takového druhu děl, ke kterému instituce vede svůj fond.⁹ Knihovní fond je dle odborné literatury sbírkou dokumentů, které knihovna spravuje, v případě NK jsou zdrojem zejm. povinné výtisky¹⁰, nicméně dle výše provedené analýzy povahy Webarchivu jakožto specializovaného knihovního fondu, bude taktéž spadat pod tuto výjimku z ochrany autorským právem. Rozmnoženinu může knihovna vytvořit tedy **pouze k takovému druhu děl, k nimž vede svůj fond** (viz výše, část Knihovní fond).

Využití výjimky v rámci připravovaného projektu:

Webarchiv tedy splňuje kritéria pro naplnění definice knihovního fondu, **může tedy sklízet český internet v rámci výjimky pro archivní a konzervační potřeby**. Takto zhotovené rozmnoženiny lze však využívat skutečně výhradně pro archivní a konzervační potřeby, **není možné je používat k jakýmkoliv dalším činnostem** (tedy např. na základě jiných výjimek nebo omezení autorského práva. Tedy není možné takto sklizené (archivované a konzervované) stránky sdílet s třetí osobou. Jiná situace nastává

⁹ Shodně Telec a Tůma, „...je instituce oprávněna pořizovat rozmnoženiny pouze takového druhu děl, ke kterému vede svůj fond.“ Komentář k § 37, Pořizování archivačních rozmnoženin.

¹⁰ § 3 zákona č. 37/1995 Sb., zákon o neperiodických publikacích; resp. § 9 zákona č. 46/2000 Sb., zákon o právech a povinnostech při vydávání periodického tisku a o změně některých dalších zákonů (tiskový zákon)

v případě děl, k jejichž užití byla knihovně udělena nositelem autorských práv majetkových k dílu licence (viz dále).

Zásadním problémem, na který narážíme, pokud uvažujeme sdílení obsahu Webarchivu s třetí osobou, je, že se v rámci fungování Webarchivu dostáváme do systémového problému s nastavením v rámci českého právního prostředí. V případě, že je Webarchiv specializovaným knihovním fondem, který shromažďuje na Internetu volně přístupné webové stránky, je možno takto činit výhradně za účelem archivace a konzervace takových knihovních dokumentů. NK může rozmnoženiny tedy výhradně archivovat a konzervovat, **nikoliv** užívat jakýmkoliv jiným způsobem, tedy zejména **dále zpřístupňovat** (tedy např. prostřednictvím webových stránek Webarchivu nebo pro užití v rámci výzkumných projektů). Toho lze dosáhnout výhradně v případě, že se jedná o zdrojové dokumenty (webové stránky), jejichž obsah je Webarchivu licencován, případně se jedná o zdrojové dokumenty, jejichž obsah je již licencován pod některou z veřejných licencí (a tedy Webarchiv jejich obsah může sklízet a dále užívat bez nutnosti uzavírat licenční smlouvu). I v tomto případě je však třeba respektovat podmínky veřejné licence (viz dále).

3. SDÍLENÍ DAT V RÁMCI PROJEKTU

Vzhledem k současnému v rámci českého právního řádu není možné bez dalšího zpřístupňovat, sdílet, nahrávat nebo jinak užívat obsah Webarchivu.¹¹

Je třeba opatřit si jiný právní titul – licenční smlouvu. K tomu lze užít licenční smlouvy, která je k dispozici na stránkách Webarchivu.¹² K tomu **silně doporučuji aktualizaci současné licenční smlouvy (probíhá)**. Problematickým článkem byl zejm. článek III. odst. 1, kdy poskytovatel licence uděloval Národní knihovně výhradní licenci k vytváření rozmnoženin, k trvalé archivaci a zpřístupňování (sdělování) díla, které je předmětem této smlouvy, prostřednictvím internetu.¹³ Tento článek byl již upraven, nyní je poskytována licence nevýhradní. Upravený vzor licenční smlouvy byl poskytnut ke konzultacím.

¹¹ Pozn.: Nerozporujeme zákonnou povinnost např. v případě požadované součinnosti orgány činnými v trestním řízení.

¹² <https://webarchiv.cz/cs/smlouva>

¹³ To je pro účely Webarchivu naprosto nadbytečné a je pro potenciální poskytovatele nepochybně odrazující. Taková licence jim totiž znemožňuje sdělovat dílo veřejnosti jimi samotnými. Podle této smlouvy by tak dílo měli stáhnout ze svých webových stránek a nadále jej nikde nezveřejňovat. Toto není cílem ani Webarchivu, ani provozovatele dotčených webových stránek.

3.1 CREATIVE COMMONS¹⁴

Specifickou problematikou je užívání veřejných licencí Creative Commons nositelem práv sklizené webové stránky. Inherentní vlastností licencí Creative Commons je, že dílo licencované dílo lze vždy sdílet dál, to umožňuje prvek (ač není přímo zaznačen v prvcích licence) Share. Tento prvek obsahuje každá licence Creative Commons.¹⁵

V případě, že je k obsahu webové stránky udělena licence Creative Commons, je možno stránku dále užívat v mezích, které licenční podmínky nastavují. Vždy je tedy možno sdílet obsah dále, tedy při respektování podmínek licence. Vzhledem k tomu, že sdílení díla na internetu se kvalifikuje jako jeden ze způsobů užití autorského díla, je třeba mít k takovému užití oprávnění, v tomto případě to umožňuje licence CC.¹⁶

3.2 UŽITÍ LICENCE CREATIVE COMMONS NA WEBU WEBARCHIVU

¹⁴ Více viz creativecommons.org, creativecommons.cz

¹⁵ Mezi opravňující prvky patří právo používat a šířit dílo v původní podobě, případně ho zařadit do díla souborného (Share), a právo dílo upravovat, měnit nebo doplňovat (Remix). Druhou skupinu licenčních prvků tvoří prvky omezující. Prvním z nich je nutnost uvést původ díla, která se uplatňuje vždy (Attribution, zkratka BY), dále prvek, který omezuje možnost vystavení upraveného díla pouze pod shodnou nebo slučitelnou licenci (ShareAlike, zkratka SA), může být stanovena nemožnost veškerého upravování díla (NoDerivatives, zkratka ND), případně zakázáno komerční využití (NonCommercial, zkratka NC). Tyto licenční prvky se v licencích kombinují a vytvářejí tak celkem šest používaných druhů licencí, které se odlišují množstvím práv a povinností, které s sebou nesou.

¹⁶ V případě nejliberálnější licence CC BY (Uveďte původ) požaduje licence při užití díla pouze uvedení původu díla. S nadsázkou můžeme říct, že se jedná o rozšířenou citaci. Jako pomůcka pro správné uvedení původu díla může sloužit akronym TASL (Title, Author, Source, License), v českém překladu NAZL:

- Název díla
- Autor díla
- Zdroj – odkaz na dílo – obvykle se bude jednat o URL odkaz, ze kterého bylo k dílu přistoupeno.
- Licence – název licence a odkaz na ni – nepostačuje uvést, že se jedná o licenci Creative Commons, je třeba uvést, o kterou konkrétně, a opět poskytnout k ní vedoucí URL odkaz.

Ve všech licenčních podmínkách se nachází sekce zabývající se uvedením původu, kde lze nalézt přesné požadavky dané licence. Splnit požadované podmínky lze jakýmkoli rozumným, přiměřeným způsobem, vzhledem k tomu, na jakém médiu a jakým způsobem je dílo sdíleno. Takovým způsobem se rozumí především uvedení hypertextového odkazu nebo URI (Uniform Resource Identifier), který vede na stránku s požadovanými informacemi. Mírně složitější situace nastává, pokud dochází k licencování upraveného díla. V takové situaci je nejjednodušší k atributům TASL přidat formulaci jako např. „Toto dílo je odvozeno od...“, případně, pokud dané dílo obsahuje více takových prací, pak „Toto dílo obsahuje materiál z následujícího seznamu děl...“, přičemž opět je možné využít odkazu. K licencování upravených děl však pravděpodobně nebude v rámci Webarchivu docházet.

Na webové stránce Webarchivu je v patičce uvedeno, že Webarchiv je pod licenci Creative Commons.

[Webarchiv](#) je pod licenci:

[CC BY/NC/ND](#)

Koncept a grafickou podobu připravili

[Bohdan Heblík](#) & [Luboš Svoboda](#)

Do chodu vše uvedl [Tomáš Kořínek](#)

Na podobné vyjádření je třeba dát zvláště pozor, může totiž zavdávat domněnku, že veškerý obsah Webarchivu je dostupný pod touto licenci, což by mohlo vést až k nepříjemným sporům ohledně užití obsahu dostupného na Webarchivu (který není vždy vytvořen zaměstnanci Národní knihovny), a vzhledem k tomu, že se na webu objevuje i obsah třetích stran.

V první řadě je třeba uvažovat, zda je možno obsah Webarchivu zpřístupňovat pod veřejnou licenci. Vzhledem k tomu, že zpřístupnění pod takovou licenci je právním úkonem, ke kterému je nutná možnost dispozice s autorskými majetkovými právy, není možné (vzhledem k povaze obsahu) uvést, že je „Aplikace Webarchiv“ zpřístupněna pod veřejnou licenci. To implikuje, že veškerý obsah je pod takovou licenci dostupný. Vzhledem k tomu, že NK není oprávněna neomezeně vykonávat autorská práva majetková ke všemu obsahu, který je na webových stránkách Webarchivu, není ani možné uvést takovéto prohlášení. NK by se dostávala do situace, kdy neoprávněně zveřejňuje autorský obsah třetích stran (což by mohlo vést k odpovědnosti za přestupky nebo trestné činy na poli autorského práva).

Je třeba rozlišit následující:

- 1) Obsah, který je autorským dílem NK, resp. jejích zaměstnanců
- 2) Obsah, který již je zveřejněn pod veřejnou licenci
- 3) Obsah, který má NK svolení zveřejnit pod veřejnou licenci
- 4) Obsah, který nesplňuje ani jedno z výše uvedeného

V prvních třech případech je možno dále zveřejňovat pod veřejnou licenci, pokud tak NK v souladu se zákonem rozhodne (1), pokud budou zachovány licenční podmínky, které ukládá daná veřejná licence (2), pokud tak stanoví licenční smlouva mezi NK a nositelem majetkových práv autorských k danému dílu (3). V případech, kdy všechno obsah na stránce není dostupný v jednom režimu, nedoporučuje se uvádět podobná prohlášení v patičce stránky, neboť, jak zmiňuji výše, to zavdává domněnku, že se taková informace vztahuje na veškerý obsah na webové stránce dostupný.

Osvědčeným postupem je v podobných případech uvádět licenci u každého jednotlivého dokumentu. Tj. u každého zdroje doplnit licenci, pod kterou je zdroj dostupný. Je to o to důležitější, že často ani z původního dokumentu není jasné, pod jakou licenci je dostupný

Zároveň zmiňované uvedení licence jako CC BY-NC-ND je nedostačující, existuje několik vývojových variant licencí, je třeba tedy specifikovat, zda se jedná (nejčastěji) o verzi 3.0 nebo 4.0. U verzi 3.0 je potom třeba uvést i lokalizovanou, případně mezinárodní doložku (CZ/International). Korektní uvedení licence je v tomto konkrétním případě „Dokument je dostupný pod licencí [CC BY-NC-ND 4.0 Mezinárodní](#)“.¹⁷

Copyrightovou výhradu (©) je možno do patičky uvést, opět je vhodné specifikovat, co pod ni spadá. Obecně se totiž používá v případech, kdy chce provozovatel stránky (a vykonavatel majetkových práv) dát návštěvníkům najevo, že webová stránka podléhá klasickému autorskoprávnímu režimu, tj. obsah je možno užívat pouze za podmínek AutZ v rámci zákonných výjimek a omezení, případně po udělení souhlasu s užitím.

Nejdůležitější je v tomto případě srozumitelně formulovat, co chce NK návštěvníkům svých stránek sdělit. Může se jednat i o delší prohlášení, případně může být v patičce uveden odkaz na podrobnější popis možností užití ve formě samostatného dokumentu. Takové delší prohlášení může být součástí stránek do budoucna, v současní době se ale nejedná o prioritu.

3.3 ZPŘÍSTUPŇOVÁNÍ FONDŮ „NA MÍSTĚ SAMÉM“

Na základě ustanovení čl. 5 odst. 3 směrnice 2001/29/ES o harmonizaci určitých aspektů práva autorského a práv s ním souvisejících v Informační společnosti stanovil autorský zákon novelizací v roce 2006 nový rozměr bezúplatné zákonné knihovní licence, a to ve formě **terminálového zpřístupnění fondů**. Podmínkou, kterou zákon ukládá, je, že účelem zpřístupnění je výhradně soukromé studium fondů veřejností, případně výzkumné účely, a to **výhradně** v prostorách (Národní) knihovny.

Co se týče zpřístupňování obsahu třetích osob, respektive obsahu, získaného Národní knihovnou na **základě obsahu knihovní licence, jedná se o užití, ke kterému je třeba souhlasu nositele práv**. Souhlas nositele práv je potřebný nezávisle na tom, zda se jedná o obsah zpřístupňovaný jiným subjektem on-line Národní knihovně na základě licenční smlouvy nebo obsah, který byl získán Národní knihovnou v rámci její archivační činnosti – tj. zejména v rámci činnosti Webarchivu.

Zákon současně umožňuje, aby součástí procesu zpřístupnění na terminálu bylo vytvoření rozmnoženiny díla, pokud je takový krok nezbytný pro zpřístupnění na terminálu. Knihovna je však současně povinna zajistit, aby si uživatel takto

¹⁷ V současné době již je k dispozici oficiální český překlad licencí Creative Commons v. 4.0.

zpřístupněného díla nemohl sám pořídit rozmnoženinu takto, na terminálu zpřístupněného díla.

Zpřístupňování v této podobě je možno pouze v případě, že nositel majetkových práv udělil Národní knihovně buď **licenci k dalšímu šíření díla**, nebo je obsah dostupný pod některou z **veřejných licencí** Creative Commons (nebo obdobnou licenci).

Aby bylo možno obsah Webarchivu terminálově zpřístupňovat, je třeba, aby bylo dílo součástí sbírek NK (avšak nikoliv jako rozmnoženina pro archivační a konzervační potřeby!) nebo na základě licenční smlouvy (včetně veřejných licencí).

3.4 SDÍLENÍ V REŽIMU VEŘEJNÝCH LICENCÍ

Všechny veřejné licence vždy umožňují sdílení licencovaného materiálu, omezením je však nutnost uvádět při takovém sdílení původ díla ve formě „jméno autora, název díla, odkaz na zdroj díla a odkaz na licenci“. Bez uvedení těchto údajů nejsou splněny podmínky licence a sdílení není možné. Nelze tedy zcela jednoduše aplikovat na případy hromadného sdílení materiálů – nelze plošně licencovaný materiál sdílet, licence Creative Commons i další veřejné licence často vyžadují jako jednu z mála podmínek uvedení původu díla (případně jiné podmínky). Nelze tedy materiál „překlopit“ na Webarchiv bez dalšího, je třeba u každého jednotlivého materiálu uvést původ díla, což vyžaduje buď manuální práci nebo strojovou extrakci potřebných informací z materiálu (a případnou alespoň manuální kontrolu.) Takové užití by vyžadovalo vytvoření seznamu původu zahrnutých děl (nebo obdobný systém.)

Zcela nesprávné je pojetí, že v případě, že je licence uvedena v původním zdrojovém dokumentu, již není potřeba ji uvádět na místě, kde je takový dokument sdílen. V případě, že je dále šířen obsah, který je původně licencován pod CC (většinou i jinou veřejnou licenci), je třeba vždy uvést původ díla, neboť se jedná o autorskoprávní užití díla, uvedení původu je nezbytnou součástí oprávnění dílo užívat. Toto je navíc jedinou povinností, kterou mají všechny licence Creative Commons společnou. Její porušení zakládá odpovědnost za neoprávněné užití díla (pokud není dílo užitó v rámci výjimek a omezení autorského práva).

3.5 MOŽNOST POŘÍZENÍ TISKOVÉ ROZMNOŽENINY PRO OSOBNÍ POTŘEBU FYZICKÉ OSOBY ČI VLASTNÍ VNITŘNÍ POTŘEBU PRÁVNICKÉ OSOBY

Možnost pořízení tiskové rozmnoženiny pro osobní potřebu fyzické osoby či vlastní vnitřní potřebu právnické osoby **nelze** opět **aplikovat na rozmnoženiny pořízené pro archivační a konzervační potřeby**.

Autorský zákon umožňuje, aby knihovna vytvořila pro zájemce na objednávku tiskovou rozmnoženinu díla zpřístupňovaného knihovnou (a to i díla zpřístupňovaného v digitální formě). Rozmnoženina musí být pořízena pro *vlastní vnitřní potřebu* uživatele (právnická osoba) nebo pro osobní potřebu (fyzická osoba). V tomto případě je tedy možno vytvořit pro partnery projektu rozmnoženinu děl, nicméně pouze ve formě tiskové rozmnoženiny, což vzhledem k množství materiálu, představuje patrně nerealizovatelnou možnost. Zde však nastávají problémy s kolektivními správci a nutností platit poplatky – opět velká administrativní náročnost. Dle informací od NK jsou s kolektivními správci (zejm. Díla v tomto případě) v současné době vedena jednání v jiné oblasti, vznesení této otázky tak není teď příliš politicky vhodné.

3.6 TEXT AND DATA MINING

Předmětem zájmu je obsah Webarchivu, který je chráněný autorským právem (pomíjíme díla, která jsou vyňata z autorskoprávní ochrany). Skládá se z materiálů, ke kterým byla buď udělena licence na základě „klasické“ licenční smlouvy, materiálů, které jsou dostupné pod licenci Creative Commons (jinou veřejnou licenci) a materiálů, ke kterým žádná licence udělena nebyla (tedy byly získány z titulu výjimek a omezení autorského práva). První dvě možnosti jsou pro další fungování ideální.

MOŽNOST DATOVÉ ANALÝZY FONDŮ NÁRODNÍ KNIHOVNY

Datovou analýzou je, dle obecně přijímané definice, která slouží jako východí pro potřeby tohoto textu, *automatizované zpracování digitálních materiálů, které mohou zahrnovat jakékoliv typy dat, za účelem odhalení nových poznatků a porozumění*.¹⁸

V případě, že na datových zdrojích neváznou žádná práva třetích osob (zejm. díla volná, tj. díla, od smrti jejichž autora uplynulo více než 70 let, a dále pak díla, která jsou vyňata z ochrany autorským zákonem, zejm. úřední dílo, pamětní knihy, výtvořky tradiční lidové kultury). V takovém případě nepředstavuje realizace datové analýzy žádný (právní) problém. V tomto případě tedy není třeba získávat oprávnění k realizaci analýzy od vykonavatele majetkových práv k analyzovaným zdrojům.

¹⁸ Definice vycházející z: TRIAILLE, Jean-Paul et al. *Study on the legal framework of text and data mining (TDM)*. Luxembourg: Publications Office, 2014, s. 17–18.

Obecně se budou zdroje shromážděné Národní knihovnou považovat za databázi v případě, že se jedná o soubor děl, údajů nebo jiných nezávislých prvků, které jsou systematicky nebo metodicky uspořádány a které jsou jednotlivě přístupné elektronickými nebo jinými prostředky.¹⁹

Záleží pak na zpracování databáze, zda se na ni bude (kromě *sui generis* databázových práv, tedy ochrany investic pořizovatele databáze před neoprávněným vytěžováním nebo zužitkováním) vztahovat i ochrana autorským právem (která se bude na databázi vztahovat v případě tvůrčího vkladu při uspořádávání databáze).²⁰ Tato práva, pokud vzniknou, náleží právě Národní knihovně - jedná se o další vrstvu ochranných autorských práv, která se k datům vztahují.²¹ Autorským právem mohou být chráněny jak jednotlivé záznamy v databázi, tak databáze jako celek.

Nicméně vzhledem k vůli Národní knihovny zpřístupňovat své fondy, budeme pro účely této analýzy uvažovat, že případná nutnost umožnit vytěžování či zužitkování databáze (tedy užívání a analyzování dat v ní obsažených, zhotovení „normalizované rozmnoženiny“ nebude ze strany Národní knihovny představovat komplikace. Právě zhotovení normalizované rozmnoženiny díla je z hlediska TDM a autorského práva problematická. Datová analýza obvykle zahrnuje rozmnožení alespoň části dat, nad kterými je prováděna. Takto zhotovované rozmnoženiny nejsou uchovávány pro další užití (nad rámec nezbytné doby pro provedení datové analýzy).

TEXT AND DATA MINING ZE STRANY TŘETÍ OSOBY

a) TDM ze strany osoby, která nemá právní titul k provádění TDM

Směrnice InfoSoc²² hovoří o mandatorním omezení dovolující „dočasné úkony rozmnožení“. Takové rozmnoženiny jsou z definice krátkodobé nebo příležitostné a slouží k realizaci technického procesu. Jejich výhradním účelem je umožnit: a) přenos v rámci sítě mezi třetími stranami uskutečněný zprostředkovatelem nebo b) oprávněné užití díla nebo jiného předmětu ochrany. Tyto rozmnoženiny nesmí mít žádný samostatný hospodářský význam. V případě datové analýzy ovšem budou mít

¹⁹ Čl. 1 odst. 2 směrnice Evropského parlamentu a Rady 96/9/ES ze dne 11. března 1996 o právní ochraně databází a § 2 odst. 2 AutZ a § 88 a násl. AutZ.

²⁰ Databázová ochrana je dvojkolejná, databáze mohou být chráněny jak autorským právem, tak v rámci zvláštního režimu (*sui generis* databázová práva)

²¹ Pozn.: Nicméně nejde zde o „autorská práva k datům“, neboť data jako taková nejsou dílem, a tedy ani předmětem ochrany autorským právem.

²² Čl. 5 odst. 1 Směrnice Evropského parlamentu a Rady 2001/29/ES ze dne 22. května 2001 o harmonizaci určitých aspektů autorského práva a práv s ním souvisejících v informační společnosti (InfoSoc směrnice).

rozmnoženiny zpravidla trvalý charakter a jejich pořízení tedy nebude kryto tímto omezením.²³

Další výjimkou, která dle InfoSoc přichází v úvahu, je užití při vědeckém výzkumu. V případě, že dojde ke kumulativnímu splnění podmínek: (a) jedná se o užití při vědeckém výzkumu, (b) je uveden zdroj včetně jména autora (je-li to možné) a to (c) v rozsahu opodstatněném sledovaným nekomerčním účelem. Je třeba zmínit, že v současné situaci, kdy neexistuje specifická úprava výjimky pro TDM, je třeba respektovat i požadavek na uvedení původu díla při užití díla v rámci výjimky z autorského práva (tj. názvu, jména autora a zdroj díla), což může být v případě TDM neuvěřitelně administrativně i finančně náročné. To se týká všech autorských děl, která jsou v rámci TDM analyzována.

Uvedené závěry opět nelze aplikovat na rozmnoženiny získané na základě licence pro archivační a konzervační účely. Jak bylo uvedeno v úvodu, jedná se o díla, která jsou dostupná buď na základě licenční smlouvy, veřejné licence nebo jako součást sbírek Národní knihovny.

NOVÁ EVROPSKÁ LEGISLATIVA V OBLASTI TDM

V rámci schválené směrnice (15. dubna 2019) o autorském právu na jednotném digitálním trhu (EU Directive on copyright in the Digital Single Market) existují dvě výjimky pro text and data mining. Pravidla, který směrnice ustanovuje, musí být transponována do národních řádů ve lhůtě 24 měsíců.

Zmiňované výjimky jsou obsaženy v článku 3 (který je zásadní pro NK) a článku 4, další informace pak najdeme zejm. v recitálech 11, 14, 15, 17 a 18. Text výjimek je následující:²⁴

Článek 3

Vytěžování textů a dat pro účely vědeckého výzkumu

1. Členské státy stanoví výjimku z práv stanovených v čl. 5 písm. a) a čl. 7 odst. 1 směrnice 96/9/ES, v článku 2 směrnice 2001/29/ES a v čl. 15 odst. 1 této směrnice s ohledem na rozmnoženiny a extrakce zhotovené výzkumnými organizacemi a institucemi kulturního dědictví za účelem vytěžování textů a dat z děl nebo jiných předmětů ochrany, k nimž mají zákonný přístup, pro účely vědeckého výzkumu.

²³ MYŠKA, Matěj a Jakub HARAŠTA. Omezení autorského práva a zvláštních práv pořizovatele databáze v případě datové analýzy. *Časopis pro právní vědu a praxi*. Brno, 2015. s. 8.

²⁴ Převzato z: <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2019-0231+0+DOC+XML+Vo//EN>

2. Rozmnoženiny a extrakce děl nebo jiných předmětů ochrany zhotovené v souladu s odstavcem 1 musí být uloženy s vhodnou úrovní zabezpečení a mohou být uchovávány pro účely vědeckého výzkumu, včetně ověření výsledků výzkumu.

3. Nositelé práv mohou uplatňovat opatření k zajištění bezpečnosti a integrity sítí a databází, v nichž jsou díla nebo jiné předměty ochrany uloženy. Tato opatření nepřekročí rámec toho, co je nezbytné k dosažení tohoto cíle.

4. Členské státy podporují nositele práv a výzkumné organizace a instituce kulturního dědictví v tom, aby k uplatňování povinností uvedených v odstavci 2 a opatření uvedených v odstavci 3 vymezily společně dohodnuté osvědčené postupy.

Článek 4

Výjimky a omezení pro vytěžování textů a dat

1. Členské státy stanoví výjimku z práv nebo omezení práv stanovených v čl. 5 písm. a) a v čl. 7 odst. 1 směrnice 96/9/ES, v článku 2 směrnice 2001/29/ES, v čl. 4 odst. 1 písm. a) a b) směrnice 2009/24/ES a v čl. 15 odst. 1 této směrnice pro rozmnoženiny a extrakce zhotovené za účelem vytěžování textů a dat z děl nebo jiných předmětů ochrany, k nimž mají zákonný přístup.

2. Rozmnoženiny a extrakce zhotovené podle odstavce 1 smějí být uchovávány pouze po dobu nezbytnou pro účely vytěžování textů a dat.

3. Výjimky nebo omezení podle odstavce 1 se použijí pod podmínkou, že si nositelé práv výslovně nevyhradili užití děl a jiných předmětů ochrany podle uvedeného odstavce vhodným způsobem, jako jsou strojově čitelné prostředky v případě obsahu zpřístupněného veřejnosti online.

4. Tímto článkem není dotčeno uplatňování článku 3 této směrnice.

Z textu přijaté směrnice vyplývá, že **členské státy budou muset zavést výjimku pro text and data mining pro výzkumné organizace a instituce kulturního dědictví**. Z toho vyplývá, že Národní knihovna, která bez dalšího spadá do kategorie výzkumných organizací²⁵ a institucí kulturního dědictví, bude moci v rámci této výjimky využívat text and data mining. Směrnice nicméně není přímo účinná, v českém právním řádu se tedy nelze odvolávat na její znění, dokud nebude výjimka zařazena do českého právního řádu nebo dokud neuplyne lhůta k implementaci.

²⁵ Čl. 2 odst. 3 CDSM Směrnice: „institucí kulturního dědictví“ se rozumí veřejně přístupná knihovna nebo muzeum, archiv nebo instituce pečující o zachování filmového nebo zvukového dědictví;

4. ZÁVĚRY A ODPOVĚDI NA ZADANÉ OTÁZKY

Generální otázka

- 1) Je možné za určitých podmínek, např. pro vědecké účely, *zpřístupnit archiv nebo jeho část* badatelům, např. univerzitám, vědcům atd., například na smluvní bázi, která by na základě sankcí zamezovala komerčnímu použití či dalšími sdílení třetím osobám?

Ačkoliv Webarchiv splňuje kritéria pro naplnění definice knihovního fondu, a **může tedy sklízet český internet v rámci výjimky autorských děl z ochrany autorským právem pro archivní a konzervační potřeby**, takto zhotovené rozmnoženiny lze využívat skutečně výhradně pro archivní a konzervační potřeby, **není možné je používat k jakýmkoliv dalším činnostem**. Tedy není možné takto sklizené (archivované a konzervované) webové stránky sdílet s třetí osobou.

Není možné zpřístupnit celý obsah Webarchiv plošně další osobě (dalšímu subjektu). Je možno zpřístupňovat **pouze materiály, ke kterým byla poskytnuta licence, která, zároveň, umožňuje další sdílení**. To může být jak Creative Commons, tak licence vytvořená speciálně pro účely Národní knihovny – dle uzavřené licenční smlouvy. Licence, vytvořená speciálně pro účely Národní knihovny je výhodná tím, že v případě, že si autor sdílených materiálů nepřeje, aby bylo jeho dílo dostupné plošně na celém internetu, může udělit licence pouze pro účely sdílení v rámci Webarchivu (nicméně pak se zde opět jedná o případ, kdy by nebylo možné poskytnout obsah k užívání třetím osobám, čili pro účelu projektu je tato možnost nevyhovující).

Pro účely projektu a sdílení třetím osobám je třeba mít k dispozici buď uzavřenou licenční smlouvu pro další užití nebo sdílená stránka musí být licencována pod některou z veřejných licencí.

Sdílení dat v rámci projektu

- 1) **Můžeme sdílet autorsky chráněná data mezi partnery projektu výhradně za účelem výzkumu a vědeckých činností (NK, SÚ, ZČÚ)?**

Pokud jde o autorsky chráněná data, podle § 2 odst. 2 a 5 Autorského zákona, tedy data, která jsou způsobem výběru nebo uspořádáním obsahu autorovým vlastním duševním výtvořem, vztahuje se na ně obdobný **režim jako na autorská díla**. V případě takové originální databáze je chráněna originální struktura, systém řazení údajů a podobně, tedy kostra, do které se samotné údaje vkládají. Tedy opět není možné sdílet je bez souhlasu nositele práv.

Pokud taková databáze nespĺňuje tyto znaky, tedy se nejedná o originální databázi, mohou k ní vznikat tzv. sui generis databázová práva, která chrání investice vynaložené

při zřízení databáze, ověření uvedených dat, nebo jejich publikaci. Práva vznikají, pokud pořízení, ověření nebo předvedení obsahu databáze představuje kvalitativně nebo kvantitativně podstatný vklad. Držitelem práv je pořizovatel databáze a těchto práv **je možno se (narozdíl od autorských práv) plně vzdát.**

K text and data mining: V současné době není možno provádět bez souhlasu nositele práv text and data mining na obsahu třetích osob. To by se však mělo změnit na základě implementace přijaté Směrnice o autorském právu na jednotném digitálním trhu. Do té doby je možné provádět TDM pouze tam, kde k takové činnosti byl udělen souhlas nebo kde jsou zdrojové soubory dostupné pod veřejnou licenci.²⁶

2) Je případně nutné uzavření nějaké další smlouvy mezi námi a partnery projektu? V jakém režimu je data možné sdílet; musí zůstat na půdě NK ČR?

Dle předchozí otázky. Data není možné plošně sdílet, a to ani po uzavření smlouvy s partnery projektu. Jediná situace, kdy je možno přistupovat k datům, která NK a Webarchiv spravují, je na základě uděleného titulu ze strany autora.

Následně je třeba takové situace řešit případ od případu - o jaká se jedná data (databázi), jak byla získána, zda jsou dostupná pod veřejnou licenci, k jakým účelům budou dále používána atp.

3) Případně bylo by možné je nahrát na zabezpečená úložiště partnerů nebo na sdílené úložiště (CESNET) a povolit přístup? Jde o čistě nekomerční využití bez dalšího šíření – pouze pro vědu a výzkum.

²⁶ Díla, na která se autorskoprávní ochrana nevztahuje, zůstávají mimo tento režim a je možno je sdílet, resp. užívat je jiným autorskoprávně relevantním způsobem. Taková díla upravuje AutZ v § 3, mimo úřední díla, u kterých je veřejný zájem, aby byla vyloučena z ochrany, se jedná také o výtvořiny tradiční lidové kultury, pokud se nejedná o anonymní díla.

Je však otázkou, zda je ekonomicky výhodné odlišovat díla vyňatá z autorskoprávní ochrany. Dle mého názoru by se jednalo o neúměrně náročnou činnost jejímu výsledku. S obsahem Webarchivu je vzhledem k jeho velikosti nakládat jako by byl tvořen výhradně autorskými díly.

Veřejné licence (včetně Creative Commons) umožňují téměř bez výjimky vždy sdílení a vytváření kopií licencovaného díla, což je nejzásadnějším atributem pro možnost s takovým dílem jakkoliv nakládat pro účely Webarchivu. Nejvhodnější licence pro dané užití je licence CC BY, která nabyvatele nejméně omezuje v nakládání s dílem. Je však třeba respektovat (v případě autorskoprávně relevantního užití, tedy např. vytváření kopií (rozmnoženin) díla, jejich úpravy a další sdílení) licenční prvek BY, který vyžaduje uvedení původu díla v rozsahu Název, Autor, Zdroj a Licence.

Ohledně jiných licencí je třeba vždy jednotlivě uvážit konkrétní licenční podmínky. Literárních nebo jiných autorských děl (mimo software) je však pod odlišnou veřejnou licenci licencováno minimum. Můžeme se setkat i s licenčními podmínkami, které neodpovídají žádné ze standardizovaných veřejných licencí, jedná se však o minoritní záležitost.

V žádném případě není možné data někam nahrávat (opět, pokud k tomu nebyl udělen souhlas), jednalo by se o přímé porušení autorského práva. V tomto případě není možné se spoléhat na účel poskytovaných materiálů – nekomerční užití k vědeckým účelům, neboť výjimky autorského práva je třeba vykládat restriktivně. Technicky může být úložiště mimo NK (mimo prostory NK), nicméně to nemění nic na možnostech užití.

Zpřístupnění a zpracování metadat

1) Základní otázkou je, co je možné zveřejnit a v jaké formě?

Metadata jako taková nejsou chráněná autorským právem, v případě, že je vytvořen index/seznam metadat, je možné taková metadata volně sdílet. Problematická je otázka jejich získávání (TDM, viz výše), případně pokud metadata budou naplňovat znaky databáze, pak vznikají buď autorská nebo *sui generis* databázová práva.

2) V případě, že je index (seznam archivovaných URL adres v archivu) na jedné straně spektra a extrakce jmenných entit na straně druhé, kde je hranice?

Hranice leží v okamžiku, kdy dochází k datové analýze, kdy software na základě zadaných kritérií prochází obrovská množství dat, pro tyto účely si dělá jejich kopie a indexuje jejich obsah. V tomto případě již dochází k text and data mining, a vzhledem k tomu, že v České republice neexistuje (prozatím) výjimka pro tento účel, je třeba postupovat výhradně v rámci výjimky pro výzkumné účely. Pro tento účel však situaci komplikuje nutnost uvádět původ *všech* analyzovaných materiálů.

Zpřístupnění dat

1) Je možné sdílet data nebo umožnit plný přístup univerzitám, vědcům atd. pro výzkumné účely? Pokud ano, tak za jakých podmínek?

Opět, zde narážíme na problémy zmiňované v prvním oddílu otázek. Vzhledem k současnému nastavení není zpřístupnění obsahu veřejnosti, který není původním obsahem NK, směrem k veřejnosti, ani pro nekomerční, výzkumné účely možné. V tomto případě neočekáváme ani žádnou změnu v rámci evropské legislativy.

DOPORUČENÍ:

- 1) Za účelem zveřejňování v online veřejně přístupném Webarchivu uzavírat s autory (nosiči práv) webových stránek sklizených do Webarchivu licenční smlouvu o užití v rámci Webarchivu. Tato smlouva musí mít doložku o možnosti NK sdílet materiál třetím stranám (alespoň) pro vědeckovýzkumné a nekomerční účely, aby bylo možno užívat tento obsah v projektu.
- 2) Evidovat stránky podle možností užití. Rozlišovat, kde se jedná o klasický autorskoprávní režim, užití na základě licenční smlouvy s NK, užití na základě licence Creative Commons, díla vyňatá z ochrany autorským právem. Zároveň umožnit možnost filtrace podle těchto kritérií.
- 3) Umožnit partnerům projektu datovou analýzu (TDM) na materiálech, které jsou dostupné na základě licenční smlouvy s NK nebo na základě licence Creative Commons.
- 4) Nevytvářet metadata pomocí datové analýzy na materiálech třetích osob bez právního titulu.

5. ZDROJE A LITERATURA

- CHALOUPKOVÁ, Helena a Petr HOLÝ. *Autorský zákon: komentář*. 4. vyd. Praha: C.H. Beck, 2012, xii, 477 s. Beckovy komentáře. ISBN 978-80-7400-432-2.
- MYŠKA, Matěj a Jakub HARAŠTA. Omezení autorského práva a zvláštních práv pořizovatele databáze v případě datové analýzy. *Časopis pro právní vědu a praxi*. Brno, 2015.
- Směrnice Evropského parlamentu a Rady 2001/29/ES ze dne 22. května 2001 o harmonizaci určitých aspektů autorského práva a práv s ním souvisejících v informační společnosti. In EUR-lex [právní informační systém]. Úřad pro publikace Evropské unie. <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32001L0029:CS:HTML>.
- Směrnice Evropského parlamentu a Rady 96/9/ES ze dne 11. března 1996 o právní ochraně databází. In EUR-lex [právní informační systém]. Úřad pro publikace Evropské unie. <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:CS:HTML>.
- Legislativní usnesení Evropského parlamentu ze dne 26. března 2019 o návrhu směrnice Evropského parlamentu a Rady o autorském právu na jednotném digitálním trhu http://www.europarl.europa.eu/doceo/document/TA-8-2019-0231_CS.html
- TELEC, Ivo a Pavel TŮMA. *Autorský zákon: komentář*. V Praze: C.H. Beck, 2007, xviii, 971 s. Velké komentáře. ISBN 978-80-7179-608-4.
- TRIAILLE, Jean-Paul et al. Study on the legal framework of text and data mining (TDM). Luxembourg: Publications Office, 2014.
- <https://Webarchiv.cz/static/www/download/collection-policy-2017.pdf>
- Zákon č. 37/1995 Sb., zákon o neperiodických publikacích;
- Zákon č. 46/2000 Sb., zákon o právech a povinnostech při vydávání periodického tisku a o změně některých dalších zákonů (tiskový zákon)
- Zákon č. 89/2012 Sb., občanský zákoník, ve znění pozdějších předpisů
- Zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů

PŘÍLOHA Č. 1 - DŮLEŽITÉ POJMY

Autorské dílo	Předmětem práva autorského je dílo literární a jiné dílo umělecké a dílo vědecké, které je jedinečným výsledkem tvůrčí činnosti autora a je vyjádřeno v jakékoli objektivně vnímatelné podobě včetně podoby elektronické, trvale nebo dočasně, bez ohledu na jeho rozsah, účel nebo význam.
Autor	Autorem je fyzická osoba, která dílo vytvořila. Autor má k dílu práva osobnostní (právo být uváděn jako autor, právo na nedotknutelnost díla, na první zveřejnění díla) a majetková (zejména právo dílo užít)
Nositel	
majetkových práv	Nositel (případně vykonavatel) majetkových práv je v základním nastavení autor. V případě zaměstnaneckého díla je (pokud není ujednáno jinak) vykonavatelem majetkových práv autorských zaměstnavatel.
Výjimky a omezení	
autorského práva	Limity autorskoprávní ochrany, které umožňují užívat dílo bez právního titulu (typicky bez smluvního oprávnění).
Text and Data Mining	Automatizované zpracování digitálních materiálů, které mohou zahrnovat jakékoliv typy dat, za účelem odhalení nových poznatků a porozumění
Databáze	Soubor nezávislých děl, údajů nebo jiných prvků, systematicky nebo metodicky uspořádaných a individuálně přístupných elektronickými nebo jinými prostředky, bez ohledu na formu jejich vyjádření. Databáze může být chráněna v případě splnění podmínek ochrannými prostředky autorského práva.
Metadata	Data, která poskytují informace o jiných datech. Vzhledem k jejich povaze nejsou chráněna jako autorské dílo.