



AKVIZICE ČESKÝCH WEBOVÝCH ZDROJŮ¹

Marie Haškovcová, Monika Holoubková, Jaroslav Kvasnica, Markéta Hrdličková

The Acquisition of Czech Web Resources

Abstract: Since human knowledge has become today an intrinsic part of the online space, an important question has emerged in relation to its preservation for future generations. Given that electronic documents hold the contemporary cultural heritage, the web archives, which seek to collect and store data on a long-term basis, are an irreplaceable source of information for studying the recent past. There is an international debate about how to make accessible the archived web pages and how to use them.

The present paper outlines various approaches to the acquisition and archiving of the Czech web sources, and further on draws attention both on the possibility of data accessibility within Czech legislation in the digital archive of Czech web resources administered by the Czech National Library and to problems related to the provision of metadata.

It examines the role of a curator in relation to electronic resources acquisition and the significance of topic collections containing data which are structured based on predetermined parameters to larger logical units. The article strives to summarize current knowledge and points out some possible approaches to web resources archiving.

Keywords: web archiving – copyright act – digital curator – legal deposit act – acquisition

Contact: Mgr. Marie Haškovcová, Národní knihovna České republiky, Klementinum 190, Praha 1; marie.haskovcova@nkp.cz
Mgr. Monika Holoubková, Národní knihovna České republiky, Klementinum 190, Praha 1; monika.holoubkova@nkp.cz
Mgr. Jaroslav Kvasnica, Národní knihovna České republiky, Klementinum 190, Praha 1; jaroslav.kvasnica@nkp.cz
Mgr. Markéta Hrdličková, Národní knihovna České republiky, Klementinum 190, Praha 1; marketa.hrdlickova@nkp.cz

Online prostor zahrnuje data většiny oblastí a aktivit lidského života, je platformou pro projevy současné kultury, místem veřejné a soukromé komunikace i sociální interakce. Dynamicky se vyvíjející sociální média propojují jednotlivce i instituce. Rychlý rozvoj komunikačních technologií umožňuje nebývalým způsobem hromadit, třídit a sdílet informace. Do popředí se tak dostává otázka jejich zachování a ochrany pro další generace. Za dobu od rozšíření internetu v 90. letech 20. století dramaticky narostlo množství elektronických dokumentů, bez jejichž studia si lze již jen těžko představit výzkum současné společnosti. Tyto dokumenty se stávají nositeli soudobého kulturního dědictví, které se webové archivy snaží uchovat.² Je zřejmé, že hodnocení událostí konce 20. a počátku 21. století se bez nich již neobejde. Vzhledem k nepředvídatelnosti a mimořádné proměnlivosti struktur dat, jejichž objem roste nevídaným tempem, přičemž stejně rychle jsou tato data měněna, přesouvána anebo úplně mazána, roste i význam archivů webového obsahu jako novodobých nástrojů uchování paměti. Stávají se tak cenným pramenem pro studium současnosti a blízké minulosti. Významná webová úložiště, knihovny a další organizace, které se zaměřují na sběr, archivaci a zajiš-

tění trvalého přístupu ke sbírkám digitálních dokumentů, sdružuje mezinárodní konsorcium IIPC – International Internet Preservation Consortium,³ jehož členem je Český webový archiv Národní knihovny ČR (dále jen Webarchiv) od roku 2007.

Důležitým úkolem webových archivů není jen sběr a dlouhodobé uchovávání dat, ale také jejich zpřístupnění a další využití. Shodují se na tom, že je potřeba zapojit odborníky různých oborů a propojovat tak jejich specifické perspektivy, vytvořit komunitu zahrnující historiky, počítačové experty, knihovníky, výzkumníky v oblasti informační vědy a další badatele, která by spolupracovala na vývoji způsobů a nástrojů pro práci s daty.⁴ Přístup k dokumentům však podstatně ovlivňují nejen limity technické, ale i právní či metodické. Zatímco používání digitalizovaných dokumentů či práce s online publikovanými odbornými zdroji se ve vědeckém prostředí staly běžnou součástí výzkumné činnosti, samotný webový obsah je pro akademickou obec coby relevantní primární zdroj stále obtížně uchopitelný. Cílem webových archivů však není jen zpřístupnění dat výzkumníkům různých oborů,⁵ ale i snaha o otevření se co nejvíce široké veřejnosti. Význam i bohatost svých dat webové archivy

¹ Realizováno v rámci institucionálního výzkumu Národní knihovny České republiky financovaného Ministerstvem kultury ČR v rámci Dlouhodobého koncepčního rozvoje výzkumné organizace.

² Julien MASANÈS (ed.), *Web Archiving*, Heidelberg 2006.

³ International Internet Preservation Consortium [online]. [cit. 2017-06-15]. Dostupné z: <http://www.netpreserve.org>.

⁴ Ian MILLIGAN, Archives Unleashed: A Series of Datathons for Cultural Heritage, Web Archives for Historians [online], 2017 [cit. 2017-07-16]. Dostupné z: <https://webarchivehistorians.org/2017/05/23/archives-unleashed-a-series-of-datathons-for-cultural-heritage/>.

⁵ Jaroslav KVASNICA – Barbora RUDIŠINOVÁ – Rudolf KREIBICH, *Vědecké využití dat z webových archivů*, Knihovna: knihovnická revue, 2016, roč. 27, č. 2, s. 24–34.

názorně ukazují mimo jiné i realizaci a prezentováním rozmanitých případových studií.⁶ Při hledání způsobů, jak archivovaný obsah smysluplně nabídnout k dalšímu užití, se začaly zaměřovat na tvorbu sbírek, které obsahují data sestavená do předem definovaných celků. Od původního zámeru spočívajícího ve sběru co největšího objemu dat, se začínají stále více přiklánět k názoru, že pro jejich další užití, analýzu a výzkumné účely mají hodnotu především data vybraná podle stanovených kritérií.

Jednou z klíčových otázek pro možnost zpřístupnění archivovaných dat se stává legislativa, přičemž evropské pojetí je značně odlišné od amerického, které je v tomto ohledu mnohem otevřenější. Oproti americkému přístupu k autorskému právu, které umožňuje digitální obsah volně zpřístupnit, dokud se proti tomu vydavatel neohradí, evropský přístup je takový, že vyžaduje nejprve jeho souhlas. České legislativě ve vztahu k Webarchivu bude věnována jedna z následujících kapitol. Podobně jako další webové archivy se proto Webarchiv soustředí i na hledání postupů a metodik, jakým způsobem ve stávajícím, značně limitujícím legislativním rámci zveřejňovat alespoň část archivu, která není autorsky chráněna, zejména nejruznější metadata. V rámci organizace OCLC – Online Computer Library Center⁷ vznikla v letošním roce pracovní skupina, která se zabývá hledáním co nejefektivnějšího zpřístupnění metadat na základě analýzy potřeb uživatelů.⁸ O možnostech zpřístupnění archivovaných dat i jejich dalšího využití se diskutuje především v rámci IIPC, které podporuje vývoj technologií, nástrojů i příslušných standardů.

Webarchiv – památník českého internetu

Hlavním cílem webových archivů je uchovávat webový obsah. Co však má být archivováno? Veškerý internet archivovat nemohou. Na co by se měly tedy webové archivy, které jsou mnohdy limitovány technicky i finančně, zaměřit? Webarchiv, který patří k nejstarším v Evropě, se coby součást Národní knihovny České republiky soustředí na archivaci národního webu. Cílí na unikátní elektronické dokumenty s dlouhodobou kulturní, vědeckou či historickou hodnotou, které mají bohemikální charakter – ať už z hlediska místního vymezení (publikované na území České republiky), jazyka (v češtině), autorství (autoři původem z Česka) nebo obsahu (pojednávající o České republice). Přístup Webarchivu konvenuje proměňující se strategii webových archivů obecně. Ve své akviziční politice tak kombinuje obě v úvodu naznačené linie. Ve snaze postihnout co nejkomplexněji český web provádí jednou až dvakrát ročně tzv. celoplošnou sklizeň. Jedná se o automatický sběr dat, kdy díky dlouholeté spolupráci s národním správcem české domény CZ.NIC archivuje všechny webové stránky s národní doménou *.cz, a představuje otisk českého webu v danou chvíli.



Obrázek 1: Katalog stránek řazených podle Konspektu. Zdroj: Webarchiv [online]. [cit. 2017-07-24]. Dostupné z: <https://www.webarchiv.cz/cs/katalog-stranek>

Druhá akviziční linie spočívá ve shromažďování datových souborů na základě kurátorského výběru (rolí kurátora a jejím významem při budování sbírek se ještě text bude důkladněji zabývat), a to dvěma způsoby.

Prvním typem jsou tzv. výběrové sklizeň koncipované na základě selekčních kritérií, jejichž cílem je snaha o postihnutí reprezentativního vzorku českého kulturního dědictví na webu napříč všemi oblastmi lidského vědění. Při budování fondů se Webarchiv opírá o metodu Konspektu,⁹ která umožňuje mezinárodně srozumitelný popis a hodnocení sbírek. Základním kritériem je předmětová kategorizace zdrojů, která spočívá v rozčlenění fondu na základě obsahových charakteristik do tematických skupin. Elektronické zdroje jsou tak tříděny do 26 předmětových kategorií a jim podřízených podkategorií, tzv. skupin Konspektu, přičemž snahou kurátorů je rovnoměrné pokrytí všech oblastí. Jsou registrovány v České národní bibliografii, při tvorbě záznamů se řídí katalogizačními pravidly RDA (Resource Description and Access) a struktura zápisu odpovídá mezinárodnímu metadatovému standardu pro popis dokumentů v knihovních informačních systémech MARC (Machine-Readable Cataloging).¹⁰ Druhý přístup k výběru zdrojů představuje budování tematických sbírek, v nichž se kurátoři snaží postihnout důležitá celospolečenská témata, která aktuálně rezonují v prostoru českého webu.

Vedle celoplošného sběru dat a kurátorsky připravovaných kolekcí se Webarchiv, podobně jako další webové archivy, vyrovnává i s problematikou sociálních médií rychle reagujících na veškeré dění. Jejich archivaci však provází nejen technologické, legislativní, ale i etické překážky. Archivy se začínají zaměřovat na přístupy, které by umožňovaly kombinovat tradiční kurátorské metody s postupy archivace sociálních médií.¹¹ Zatím neexistuje jediná cesta, jak k archivaci přistupovat, stále jsou hledány a testovány vhodné metody.

⁶ Jane WINTERS, *Breaking in to the mainstream: demonstrating the value of internet (and web) histories*, Internet Histories. Digital Technology, Culture and Society [online], 2017 [cit. 2017-07-27], Vol. 1, Iss. 1–2. Dostupné z: <http://www.tandfonline.com/doi/full/10.1080/24701475.2017.1305713>.

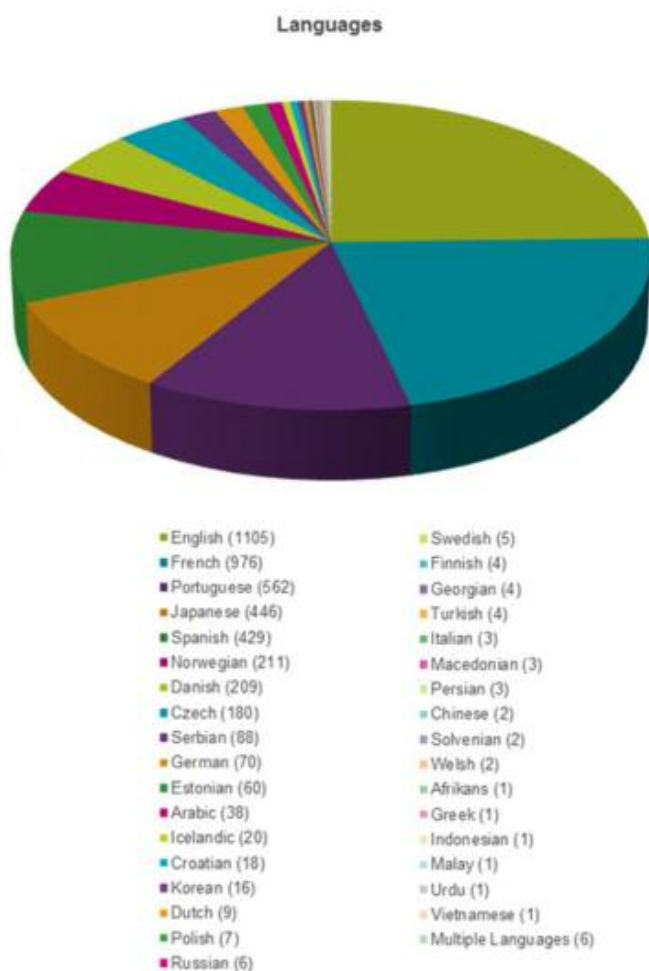
⁷ OCLC [online]. [cit. 2017-06-15]. Dostupné z: <https://www.oclc.org>.

⁸ OCLC Research. *Web Archiving Metadata Working Group* [online], 2016 [cit. 2017-08-02]. Dostupné z: <http://www.oclc.org/research/themes/research-collections/wam.html>.

⁹ Konspekt: Fondy Národní knihovny ČR. [online]. [cit. 2017-06-20]. Dostupné z: <http://konspekt.nkp.cz/>.

¹⁰ Národní knihovna České republiky. *Katalogizační politika* [online]. [cit. 2017-06-20]. Dostupné z: <https://www.nkp.cz/o-knihovne/odborne-cinnosti/zpracovani-fondu/katalogizacni-politika/>.

¹¹ Ian MILLIGAN – Nick RUEST – Jimmy LIN, Content selection and curation for web archiving: The gatekeepers vs. the masses, 2016 [cit. 2017-07-29]. Dostupné z: <http://ieeexplore.ieee.org/document/7559571/>.



Obrázek 2: Podíl jednotlivých jazyků na kolekci k Olympijským a paralympijským hrám 2016. Zdroj: IIPC: HYPERLINK "http://netpreserve.org/"Netpreserve.org [online]. [cit. 2017-07-24]. Dostupné z: <https://netpreserveblog.wordpress.com/tag/olympic-games/>

Tematické kolekce

Budování tematických souborů patří v současnosti k důležitým a i v mezinárodním kontextu progresivním akvizičním strategiím. Jejich význam pro další využití spočívá ve shromažďování relevantních zdrojů do celků podle předem definovaných kritérií. Vedle tematických sbírek, vytvářených na základě aktuálních událostí, ať už plánovaných (například volby, výročí 1. světové války, 700. výročí narození Karla IV.), nebo nenadálých (povodně, vládní krize, útok na redakci Charlie Hebdo), se Webarchiv začíná zaměřovat na vznik dlouhodobě budovaných kolekcí. Tyto kolekce nejsou podmíněné aktuálností nebo časovým ohraničením tématu – například kolekce univerzitních webů

(zdroje Univerzity Karlovy nebo Masarykovy Univerzity), soubor zdrojů vystavených pod otevřenou licenci Creative Commons a podobně. V případě aktuálních událostí se kurátoři Webarchivu snaží zachytit kromě příslušných tematických zdrojů i reflexi událostí v různých typech médií, reakce přírodních institucí a organizací či ohlasy na veřejných profilech sociálních sítí. Ve snaze o co nejúplnější obraz probíhá archivace zdrojů zpravidla v několika kolech.

Vedle vlastních tematických sbírek se Webarchiv podílí i na budování mezinárodních kolekcí iniciovaných International Internet Preservation Consortium, tzv. Collaborative collections,¹² například na vzniku sbírky věnované Olympijským a paralympijským hrám 2016 (obr 2). IIPC vyzývá členské instituce k vytvoření seznamu URL adres zdrojů, které dané téma reflektují, následně se na základě tohoto dokumentu stará o technické sklizení obsahu webů a archivní kopie zpřístupňuje na svém portálu. Výsledkem je vznik komplexních, mezinárodně relevantních sbírek s širokým pokrytím tématu. Do této kolekce¹³ se kurátoři Webarchivu pokusili zachytit, jakým způsobem zaznamenala masmédiá i specializované zdroje a profesní organizace (například Český olympijský výbor, Antidopingový výbor ČR) nejen sportovní výsledky a atmosféru her v českém prostředí, ale i jejich společensko-ekonomický kontext, sbírka zahrnuje stránky jednotlivých sportovců, jejich profily na sociálních sítích i weby sportovních svazů a asociací. Jakým způsobem jednotlivé tematické sbírky¹⁴ vznikají a jaké zdroje zahrnují, je možné nalézt na stránkách Webarchivu.

Legislativa

Z legislativního hlediska – v souladu s výkladem zákona a běžnou praxí v Národní knihovně – ovlivňuje činnost Webarchivu zákon č. 121/2000 Sb., Zákon o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů neboli Autorský zákon.¹⁵ Paragraf 37 Knihovní licence v odstavci číslo 1 písmene a Webarchivu umožňuje vytváření digitálního archivu, avšak jeho zpřístupnění veřejnosti přes internet již nikoli (v případě Webarchivu se tedy jedná o archivní kopie všech webových stránek, umístěných na doméně *.cz):

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

a) zhotoví-li **rozmnoženinou díla, která neslouží k přímému nebo nepřímému hospodářskému nebo obchodnímu účelu, pro své archivní a konzervační potřeby, a to v počtech a formátech nezbytných pro trvalé uchování díla.**^{16,17}

¹² International Internet Preservation Consortium. *Collaborative collections* [online]. [cit. 2017-05-10]. Dostupné z: <http://netpreserve.org/projects/collaborative-collections/>.

¹³ Webarchiv: památník českého internetu. *Olympijské a paralympijské hry 2016* [online]. [cit. 2017-05-10]. Dostupné z: <https://www.webarchiv.cz/cs/tematicke-kolekce/olympijske-a-paralympijske-hry-2016>.

¹⁴ Webarchiv: památník českého internetu. *Tematické sbírky* [online]. [cit. 2017-05-10]. Dostupné z: <https://www.webarchiv.cz/cs/tematicke-kolekce>.

¹⁵ ČESKO. Zákon č. 121/2000 Sb. *Zákon pro lidi.cz* [online]. © AION CS 2010-2017 [cit. 14. 7. 2017]. Dostupné z: <https://www.zakonyprolidi.cz/cs/2000-121>.

¹⁶ Tamtéž.

¹⁷ Zvýraznění provedli autoři textu.

Se zpřístupňováním veřejnosti je výklad autorského zákona složitější. Umožňuje tyto kopie webových stránek zpřístupňovat, ale pouze prostřednictvím takového technického zařízení, které se nachází ve vlastních prostorách budovy Národní knihovny a které je navíc chráněno takovým způsobem, že si z něj uživatel není schopen zajistit kopie pro vlastní potřebu. V případě Webarchivu se jedná o stolní počítač v Referenčním centru¹⁸ v budově Klementina Národní knihovny. Není možné v něm pořizovat screenshoty obrazovky, tisknout z něj ani vkládat zapisovatelné USB disky. Uživatelé mají zároveň zákaz používání fotoaparátů a focení mobilními telefony.

V případě zpřístupnění zdrojů v budově Národní knihovny zákon neukládá povinnost finančně odměňovat autory 1 296 478¹⁹ webových stránek, které se v současnosti²⁰ nacházejí na doméně *.cz. Pokud by byly zpřístupněny archivní verze těchto stránek v online podobě, musela by Národní knihovna autorům zaplatit. Kromě toho, že by se jednalo o velký finanční obnos, jsou autoři v mnoha případech nedohledatelní. Dále by bylo nutné řešit omezení a správu přístupu po technické stránce, např. pomocí Digital rights management (DRM).²¹

Kromě využití Referenčního centra, ve kterém je umožněn přístup ke všem archivovaným kopiím na doméně *.cz, mají uživatelé ještě jednu možnost, jak se dostat alespoň k některým z nich. Na stránkách Webarchivu jsou zveřejněny a zpřístupněny archivní verze několika tisíc webových stránek. Aktuálně se jedná o 5 152 online zdrojů.²² Jde o zdroje, které nutně nemusí být na doméně *.cz, ale mají již zmíněný bohemikální charakter.

Současný stav české autorskoprávní legislativy, podobně jako u dalších evropských zemí, nedovoluje online zpřístupnění archivovaných webových dokumentů bez souhlasu vydavatele. Avšak v některých z nich (např. Velká Británie,²³ Dánsko, Francie a další)²⁴ je již několik let zavedený tzv. legal deposit neboli povinný výtisk pro elektronické dokumenty. Ten je sice restriktivní pro zpřístupnění, ale upřesňuje a stanovuje práva pro akvizici webových zdrojů v dané zemi. Jejich webové archivy tak nacházejí pro archivaci částečnou oporu v zákoně a vydavatelé elektronických dokumentů jsou povinni informovat národní knihovny o svých publikacích (v začátcích probíhaly velké diskuze o tom, co přesně si představit pod pojmem elektronická publikace²⁵ – aktuálně tedy jde o elektronické publikace, které jsou často ekvivalentem publikací tištěných, ale i o born digital dokumenty²⁶). V zemích, kde je elektronický povinný výtisk zavedený a vydavatelé kurátory sami informují o vzniku nového zdroje, je tak archivace elektronických dokumentů mnohem jednodušší

a rychlejší. Bez něj kurátoři mnohdy složitě pátrají po relevantních zdrojích.

Z výše uvedeného vyplývá, že je nutné od vydavatelů získávat souhlas se zpřístupněním archivních kopií jejich zdrojů mimo budovu Národní knihovny. Lze ho získat prostřednictvím smlouvy o poskytování elektronických online zdrojů nebo využitím licence Creative Commons.

1. Smlouva

Jednou z možností, jak lze archivní verze zdrojů legálně zpřístupnit online, je uzavření licenční smlouvy s vydavatelem. Smlouva specifikuje možnosti archivace a zpřístupňování zdrojů (zpracování zdrojů probíhá za účelem veřejného prospěchu, je bezúplatné a Národní knihovna může zpracovávat pouze domény a poddomény uvedené v předmětu smlouvy), autorská práva a dobu trvání smlouvy. Tento proces obnáší značnou administrativní zátěž jak pro vydavatele, tak i pro Webarchiv. Počet zdrojů, se kterými má podepsanou smlouvu, představuje v současnosti zhruba 87 % z celkového počtu archivovaných webů.

2. Licence Creative Commons²⁷

Další možností je využití licence Creative Commons. Jedná se o soubor oprávnění, která umožňují legální sdílení a využívání autorských děl. Vznikla z potřeby právně upravit vztah mezi vydavatelem a uživatelem v případě, že vydavatel chce, aby jeho dílo mohla za určitých podmínek používat široká veřejnost. Jde celkem o šest variant licencí, kdy si sám vydavatel určí, zda je například nutné uvést pouze autora, nebo zda je důležité, aby dílo nebylo používané ke komerčním účelům. Práce s tímto typem licence je jak pro vydavatele, tak i pro Webarchiv mnohem rychlejší a jednodušší, stačí, když vydavatel své stránky vystaví pod jednu z vybraných licencí. Zdroje, které jsou vystavené pod některou z licencí Creative Commons, představují v současnosti zhruba 13 % z celkového počtu archivovaných webů. V posledním roce však množství těchto zdrojů stoupá.

Přestože otevřených licencí existuje několik typů (např. Open Data Commons nebo Open Publication Licence²⁸), česká legislativa v současnosti zná pouze Creative Commons. „*Creative Commons nejsou popřením klasického pojetí autorského práva, jsou jeho nadstavbou a jako takové vycházejí z občanského zákoníku (§ 2358 – 2389 Zákona č. 89/2012 Sb., občanský zákoník)*“.²⁹ Licence Creative Commons je zároveň snadno využitelná a přizpůsobitelná potřebám tvůrců obsahu.

¹⁸ Referenční centrum. *Národní knihovna České republiky* [online]. Praha, 2013 [cit. 2017-07-24]. Dostupné z: <https://www.nkp.cz/sluzby/pc34>.

¹⁹ Statistika. *CZ.NIC: správce domény cz* [online]. c2017 [cit. 2017-07-24]. Dostupné z: <https://stats.nic.cz/>.

²⁰ K červenci roku 2017.

²¹ Digital rights management. *Wikipedie: Otevřená encyklopedie* [online]. c2016 [citováno 25. 07. 2017]. Dostupný z: https://cs.wikipedia.org/w/index.php?title=Digital_rights_management&oldid=14142081.

²² K červenci roku 2017.

²³ Legal deposit for websites and electronic publications. *British Library* [online]. [cit. 2017-07-25]. Dostupné z: www.bl.uk/aboutus/legaldeposit/websites.

²⁴ Legal deposit. *IIPC: International Internet Preservation Consortium* [online]. c2017 [cit. 2017-07-25]. Dostupné z: <https://netpreserve.org/web-archiving/legal-deposit/>.

²⁵ THE LEGAL DEPOSIT OF ELECTRONIC PUBLICATIONS. *UNESCO: United Nations Educational, Scientific and Cultural Organization* [online]. Working Group of the Conference of Directors of National Libraries, 1997 [cit. 2017-07-25]. Dostupné z: <http://www.unesco.org/webworld/memory/legaldep.htm#Section1>.

²⁶ Dokument, který vznikl elektronicky, bez analogového ekvivalentu (např. webová stránka).

²⁷ Creative Commons. *Webarchiv* [online]. [cit. 2017-07-24]. Dostupné z: <http://webarchiv.cz/cs/creative-commons>.

²⁸ Open Publication Licence. *Opencontent.org* [online]. 1999 [cit. 2017-07-25]. Dostupné z: www.opencontent.org/openpub/.

²⁹ Úvod. *Creative Commons: Česká republika* [online]. 2017 [cit. 2017-07-25]. Dostupné z: <https://www.creativecommons.cz/uvod/>.

Označení licence	Práva	Povinnosti	Název licence
BY			Uveďte autora
BY-SA			Uveďte autora – Zachovejte licenci
BY-ND			Uveďte autora – Nepracovávávejte
BY-NC			Uveďte autora – Neužívejte komerčně
BY-NC-SA			Uveďte autora – Neužívejte komerčně – Zachovejte licenci
BY-NC-ND			Uveďte autora – Neužívejte komerčně – Nepracovávávejte

Obrázek 3: Varianty licence Creative Commons. Zdroj: Webarchiv [online]. [cit. 2017-07-24]. Dostupné z: <https://www.webarchiv.cz/cs/creative-commons>

Role kurátora

Role kurátorů je v oboru webové archivace velmi důležitá. Kurátoři Webarchivu vedou komunikaci Národní knihovny s vydavateli a zajišťují uzavírání smluv, připravují výše zmíněné výběrové sklizně či koncipují tematické kolekce, na jejichž výsledné podobě se svým výběrem podílejí. Kolekce jsou tak specifickým předvýběrem pro koncového uživatele. Aktuálně se jedná celkem o tři skupiny uživatelů³⁰ s různými potřebami:

1. Individuální uživatelé – širší veřejnost, kdokoliv, kdo má přístup k internetu a webovému prohlížeči a stránky Webarchivu navštěvuje, protože má osobní zájem o prohlížení archivních verzí webových zdrojů.
2. Institucionální uživatelé – instituce, které data Webarchivu využívají ke své vlastní činnosti, například policie, soudy atp.
3. Výzkumníci a vědci – uživatelé, kteří požadují komplexnější soubor dat. Nespokojí se s archivním záznamem online zdroje, ale často mají zájem například i o metadata.

Všechny výše uvedené skupiny uživatelů mají specifické potřeby a následně i požadavky směrem k Webarchivu. Individuálním uživatelům stačí ve většině případů přístup k archivním verzím různých webů, které jsou dostupné na stránce www.webarchiv.cz³¹ prostřednictvím zobrazovacího rozhraní Wayback Machine.³² Institucionální uživatelé požadují kromě klasického přístupu i oficiální potvrzení, že se skutečně jedná o historické zobrazení konkrétní verze webové

stránky. K výzkumníkům Webarchiv přistupuje individuálně – badatel například potřeboval soubor metadat o cca 1 000 webových stránkách vzniklých mezi lety 2012–2016, atp. Kurátoři tak musí počítat s mnoha variantami požadavků.

Aby bylo možné podívat se, kam role kurátorů Webarchivu bude směřovat, je potřeba si nejdříve definovat roli digitálních kurátorů. Česká terminologická databáze knihovnictví a informační vědy definuje digitální kurátorství jako „správu digitálních dat a jiných digitálních materiálů, jejímž smyslem je zejména dlouhodobé uchování a současné i budoucí zpřístupnění digitálních objektů“.³³

Digitální kurátorství je podle Neila Beagrieho „soubor činností nutných k uchování dat a jiných digitálních materiálů a jejich následné zpřístupnění uživatelům a budoucím generacím“.³⁴

Digital curation center (DCC) uvádí, že „digitální kurátorství spočívá v zachování digitálních objektů, zpřístupnění jejich obsahu a zároveň přidání hodnoty vsazením do kontextu“.³⁵

Digitální kurátor tedy musí zhodnotit digitální data, vybrat pouze ta relevantní (v závislosti na předem definovaných pravidlech), zajistit k nim uživatelům jednoduchou a bezpečnou možnost přístupu a dodání kontextu. Pokud je tedy takto definované digitální kurátorství aplikováno na Webarchiv, jednalo by se o následující pracovní postup:

- akvizice webových zdrojů
- hodnocení webových zdrojů jak z hlediska obsahového, tak technického

³⁰ Jaroslav KVASNICA – Barbora BJAČKOVÁ. Profilování kolekce a stanovení určené skupiny WebArchivu Národní knihovny ČR. *ProInflow* [online]. 2014, roč. 6, č. 1, s. 13–25 [cit. 2017-07-25]. ISSN 1804–2406. Dostupné z: <http://www.phil.muni.cz/journals/index.php/proinflow/article/view/942/1072>.

³¹ *Webarchiv: památník českého internetu* [online]. Praha, 2017 [cit. 2017-07-24]. Dostupné z: <http://webarchiv.cz>.

³² The Wayback Machine. *Archive.org* [online]. [cit. 2017-07-24]. Dostupné z: https://archive.org/about/faqs.php#The_Wayback_Machine.

³³ Jaroslava HAVLOVÁ. Digitální kurátorství. *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online]. Praha: Národní knihovna ČR, 2003- [cit. 2017-08-08]. Dostupné z: http://aleph.nkp.cz/F/?func=direct&doc_number=000016392&local_base=KTD.

³⁴ Neil BEAGRIE. Digital Curation for Science, Digital Libraries, and Individuals. *The International Journals of Digital Curation* [online]. 2006, vol. 1, no. 1, pp. 3–16 [cit. 2017-07-25]. ISSN 1746-8256. Dostupné z: <http://ijdc.net/index.php/ijdc/article/view/6/2>.

³⁵ What is digital curation? *DCC* [online]. c2004-2017 [cit. 2017-07-24]. Dostupné z: <http://www.dcc.ac.uk/digital-curation/what-digital-curation>.

- komunikace s vydavatelem
- uzavření smlouvy / vystavení zdroje pod licenci Creative Commons
- katalogizace webových zdrojů – pro každý webový zdroj zařazený do výběrové sklizně je vytvořen vlastní katalogizační záznam, v této souvislosti se Webarchivu podařilo vyvinout nástroj pro semi-automatické vytváření katalogizačního záznamu (WA-KAT)³⁶
- zpřístupnění archivních verzí webových zdrojů přes webovou aplikaci Wayback Machine
- další dohled nad webovými zdroji (může dojít ke změně vydavatele, URL, zániku zdroje, vypovězení smlouvy a podobně)

V současné době informačního přetížení a všudypřítomného datového smogu³⁷ je práce digitálních kurátorů Webarchivu velmi potřebná. Z jednotlivostí dávají dohromady smysluplný celek, balík informací – oborově řazené soubory webových stránek nebo tematické kolekce zaměřené na události, které mají širší ohlas v prostředí internetu (viz Tematické kolekce). Tyto kolekce proto mohou být pro uživatele a čtenáře užitečné, a to ať už jde o výběrové sbírky (katalog zdrojů, se kterými je uzavřena smlouva, případně ty, které jsou vystaveny pod licenci CC), nebo o tematické kolekce, které kurátoři začali připravovat v rámci monitorování celospolečenských debat, u kterých je předpoklad, že zaujmou významnější místo z hlediska dějin České republiky.

Jak již bylo zmíněno, Webarchiv by rád s novými tematickými kolekcemi vyšel vstříc potřebám uživatelů, vědců a badatelů. Vytváří úzce zaměřené odborné tematické sbírky, jako je například kolekce webových stránek Univerzity Karlovy, která aktuálně čítá dohromady přes 800 webových zdrojů. Do budoucna plánuje přinášet datasey vědcům a odborníkům z odborných pracovišť dle jejich specifických potřeb a definovaných požadavků.

Závěr

V dnešní době již uživatelé začínají objevovat webové archivy a zjišťovat, že se bez nich do budoucna neobejdou. Každá skupina uživatelů samozřejmě vyžaduje ty „své“ in-

formace a očekává, že jim je webové archivy poskytnou. Ať už se jedná o běžné uživatele, kteří hledají ztracené dokumenty, webové stránky ze svých záložek, nebo současné historiky, kteří chtějí zkoumat devadesátá nebo nultá léta. To přináší tlak na webové archivy, které se nemohou pouze starat o sběr a ochranu dat, k nimž ale nikdo nebude mít přístup.

Správné nastavení akvizice je jen jedním z klíčových faktorů, jak přizpůsobit současné webové archivy, aby co nejlépe sloužily svým uživatelům. Hledání nových akvizičních řešení, vytváření oborových tematických kolekcí nebo archivování univerzitních webových stránek přináší uživatelům data v nových souvislostech, se kterými mohou pracovat. Otevírají jim nové možnosti pro jejich další badatelskou činnost.

Nicméně kvůli legislativním překážkám, zejména na evropské půdě, které komplikují smysluplné využití archivovaných dat pro jakékoliv uživatele, ať už individuální, nebo výzkumníky, se webové archivy stávají tzv. *dark archivy*,³⁸ pouze úložišti bez uživatelů. Webové archivy hledají cesty, jak svůj archiv uživatelům otevřít. Snaží se zpřístupňovat volně šiřitelný obsah, který jeho tvůrci vystavují pod Creative Commons licenci nebo uzavírají licenční smlouvy s vydavateli, kteří jsou ochotni na takové podmínky přistoupit. V posledních letech se také objevuje snaha zpřístupnit alespoň metadata, která nejsou chráněna autorským zákonem, ale takový přístup je technicky i finančně velmi náročný a stejně uživateli přináší pouze zlomek z celého archivu.

Snad jen díky americkému internetovému archivu (Internet Archive), který se snaží archivovat kompletní internet, se dnes mohou uživatelé jednoduše dostat k archivním verzím webových stránek. Zpravidla ale jednotlivé evropské národní archivy mají podrobnější pokrytí těch částí internetu, na které jsou zaměřeny, a také disponují kvalifikovanými kurátory, kteří se starají o akviziční politiku. Podobně jako webové archivy po celém světě usilují o změnu místních legislativ, snaží se i Webarchiv dlouhodobě o zavedení povinného elektronického výtisku pro webové stránky a o zlepšení situace se zpřístupňováním archivních kopií pro uživatele. Situaci by částečně mohl pomoci nový knihovní zákon, který je ale teprve v legislativním procesu a který by elektronický povinný výtisk měl zavést.

³⁶ Webarchiv CZ: WA-KAT. *GitHub* [online]. c2017 [cit. 2017-07-24]. Dostupné z: <https://github.com/WebarchivCZ/WA-KAT>.

³⁷ Julie HALADOVÁ, Role informačního profesionála v kontextu informačního přetížení [online]. 2013 [cit. 2017-06-15]. Dostupné z: https://is.muni.cz/th/382364/ff_b/role_IP_v_kontextu_IO.pdf.

³⁸ Dark archiv je mezinárodní označení pro webové archivy, které nejsou přístupné pro uživatele online přes internet.