

Univerzita Pardubice
Fakulta filozofická

Webové stránky jako archiválie?

David Kacrda

Bakalářská práce

2009

ZADÁNÍ BAKALÁŘSKÉ PRÁCE
(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **David KAKRDA**
Studijní program: **B7105 Historické vědy**
Studijní obor: **Spisová a archivní služba**

Název tématu: **Webové stránky jako archiválie?**

Z á s a d y p r o v y p r a c o v á n í :

1. Objasnění pojmu webových stránek
2. Výběr informací z webů
3. Webové stránky z hlediska archivnictví
4. Trvalé uložení webů
5. Výhledy do budoucna

Rozsah grafických prací:

Rozsah pracovní zprávy:

Forma zpracování bakalářské práce: **tištěná/elektronická**

Seznam odborné literatury:

periodika: Archivní časopis, Computer, Sborník archivních prací, Veřejná správa

internetové

stránky:

www.issz.cz, www.mvcr.cz, www.cesarch.cz, www.archive.org, webarchiv.nkp.cz

Vedoucí bakalářské práce:

Mgr. Helena Pochobradská

Katedra historických věd

Datum zadání bakalářské práce: **30. dubna 2008**

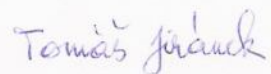
Termín odevzdání bakalářské práce: **31. března 2009**



prof. PhDr. Petr Vorel, CSc.

děkan

L.S.



doc. PhDr. Tomáš Jiránek, Ph.D.

vedoucí katedry

V Pardubicích dne 30. listopadu 2008

Souhrn

Tato práce se zabývá dlouhodobým uchováváním webových stránek jako archivních dokumentů. Část práce je věnována popisu sítě Internet a webových technologií. Dále je poukazováno na činnost nearchivních institucí, které se zabývají archivací webů. Jádrem práce je objasnění problematiky webových stránek v oblasti spisové a archivní služby a průzkum na dané téma.

Klíčová slova

Internet, webové stránky, archivace, Internet Archive Národní knihovna ČR, WebArchiv, spisová a archivní služba, průzkum, digitální dokument, Národní digitální archiv

Title

Web items as archival documents?

Abstract

This bachelor paper deals with long-term preservation of Web items Web as archival documents. Part of this work describes Internet and utility of world wide web. Next part dwell on non-archive institutions which have been working on long-term preservation of Web items. The core of the work is explication of the websites role in the system of records management and records keeping and the survey about the theme.

Keywords

The Internet, Web items, preservation, Internet Archive, National library of Czech republic, WebArchiv, records management and records keeping, survey, digital record, National digital archive

Na tomto místě bych rád poděkoval vedoucí mé bakalářské práce Mgr. Heleně Pochobradské za zajímavé téma, odbornou pomoc a trpělivost. Dále bych chtěl poděkovat všem respondentům mého průzkumu za jejich čas a zajímavé názory.

Obsah

1	Úvod.....	1
2	Úvod do Internetu, technologií a struktury webových stránek	3
2.1	Internet	3
2.1.1	Historie Internetu.....	4
2.2	Služba WWW.....	6
2.3	Webový prohlížeč.....	7
2.4	Webové technologie	8
2.4.1	Kaskádové styly.....	9
2.4.2	Skripty.....	9
2.4.3	DHTML a objekty prohlížeče	10
2.4.4	Flash animace a Java applety	11
2.4.5	Databáze.....	11
2.5	Webhosting a domény.....	12
2.5.1	Webhosting.....	12
2.5.2	Domény.....	14
2.6	Alternativní řešení	15
3	Nearchivní instituce a projekty zabývající se uchováváním webových stránek	17
3.1	Webové archivy.....	18
3.2	Knihovny.....	20
3.2.1	Legislativa.....	22
3.2.2	Co a jak uchovávat.....	28
3.2.3	Národní knihovna České republiky a dlouhodobé uchovávání webů	31
4	Webové stránky z pohledu spisové a archivní služby	42
4.1	Úvod do archivnictví	42
4.2	Archivy a dlouhodobé uchovávání webových stránek	44
4.2.1	Průzkum	46
4.2.2	Webové stránky v kontextu spisové služby	49
4.3	Webové stránky v archivní péči	53
5	Závěr	56

6	Seznam pramenů a literatury	58
7	Přílohy	63
8	Resumé	65

1 Úvod

Informační a komunikační technologie pronikají téměř do všech oblastí veřejného a soukromého života. Stále více a více oborů využívá jejich služby. Nejinak je tomu v oblasti spisové a archivní služby, kde vzniká a uplatňuje se stále větší množství dokumentů v digitální podobě. Tyto dokumenty mají jiné vlastnosti a požadavky než tradiční papírové, proto vnášejí do praxe předarchivní i archivní péče určité změny. Za naprosto specifický druh těchto digitálních dokumentů mohou být považovány i webové stránky.

V dnešní době představují webové stránky jeden z nejrozšířenějších způsobů prezentace informací a oslovení širokého spektra veřejnosti. Jejich výhoda spočívá především v jejich dostupnosti a schopnosti obsáhnout poměrně velký objem informací. Drtivá většina veřejnoprávních původců se v současnosti prezentuje na internetu pomocí webových stránek, či je využívá k různým úkonům. Přenášením některých úkonů na web vzrůstá i informační hodnota takových stránek a tím se dostávají do popředí zájmu archivářů, jako možné dokumenty trvalé hodnoty neboli archiválie.

Důvodem vzniku této bakalářské práce je chybějící legislativa a metodiky zabývající se weby v archivní sféře a také podle mne nedostatek odborné diskuze na toto téma. Základním podkladem práce je pak článek pracovníků Odboru archivní správy a spisové služby ministerstva vnitra Oskara Macka a Michala Wannera, věnující se problematice dlouhodobého uchovávání webů.¹ Jako další prameny jsou použity zejména elektronické články uveřejněné na internetu, specializované webové stránky, odborná periodika a publikace. Dále platná legislativa a v neposlední řadě i vlastní průzkum, ze kterého jsem čerpal hlavně poznatky pro zmapování názoru původců a archivů na danou problematiku a také v jakém postavení se nacházejí weby jednotlivých původců.

¹ MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchovávání webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 241-254.

Cílem práce je:

- Uvést technologie webových stránek a Internetu včetně základních pojmů.
- Seznámit se s nearchivními institucemi a projekty, zabývajícími se ukládáním a zpřístupňováním webů z hlediska jejich obsahové hodnoty.
- Provést analýzu postavení webových stránek ve veřejné správě a u vybraných původců.
- Popsat možné způsoby práce se stránkami v činnostech spisové a archivní služby s přihlédnutím v archivní legislativě a praxi.

2 Úvod do Internetu, technologií a struktury webových stránek

Internetové neboli webové stránky pro nás znamenají zejména přístup k informacím, zábavě a službám. Obsah stránek se neustále zpestřuje, kromě samotného textu se na stránkách objevují obrázky, videa i zvukové soubory. Pomocí stránek si můžeme prohlížet databáze (katalogy knihoven), nakupovat (tzv. E-shopy²) či sdílet údaje (dotazníky). Pro tyto účely jsou při tvorbě stránek používány různé technologie.

V této části se pokusím osvětlit základy webových stránek tak, aby čtenář získal potřebné základní znalosti a mohl se orientovat v odborných pojmech. Řekneme si také i něco o vývoji Internetu. Ačkoliv jsem se při psaní této kapitoly snažil být stručný a danou věc objasnit tak, aby ji pochopil především laik, je nutné, aby měl čtenář alespoň obecnou počítačovou gramotnost. I když by se mohlo zdát, že tato problematika přímo nesouvisí s odbornou archivní praxí, je pro účel této práce nezbytná. Z výše zmíněných důvodů jsem vycházel především z odborné literatury určené pro širší veřejnost³ a internetových zdrojů.⁴ Také jsem čerpal z bakalářské práce mého kolegy z Fakulty elektrotechniky a informatiky⁵.

2.1 Internet

Internet je celosvětová síť navzájem propojených počítačů či jiných zařízení, vzniklá za účelem sdílení dat a komunikace mezi jednotlivými uživateli. Pokud jsme připojeni k Internetu, je nám k dispozici téměř celý jeho obsah a přístup ke všem počítačům, ať už se nacházejí kdekoliv. Počítače připojené k Internetu lze rozdělit na typ **server** a **klient**. Server si lze představit jako speciální počítač, připojený k Internetu, který je neustále v provozu a na něm jsou uloženy samotné webové stránky. Klientem může být třeba náš osobní počítač

² E-shop je anglický výraz pro internetový obchod.

³ BROŽA, Petr. *Tvorba WWW stránek pro úplně začátečníky*. 5. aktualiz. vyd. Brno: Computer Press, a.s., 2006. 149 s. ISBN 80-251-1300-0.

BROŽA, Petr, BURANSKÝ, Imrich. *Programování WWW stránek pro úplně začátečníky*. 3. aktualiz. vyd. Brno: Computer Press, 2003. 202 s. ISBN 80-7226-818-X.

⁴ *Wikipedie, otevřená encyklopedie* [online]. 2002, 29. ledna 2009 [cit. 2009-01-29]. Dostupný z WWW: <http://cs.wikipedia.org/wiki/Hlavn%C3%AD_strana>.

JANOVSKÝ, Dušan. *Jak psát web, návod na html stránky* [online]. [2008], Poslední aktualizace 27. prosince 2008 [cit. 2009-01-29]. Dostupný z WWW: <<http://www.jakpsatweb.cz/>>. ISSN 1801-0458.

⁵ KOUDELA, Vlastimil. *Analýza WWW prezentací veřejné správy*. Pardubice, 2005. 85 s. Univerzita Pardubice, Ústav systémového inženýrství a informatiky. Vedoucí bakalářské práce Miloslava Bartoničková.

s možností přístupu na Internet. Mimo klasických serverů poskytujících data existují i specializované servery jako například poštovní server starající se o doručení zpráv elektronickou poštou anebo tzv. **DNS**⁶ servery, které mají za úkol rozpoznat adresu počítače a nasměrovat komunikaci, převádějí i slovní doménovou adresu na číselnou **IP adresu**⁷.

V Internetu funguje několik služeb. Asi nejznámější a nejpoužívanější je služba **WWW**⁸, která představuje v podstatě webové stránky a jejich vyhledávání a prohlížení. Další neméně známou službou je **E-mail** neboli elektronická pošta. Dále pak služby **FTP**⁹ pro přenos souborů, **Instant messaging**¹⁰ zajišťující komunikaci mezi uživateli v reálném čase, **VoIP**¹¹, pomocí které můžeme telefonovat přes Internet a další.

Abychom mohli používat Internet se všemi jeho službami, je potřeba být fyzicky k Internetu připojen. Možností připojení je několik a to buď přes telefonní linku, nebo přípojku od kabelové televize. Druhým typem je bezdrátové připojení, které může být realizováno prostřednictvím satelitní sítě, mobilní telefonní sítě a také přes dnes velmi rozšířené **Wi-Fi**¹². Toto připojení nám zrealizuje poskytovatel internetového připojení, což je společnost s přístupem k určité infrastruktuře či technologii, kterou budeme pro připojení k Internetu využívat, např. to může být telefonní operátor.

2.1.1 Historie Internetu

Počátek vývoje Internetu se datuje od roku 1957, kdy se Spojené státy americké pod tlakem aktivit bývalého Sovětského svazu rozhodly založit grantovou agenturu **ARPA** (Advanced Research Projects Agency). Hlavní náplní této agentury byl výzkum nových

⁶ DNS neboli Domain Name System je služba převádějící slovní adresy počítačů na číselné IP adresy a naopak. Podrobněji viz kapitola **2.5.2 Domény**.

⁷ IP adresa je jedinečná číselná adresa. Mají ji všechny počítače připojené k Internetu. Podrobněji viz kapitola **2.5.2 Domény**.

⁸ WWW- World Wide Web je název pro jednu ze služeb Internetu. Jedná se v podstatě o síť propojených hypertextových dokumentů, často označovaných jako webové stránky. Podrobněji viz kapitola **2.2 Služba WWW**.

⁹ FTP znamená File Transfer Protocol, což je jeden z typů protokolů používaný k přenosu dat mezi klientem a serverem.

¹⁰ Instant messaging je služba pro téměř okamžitou komunikaci svých uživatelů pomocí textových zpráv, posílání dat apod. Na rozdíl od jiných komunikací u instant messagingu máme přehled o tom, zda je konkrétní uživatel připojen a je tak k dispozici. Tuto službu zprostředkovávají speciální programy k tomu vytvořené, často nazývané jako IM klienti. Mezi nejznámější IM klienty patří třeba Windows Live Messenger nebo ICQ.

¹¹ VoIP (Voice over Internet Protocol) je technologie sloužící pro přenos digitalizovaného hlasu přes počítačovou síť (telefonování přes Internet). Hlasové spojení se provádí pomocí VoIP klienta (speciální program), jako např. Skype.

¹² Wifi označuje standart pro bezdrátovou síť. Může sloužit jako klasická lokální síť nebo pro připojení k Internetu.

technologií použitelných pro vojenské účely. Dochází k vývoji komunikační sítě společnosti RAND Corporation jako nejlepšího řešení na požadavek spolehlivého komunikačního systému v případě jaderné hrozby. Roku 1969 tak byla vytvořena první síť nazvaná **ARPANET**. K této síti bylo zpočátku připojeno jen několik univerzit, později přibývala další především výzkumná a vojenská střediska. V roce 1973 se k ARPANETU připojily první evropské instituce ve Velké Británii a Norsku. Později vznikla síť **Bitnet** pro americké střední a vysoké školy a roku 1983 se z ARPANETU vydělila armádní síť **Milnet**. Pro jednotnou komunikaci ve všech sítích byl v roce 1982 vytvořen protokol **TCP/IP**¹³ a je používán dodnes.

Internet však stál mimo zájem komerční sféry a sloužil pouze k výměně informací mezi jednotlivými složkami propojenými v síti. Síť sloužila pouze několika málo svým uživatelům a nebyla veřejně dostupná. V roce 1989 došlo k zásadnímu přelomu ve vývoji Internetu, když švýcarský fyzik **Tim Berners Lee** vymyslel nový způsob komunikace – **hypertextové dokumenty**, tedy dokumenty odkazující na další dokumenty, které mohou být umístěny kdekoli v síti. Hypertextové dokumenty sloužily původně pro vnitřní potřebu firmy **CERN**¹⁴ zabývající se jaderným výzkumem, ale brzy se rozšířily dále a dnes nazýváme tento způsob komunikace World Wide Web. Okolo roku 1990 se začíná objevovat souhrnný název pro síť - Internet. Páteřní síť ARPANET byla nahrazena sítí **NSF-NET** organizace **NFS** (National Science Foundation), která vznikla už roku 1986. Od roku 1992 probíhá vývoj moderního Internetu, jak ho známe dnes. Společnost NFS umožnila připojení komerčním subjektům a počet hostitelských počítačů překročil jeden milion.¹⁵

Do Čech přichází Internet až po roce 1989. První sítí, která se do tehdejšího Československa dostala, byla síť **EUNET** v roce 1990. Ve stejném roce se zde objevuje i síť **EARN** (European Academic and Research Network), což byla odnož amerického BITNETU. Oficiálně se Československá federativní republika připojuje k Internetu 13. února 1992 slavnostním aktem, který se konal na **ČVUT**¹⁶ v Praze. V Čechách panoval záměr vybudovat páteřní síť, která by propojila akademická střediska, vznikl tak projekt **FESNET** (Federal Educational and Scientific Network). Ten se však po rozdělení Československa přejmenoval

¹³ Transmission Control Protocol/Internet Protocol je označení pro celou skupinu protokolů pro komunikaci v počítačové síti a je to také hlavní protokol používaný v síti Internet.

¹⁴ Zkratka CERN (*Conseil Européen pour la recherche nucléaire*) označuje mezinárodní organizaci pro jaderný výzkum založenou v roce 1954 a sídlící v Ženevě. <http://public.web.cern.ch/>

¹⁵ *Historie Internetu* | *i-extra.net* [online]. c2007-2008 [cit. 2009-02-06]. Dostupný z WWW: <<http://i-extra.net/internet-a-site/historie-internetu>>. ISSN 1803-134X.

¹⁶ České vysoké učení technické. <http://www.cvut.cz/>

na **CESNET**¹⁷ (Czech Educational and Scientific Network). CESNET zajišťoval propojení jednotlivých akademických středisek. Pozdější přívod Internetu do městských aglomerací byl financován již soukromým sektorem.¹⁸

2.2 Služba WWW

WWW neboli World Wide Web je internetová služba používaná pro prohlížení webových stránek. World Wide Web by se dalo přeložit jako „celosvětová pavučina“, což vyjadřuje provázanost stránek v celém internetu. Při prohlížení stránky totiž můžeme narazit na část textu, která je zvýrazněna a kliknutím myši na ni zobrazíte další webovou stránku. Takovéto části textu se říká link neboli odkaz. Pomocí odkazů tak lze ihned zobrazit libovolné stránky, které se pod nimi skrývají, proto tedy pojem „pavučina“. WWW samozřejmě není jediná služba internetu, další často využívanou je např. e-mail.

WWW je založeno na třech základních technologiích:

- Uniform Resource Locator (URL)
- Hypertext Transfer Protocol (HTTP)
- Hypertext Markup Language (HTML)¹⁹

URL je slovní nebo také doménová adresa internetových stránek, nahrazující tzv. číselnou IP adresu, která je jedinečná a má ji přidělen každý počítač připojený k Internetu. URL jednoznačně identifikuje umístění stránky v Internetu. Příklad URL adresy: <http://www.archivzamrsk.cz>, odkazující na stránky Státního oblastního archivu v Zámrsku. O doménách a IP adresách bude ještě zmínka později²⁰.

Zkratka **HTTP** je označení pro protokol sloužící pro přenos webových stránek mezi jednotlivými počítači v Internetu. Pověštinou se jedná o server a osobní počítač uživatele neboli klient. Počítač klienta se spojí přes Internet se serverem, pomocí prohlížeče vyhledá požadovanou stránku, kterou prohlížeč zobrazí. V podstatě vždy, když si prohlížíme stránky na Internetu, používáme protokol HTTP. Jiné služby než WWW používají své vlastní

¹⁷ Je sdružení, které založily vysoké školy a Akademie věd České republiky v roce 1996, jejímž hlavním cílem je provoz a rozvoj páteční akademické sítě v ČR, která se nazývá CESNET2.

¹⁸ *Historie Internetu | i-extra.net* [online]. c2007-2008 [cit. 2009-02-06]. Dostupný z WWW: <<http://i-extra.net/internet-a-site/historie-internetu>>. ISSN 1803-134X.

¹⁹ KOUDELA, Vlastimil. *Analýza WWW prezentací veřejné správy*. Pardubice, 2005. 85 s. Univerzita Pardubice, Ústav systémového inženýrství a informatiky. Vedoucí bakalářské práce Miloslava Bartoničková.

²⁰ viz kapitola **2.5.2 Domény**

protokoly, např. pro stahování dat funguje protokol FTP (File Transfer Protocol) a e-mail využívá protokol SMTP (Simple Mail Transfer Protocol).

HTML je základním jazykem pro tvorbu stránek v systému World Wide Web a umožňuje definovat vzhled a strukturu stránky. HTML má podobu kódu, který ovlivňuje podobu stránek. Tento kód má svůj zdrojový text vytvořený programátorem nebo tvůrcem stránek. Při psaní zdrojového textu je potřeba dodržovat určitou syntaxi neboli postup, či pravidla, aby se webová stránka zobrazila v internetovém prohlížeči správně. Zdrojový text nemá žádné formátování, proto je pro jeho tvorbu ideální např. Poznámkový blok, který je součástí MS Windows. Text obsahuje, kromě vlastního textu webové stránky, také specifické značky. Tyto značky např. definují odstavce, velikost a barvu písma nebo odkazují na externí soubory, jakými mohou být obrázky, zvuky, videa apod., které jsou součástí stránek. Představme si tedy HTML jako soubor s příponou HTM nebo HTML, obsahující jednoduchý text vytvořený programátorem podle určitých pravidel. Takovýto soubor tvoří základní kámen webové stránky. Všechny ostatní soubory (např. obrázky) jsou umístěny mimo něj. Zdrojový text HTML je přeložen a zobrazen ve výsledné podobě webovým prohlížečem. Jazyk HTML je dnes základem většiny webových stránek.

Stejně jako většina informačních technologií, prošel i jazyk HTML od svého vzniku dlouholetým vývojem. Počátek jazyka se datuje do roku 1990 a v roce 1999 vznikla jeho poslední oficiální verze 4.01. Prozatímním nástupcem HTML je jazyk **XHTML** (Extensible HyperText Markup Language) jako novější forma HTML. XHTML je v podstatě postaveno na stejném principu jako HTML. Zásadní rozdíl mezi nimi je ten, že na XHTML je aplikován jazyk **XML** (Extensible Markup Language), což je obecný značkovací jazyk, který dovoluje vytvářet dokumenty v otevřeném formátu. Výhoda takových dokumentů je jejich nezávislost na konkrétní platformě a technologii.

2.3 Webový prohlížeč

Webový prohlížeč (Internetový prohlížeč) je program, který nám umožňuje po zadání adresy vyhledat požadovanou webovou stránku, prohlížet si internetové stránky a zobrazit jejich obsah. Jednak dokáže komunikovat pomocí HTTP protokolu a především zpracovat kód stránky a zobrazit stránku ve výsledné podobě tak jak ji vidíme běžně ve svém počítači. Prohlížeč také umožní zobrazit obrázky, přehrát video, stáhnout soubor umístěný na webu atd.

Pro některé složitější aplikace jako např. **Flash animace** nebo **Java applety**²¹, které mohou být součástí webové stránky, je potřeba mít nainstalované speciální doplňky nebo **zásuvné moduly (pluginy)**²². Některé prohlížeče si tyto doplňky sami vyhledají a po odsouhlasení uživatele nainstalují.

Mimo to nám prohlížeč usnadňuje práci s internetovou stránkou při jejím prohlížení. Můžeme se pomocí tlačítek vracet souvisle dozadu nebo dopředu na stránky, které jsme již předtím zobrazili. Lze si také nastavit domovskou stránku, což je stránka, kterou prohlížeč při svém spuštění automaticky zobrazí. Pomocí prohlížeče dokážeme spravovat stahované soubory nebo dokonce uložit stránku na disk do svého počítače. Asi nejpoužívanějším prohlížečem je v současné době **Internet Explorer** od společnosti **Microsoft**²³. Tento prohlížeč je součástí operačního systému Windows a je i volně stažitelný na Internetu. Dalším možná o něco méně známým prohlížečem je **Firefox** od firmy **Mozilla**²⁴, který však v poslední době svou kvalitou dohnal Internet Explorer a díky některým svým přednostem nachází oblibu u stále většího počtu uživatelů, což mohu potvrdit ze svých dlouholetých zkušeností. Nutno však říci, že souboj na poli webových prohlížečů pokračuje a stále častěji vycházejí nové a dokonalejší verze obou zmíněných prohlížečů. V poslední době se navíc objevily prohlížeče nové jako třeba **Google Chrome** od známé společnosti **Google**²⁵ a lze očekávat, že by mohl být silnou konkurencí pro dosavadní webové prohlížeče.

2.4 Webové technologie

Staré webové stránky byly tvořeny pouze jazykem HTML, který definoval jejich strukturu. Bylo možné ovlivnit jejich vzhled, barvu a velikost písma, umístit na stránky obrázky apod. Avšak jednalo se pouze o tzv. statické webové stránky, respektive uživatel nemohl ovlivnit činnost stránky, mohl se pouze proklikávat odkazy, zobrazovat a číst obsah stránky. Toto však brzy přestalo provozovatelům i uživatelům stránek stačit. Vznikl tak požadavek na stránku spolupracující s uživatelem a lépe vyhovující současným nárokům na obsah webů. Výsledkem jsou tzv. **dynamické stránky**, které dovedou komunikovat s klientem a podle zadaných údajů zpracovat výsledek. Jako příklad si uveďme třeba klasický

²¹ viz kapitola 2.4.4 **Flash animace a Java applety**

²² Soubory zajišťující koordinaci své mateřské aplikace (webového prohlížeče) s externími programy, slouží např. k přehrávání videa apod.

²³ <http://www.microsoft.com/>

²⁴ <http://www.mozilla.org/>

²⁵ <http://www.google.com/intl/en/corporate/>

e-shop, čili internetový obchod, kde se zpracovávají údaje od zákazníka. V této kapitole si pokusíme objasnit nejpoužívanější technologie při tvorbě moderního dynamického webu.

2.4.1 Kaskádové styly

Jazyk HTML je svou formou poměrně pracný a nepřináší žádné možnosti formátování textu. Veškeré definice samotného textu webové stránky (odstavce, velikost písma apod.) se provádějí speciálními značkami. Hlavní nevýhoda spočívá v tom, že je pokaždé potřeba znova definovat text např. u každého jednotlivého nadpisu zvlášť. Není možné si dopředu navolit určité styly a ty pak používat jako třeba u textového editoru MS Word. Právě tyto důvody vedly ke vzniku kaskádových stylů neboli **CSS (Cascading Style Sheets)**.

Můžeme říci, že CSS jsou jakousi nadstavbou jazyka HTML, případně XHTML. Kaskádové styly nám velmi usnadňují práci při tvorbě stránek a s jejich pomocí lze dosáhnout velice zajímavých efektů na webové stránce. Použití CSS má tři možnosti, přičemž první dva pracují se zdrojovým textem stránky a s použitím speciálních značek. Třetí způsob je vytvoření souboru s příponou `css`²⁶ a s definovaným stylem, na který je pak ve zdrojovém textu odkazováno pomocí značek. Kaskádové styly patří spolu se skripty k prvkům, které nejvíce oživují stránku.

2.4.2 Skripty

Skript je programový kód vytvořený v určitém programovacím jazyku. Tento kód je přímo součástí zdrojového textu webové stránky (HTML), případně může být uložen v externím souboru a zdrojový text na něj odkazuje. Skript je úzce propojen s prohlížečem a určuje mu, co má udělat, která data se mají zpracovat atd. V podstatě jde o to, že ovlivňuje chování stránky v reálném čase na základě požadavků od klienta, např. jako když na Internetu vyplňujeme dotazník apod.

Jak už bylo řečeno, skripty jsou vytvářeny v různých programovacích jazycích. Jednotlivé skripty se od sebe liší a mají také jiné využití. Mimo jiné se skripty dělí na ty, které se spouštějí na straně serveru a na ty, které na straně klienta. Princip spočívá v tom, že skript

²⁶ Přípona souboru je část názvu souboru oddělená od kořene názvu tečkou, např. název souboru.css. Přípona je tvořena 2-4 znaky a určuje typ souboru, podle kterého spustí operační systém příslušný program, jež dovede se souborem pracovat.

je zpracováván příslušným programem, buď přímo na serveru, čili v počítači, kde se nachází samotná internetová stránka, nebo na počítači uživatele (klienta). V obou případech je potřeba ke spuštění skriptu příslušný program. Na straně klienta je to webový prohlížeč. Nejrozšířenější skriptovací technologií na straně klienta je **JavaScript**. Dnešní webové prohlížeče, již mají v sobě obsaženou podporu JavaScriptu a umějí s ním pracovat. Mezi skripty prováděné serverem patří např. **PHP (Hypertext Preprocessor)**, **ASP (Active Server Pages)** a dříve používané **CGI (Common Gateway Interface)**. Hlavní využití serverových skriptů spočívá především ve spolupráci s databázemi, e-mailem atd.

Každá ze skriptovacích technologií má své konkrétní využití a svá specifika. Nebudeme si zde popisovat jednotlivé technologie, pro pozdější práci se stránkami je důležité určit jednotlivé skripty a uvědomit si, jaký software je nutný pro práci se skriptem, případně v jakém rozhraní skript funguje. Tyto otázky by však už měly být úkolem pro IT odborníka.

2.4.3 DHTML a objekty prohlížeče

DHTML je zkratka pro pojem dynamické HTML, které bylo před pár lety dalším stupněm ve vývoji webových stránek. DHTML nikterak neupravuje samotný jazyk HTML, ani není samostatnou technologií. Je to spíše souhrnné označení pro kombinaci kaskádových stylů (CSS) a JavaScriptu implementovaných do jazyka HTML. Dynamické HTML je právě ten způsob, jak vdechnout stránkám život.

Skripty vložené do stránek mohou měnit jejich obsah a vzhled v reálném čase. Už víme, že pomocí CSS si můžeme nadefinovat vzhled různých atributů na naší stránce, dynamické HTML umožňuje pomocí skriptů tyto definice libovolně změnit v podstatě za běhu stránek.²⁷ JavaScript zase dokáže pracovat s objektovým modelem prohlížeče a dokumentu, což je vlastně okno prohlížeče a další prvky v něm obsažené, které jsou úzce vázány na jazyk HTML. Objektový model má své jednotlivé části (objekty), které jsou definovány buď metodou (příkaz), vlastností (určitá hodnota) nebo podobjektem (může obsahovat všechny tyto tři části). Všechny objekty prohlížeče mají své jméno a určitou hierarchii.

²⁷ KOSEK, Jiří. *Dynamické HTML: Úvod* [online]. c1998 , Poslední modifikace: 24-01-98 [cit. 2009-02-15]. Dostupný z WWW: <<http://www.kosek.cz/clanky/dhtml/uvod.html>>.

2.4.4 Flash animace a Java applety

Jedná se především o různé animace, které jsou často součástí webových stránek. **Flash animace** jsou vytvářeny ve **Flash editoru**. K animaci je možné přidat zvuk a také skript, který pak ovlivňuje chování samotné animace (např. **ActionScript**, podobný JavaScriptu). Hotová animace se převede do spustitelného souboru ve formátu **SWF (Shockwave Flash)**, který je následně možné přehrávat. Animace se přehrávají v prohlížeči, je však potřeba mít nainstalovaný speciální zásuvný modul (plugin) v podobě Flash přehrávače. Zásuvné moduly jsou doplňky pro internetové prohlížeče. Prohlížeč většinou sám upozorní na absenci potřebného modulu. Flash přehrávač se nazývá **Adobe Flash Player** a pochází od firmy **Macromedia, Inc**. Dnes existuje jako plugin ve dvou verzích, jedna je určena pro webové prohlížeče Firefox, Mozilla a **Opera**. Druhá verze je specifická pro Internet Explorer, využívající technologii ovládacích prvků **ActiveX**. Adobe Flash Player je distribuován výrobcem k volnému použití a lze si ho zdarma stáhnout z oficiálního webu výrobce.

Applety jsou externí programy vytvořené v programovacím jazyku **Java**. Slouží hlavně k vytváření různých jednodušších animací, umožňují přehrávat i zvuky či otevřít nové okno prohlížeče. Applety se vkládají do zdrojového textu HTML, kde se na ně odkazuje. Pro bezproblémový běh Java appletů v prohlížeči je zapotřebí Java rozhraní²⁸ - **Java Runtime Environment** od společnosti **Sun Microsystems, Inc**²⁹. Stejně jako Adobe Flash Player je bezplatně k dispozici na webu.

2.4.5 Databáze

Databáze všeho druhu provázejí informační technologie už od jejich vzniku, ať už to bylo v podobě různých úložišť dat nebo profesionálních databází. Potřeba ukládat velké objemy dat neustále roste i v současné době. Databáze umožňují organizovat a spravovat vložená data podle parametrů uživatele a také je zpětně z databáze získávat. Na Internetu se můžeme setkat s databázemi v podobě **online**³⁰ katalogů knihoven či nabídek e-shopů.

²⁸ Rozhraní (interface) označuje jednak nějaký fyzický vstupní či výstupní bod, který spojuje dvě zařízení (např. paměť a procesor – hardwarové rozhraní). Dále existuje ještě softwarové rozhraní, jež představuje jakousi pomyslnou hranici, kde se setkávají dva programy a komunikují spolu.

²⁹ <http://www.sun.com/>

³⁰ Výraz online se užívá pro označení něčeho, co je publikované či dostupné na Internetu.

Zde si přiblížíme jednu z často využívaných databází (nejen) na webu - **MySQL**. MySQL je relační databázový systém vytvořený švédskou firmou **MySQL AB**. Jeho hlavními výhodami je možnost bezplatného užívání a použití pro více různých operačních systémů. Pro správu databází je zapotřebí dalších programů, které nám umožní databázi vytvořit, nadefinovat oprávnění pro přístup atd. V případě MySQL je takovým nástrojem **PHPMysqlAdmin**, aplikace vytvořená v jazyce PHP. PHPMyAdmin zadává příkazy databázi v jazyce **SQL (Structured Query Language)**, prostřednictvím kterého databáze komunikuje. Přes tohoto správce databáze lze MySQL spravovat i vzdáleně (z jiného počítače) pomocí PHP skriptů³¹, které jsou spouštěny přímo na serveru.

2.5 Webhosting a domény

2.5.1 Webhosting

Doposud jsme se seznámili s principy tvorby a technologiemi webových stránek. Většina z nás si jistě položí otázku: „Dobře, tak tedy mám vytvořeny nějaké stránky, ale co teď s nimi, jak je dostanu na Internet?“ Odpověď zní: „**webhosting**“. Aby byly webové stránky neustále přístupné všem uživatelům Internetu je potřeba umístit je na nějaký server, což je počítač se stálým připojením k Internetu. Bylo by jistě možné zřídit si takový sever i doma, tato varianta je však velice finančně i odborně náročná. Server musí být neustále v provozu a má své specifické požadavky na hardwarovou a softwarovou výbavu, také jeho správa a údržba by měla být svěřena do rukou odborníka. Ideálním a dnes zcela běžným řešením, kde najít místo pro své stránky, je již zmíněný webhosting. Webhosting je služba, kterou nabízejí specializované firmy. Taková firma má v provozu několik serverů, na kterých vám umožní provozovat vaše webové stránky, většinou za určitý periodický poplatek. Cena za umístění stránek může být různá podle požadavků zákazníka.

Hlavními kritérii při výběru hostingu a výsledné ceny jsou:

- Datový limit - neboli velikost prostoru pro naše stránky. Zde záleží na velikosti stránek, běžné stránky se dokážou vejít do 10 Mb. Pokud však chceme na web umístit i video či databáze, je potřeba prostoru více, tím vyšší je také ale i cena.

³¹ viz kapitola 2.3.2 Skripty

- Podpora skriptů a databází - toto je velmi důležité hledisko, pokud jsme při tvorbě stránek použili některou ze skriptovacích technologií na straně serveru nebo využíváme databázi. V takovém případě je potřeba vyžadovat od hostingu podporu té či oné technologie či databáze.
- Logování – je to služba, kdy server zapisuje informace o provozu stránek do souborů tzv. logů. Zaznamenává se IP adresa návštěvníka stránek, čas atd. Tyto informace mohou být využity k různým statistikám a zpětné vazbě k tvůrci stránek.
- Dostupnost - znamená zaručený přístup na webové stránky tak, aby se třeba nestalo, že tři hodiny v průběhu dne bude server nedostupný a na stránky se tak nikdo nedostane. Webhostingové firmy téměř vždy uvádějí v nabídce služeb zaručenou dostupnost v procentech.
- Technická podpora – měla by být samozřejmá u kvalitního webhostingu a také by měla fungovat dvacet čtyři hodin denně.
- Administrační rozhraní a kopírování souborů na server – jedná se o správu našich stránek, které jsou již umístěny na serveru. Administrační rozhraní by nemělo být příliš uživatelsky náročné a mělo by poskytovat všechny důležité úkony pro správu webových stránek. Soubory se na server kopírují většinou pomocí FTP protokolu. Od správce serveru dostaneme přidělený FTP účet, tudíž FTP adresu serveru a uživatelské jméno a heslo k našemu účtu. Abychom mohli soubory se stránkami nahrát na server, potřebujeme ještě FTP klienta, což je program, který umí pracovat s FTP protokolem. Takovýchto programů je několik a lze například použít i oblíbeného správce souborů **Total Commander**.
- Další služby – u placených hostingů existuje kromě základní nabídky ještě celá škála služeb, které jsou buď zdarma nebo za příplatek. Mezi standardní neplacené služby patří například zřízení e-mailových schránek, počítadlo přístupů, formulář na odesílání e-mailů atd. Připlatit si můžeme třeba za SSL certifikáty zabezpečení nebo za obnovu dat ze zálohy. Zde bychom měli upozornit ještě na měsíční limit přenesených dat. U kvalitních hostingů bývá tento limit neomezený, ale je dobré se o tomto pro jistotu informovat dopředu, vyhneme se tak možným pozdějším nepříjemnostem.

Webhostingové společnosti často nabízejí už připravené balíčky služeb, ze kterých si může zákazník vybrat ten, který nejvíce vyhovuje jeho potřebám. Jednotlivé balíčky se liší cenou a nabízenými službami. Pokud má zákazník specifické požadavky, je samozřejmě možné objednat si hosting šitý na míru.

2.5.2 Domény

O doménách byla zmínka, když jsme si vysvětlovali, co je to URL. Nyní si o nich povíme něco víc. Aby mohly webové stránky v Internetu fungovat, musí mít svou slovní neboli **doménovou adresu**. Tato adresa nahrazuje tzv. **IP adresu**, což je číselná adresa, kterou má přiděleno každé zařízení zapojené v síti, tedy i každý počítač v Internetu. Tato adresa je pro každý počítač nebo zařízení jedinečná, aby ho šlo jednoznačně identifikovat a navázat s ním spojení. Zkratka IP znamená **Internet Protocol**, tedy protokol pomocí kterého komunikují všechna zařízení v Internetu. IP adresa je tvořena čtveřicí čísel od 0 do 255 a může mít například takovouto podobu 192.168.0.1. Vzhledem k tomu, že každá IP adresa musí být jedinečná, je v současnosti možno připojit k Internetu asi přes 4 miliardy počítačů. Podle dlouhodobých prognóz však nebude tento počet v poměrně blízké budoucnosti stačit. Z toho důvodu se plánuje přechod ze současné verze IP adres **IPv4** na verzi **IPv6**, která zvýší počet možných adres na stovky sextiliónů.

Vyhledávat však stránky podle jejich IP adres by bylo velice nepraktické a nepohodlné. Představme si, že bychom si museli zapamatovat několik takovýchto IP adres a ještě k jakým konkrétním stránkám se ta která IP adresa vztahuje. Za účelem snadnějšího a uživatelsky daleko přívětivějšího vyhledávání webových stránek a orientace v Internetu vůbec vznikla služba **DNS (Domain Name System)**. Systém DNS umožňuje přiřadit k číselné IP adrese slovní adresu neboli doménové jméno, což Internet podstatně „zlidštilo“. Výhoda slovní adresy tkví nejenom ve snadnější zapamatovatelnosti, ale i v intuitivním vyhledávání, např. pokud známe název nějaké instituce, lze zkusit zadat rovnou do prohlížeče: www.jmenoinstitute.cz (bez mezer a diakritiky) a můžeme se dostat rovnou na požadované stránky, navíc umožňuje provozovatelům stránek zachytit ve slovní adrese jejich název či klíčové slovo vztahující se k obsahu stránek.

Doménový název stránek je rozdělen podle určité hierarchie. Vezměme si například adresu Univerzity Pardubice www.upce.cz, kde **cz** znamená doménu první úrovně neboli také generickou doménu, **upce** doménu druhé úrovně a **www** doménu třetí úrovně, této poslední doméně se také říká subdoména. Jednotlivé domény jsou od sebe odděleny tečkami. Doména první úrovně vyjadřuje většinou příslušnost k určitému státu, tedy v ČR se používá cz a na Slovensku sk apod. Dále existují domény všeobecné jako třeba známé com, dále gov, edu, info, net, biz apod. Doménu druhé úrovně si tvoří registrátor domény (většinou majitel stránek) podle vlastního uvážení, nejčastěji se vztahuje k názvu nebo obsahu stránek. Doména třetí úrovně je méně podstatná než zbývající dvě a většinou je v podobě www, což je dáno

pouze zvykem z doby, kdy ještě internet nebyl tolik rozšířen. Registrátor si může požádat o svou vlastní doménu třetí úrovně. Jedna doména druhé úrovně může obsahovat více domén třetí úrovně, příkladem je třeba již zmíněná adresa www.upce.cz a adresa khv.upce.cz, což jsou stránky Katedry historických věd Univerzity Pardubice. Doména druhé úrovně tedy **upce** je u obou adres stejná, mění se však doména třetí úrovně, tedy subdoména.

Byl zde řečen pojem registrátor domény, z čehož vyplývá, že doménu je před použitím potřeba zaregistrovat. Doménu vám zaregistruje, podobně jako u webhostingu, specializovaná firma nabízející tyto služby, často může poskytovat jak služby hostingu tak i registraci domén. Zaregistrování domény není samozřejmě zadarmo a je potřeba uhradit určitý zřizovací poplatek na určitou dobu a uplynutí této doby se platí obnovovací poplatek, pokud chceme doménu dál používat. Není možné používat doménu, která už byla zaregistrována kdekoliv na světě a je stále v platnosti, respektive nemohou existovat dva weby se stejnou doménou. Zda je požadovaná doména volná, zjistíme u registrátora domény, který nám také řekne pravidla pro tvorbu domény. Mohou však být zaregistrovány dvě stejné domény druhé úrovně, ale každá s jinou doménou první úrovně, např. když si budeme chtít zaregistrovat doménu www.archiv.cz, ale ta bude již obsazena, můžeme zkusit www.archiv.com (případně jinou všeobecnou doménu). Pokud se ani tak nedobereme k námi vytoužené doméně, není jiná možnost než pokusit se vymyslet jiný název nebo počkat až majiteli vyprší platnost a on si ji včas neprodlouží (nebo ji nechce prodloužit) či zkusit se s majitelem domluvit na odkoupení domény. Informace o majiteli a době platnosti konkrétní domény si lze vyhledat na Internetu v registrech domén.

2.6 Alternativní řešení

Způsoby tvorby webových stránek založené na technologiích, které jsme si popsali v předchozích kapitolách, jsou profesionálními způsoby jak vytvořit web. K tvorbě takovýchto stránek je potřeba znát a ovládat některé ze zmíněných technologií, nejméně alespoň jazyk HTML. Naučit se však pracovat s jazykem HTML může být poměrně pracné i časově náročné, nemluvě o dalších technologiích. Leckoho by mohla tato překážka od záměru vytvořit si svůj vlastní web odradit. Nabízí se tedy otázka, zda neexistuje snadnější způsob, jak si zhotovit webové stránky.

Alternativních způsobů tvorby webových prezentací je celá řada. Jedním z nich jsou tzv. **wysiwyg (What you see is what you get)** editory, což jsou programy, které nám

umožňují pracovat rovnou s grafickou podobou stránek, tedy přímo vidíme, jak bude stránka v prohlížeči vypadat. Výslednou podobu pak převádějí do HTML kódu, ten je pak možno zpětně upravit. Výhodou těchto editorů je nepochybně to, že s nimi mohou pracovat i uživatelé bez znalosti HTML. Nejznámějšími zástupci wysiwyg editorů jsou **MS FrontPage** a **Macromedia Dreamweaver**. Dále je možno použít některou z volně dostupných služeb na Internetu pracujících na podobném principu jako předešlé editory. Například www.estranky.cz, kde možné vytvořit si vlastní stránky šité na míru. K snadné tvorbě webu je zde široká paleta šablon a navíc je možné zvolit si typ webových stránek, zda to budou jen osobní stránky či stránky firmy nebo instituce. Na výběr máme z několika balíčků od základního pro nekomerční využití zdarma až po placené pro firemní účely. Vytvořením stránek však nabídka nekončí, se stránkami je také automaticky spojený webhosting a doména s tím, že si vybereme svou doménu třetího řádu a doména druhého a prvního řádu bude estranky.cz. Celá adresa pak může vypadat takto www.nase-domena.estranky.cz. Je totiž možné, aby doména druhého řádu obsahovala více domén třetího řádu. Vytvoření si vlastního webu nabízejí i někteří poskytovatelé webhostingu a domén. Samozřejmě je možné najít i poskytovatele bezplatných hostingů pro naše stránky.

Alternativní způsoby tvorby stránek a bezplatného webhostingu mohou být velmi užitečné pro uživatele, kteří nemají potřebné znalosti webových technologií anebo nechtějí investovat úsilí, čas a finance spojené s tvorbou i provozem stránek. Tyto způsoby však sebou téměř vždy nesou méně či více omezení. Uživatel nemá nad stránkami plnou kontrolu a často může docházet k drobným závadám a komplikacím. Profesionální webové stránky jsou tvořeny za pomoci nejúčinnějších technologií, aby bylo dosaženo kýženého efektu. Neméně důležitý je i kvalitní webhosting, aby vše bez problémů fungovalo.

3 Nearchivní instituce a projekty zabývající se uchováváním webových stránek

Tato práce se ve své podstatě zabývá archivací webových stránek a od této kapitoly se budeme věnovat již tomuto tématu a to v různých spektrech. Archivací webových stránek rozumíme dlouhodobé uchovávání a zpřístupňování webů.

Jejími hlavními činnostmi a cíli jsou:

- Vytváření digitálních kopií určených webových stránek a jejich uložení na paměťových médiích.
- Zpřístupnění uložených dokumentů
- Zajištění dlouhodobého uchování a trvalého přístupu k archivovaným dokumentům
- Kontinuální vytváření fondu či sbírky archivovaných webů a její organizace za účelem zajištění online vyhledávání uvnitř sbírky.³²

Dříve než se pustíme do samotné problematiky této práce, nabízí se otázka: „Proč se vlastně zabývat archivováním webových stránek, když je vše stále přístupné na Internetu takřka kdykoliv a kdekoliv?“. Zde je potřeba uvědomit si, že obsah webových stránek a celého Internetu vůbec je velice proměnlivý a nestálý. Přesto, že počet nových webových dokumentů roste ohromným tempem, u velké většiny webů nemáme jistotu, že budou existovat např. i za deset let, navíc se často mění i samotný obsah webů. Podle odhadů americké společnosti **Alexa Internet**³³ je průměrná délka životnosti webové stránky 75 dnů.³⁴ Takto se může lehce stát, a stává se, že navždy ztratíme hodnotné informace a díla, která nemají jinou podobu, než jsou elektronická data zpřístupněná na webu nebo jsou takové povahy, která jiné než online publikování nedovoluje.

Archivy nejsou zdaleka jedinými institucemi, které mají nebo by mohly mít zájem na trvalém uložení webů z hlediska jejich informační hodnoty. V této oblasti vyvíjejí aktivitu již delší dobu především knihovny a jsou také mnohem zkušenější než archivy. Obě instituce mají k webům odlišný přístup, a to zejména v souvislosti s legislativou, kterou se řídí.³⁵

³² CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 8 s. ISBN 978-80-7050-562-5.

³³ <http://www.alexa.com/>

³⁴ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 6 s. ISBN 978-80-7050-562-5.

³⁵ MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchovávání webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 242.

Přesto, že knihovny a archivy se nebudou ubírat zcela stejnou cestou, jejich cíl je v principu stejný - uchování a zpřístupnění webu, proto se archivy mohou od knihoven leccemu přiučít. V této kapitole si postupně povíme o legislativě, která práci knihoven ovlivňuje, metodách, kritériích a technologiích sběru webových stránek a nakonec se seznámíme s konkrétními projekty knihoven v České republice. Nejprve se však podívejme, kdo se jako možná vůbec první začal archivací webu ve světě zabývat.

3.1 Webové archivy

V oblasti uchovávání webů působí již velmi dlouhou dobu, organizace **Internet Archive**.³⁶ Tato nezisková organizace založená roku 1996 a její stejnojmenný projekt se snaží o zachování co nejširšího světového webu ruku v ruce s budováním internetové knihovny. Ve své činnosti spolupracuje Internet Archive s mnoha institucemi jako je např. Library of Congress, což je nejstarší federální knihovna v USA s největší sbírkou na světě³⁷ a mnoha dalšími knihovnami a muzei po celém světě. Internetová knihovna by měla především zajistit stálý přístup pro badatele, historiky a vědce ke svým sbírkám v digitální podobě. Na sbírky webů pohlížejí představitelé Internet Archive jako na součást digitálního kulturního dědictví lidstva. Archivované weby mohou mít využití jednak pro historiky jako historický pramen, jednak pro lingvisty při sledování jazykových změn v čase, dále pak pro sociology, novináře jako obraz společnosti, pro ekonomy politology atd. Jistě se některým z nás už stalo, že jsme klikli na nějaký internetový odkaz a objevilo se chybové hlášení: „404 - Page Not Found“ neboli stránka nebyla nalezena. Díky Internet Archive je možné oživení takového mrtvého odkazu.³⁸

Archiv v současnosti obsahuje přes 85 miliard webových stránek, přičemž se ukládají kompletní webové stránky včetně obrázků, audia, videa a nejrůznějšího softwaru.³⁹ Jsou vytvářeny i různé tematické kolekce webů třeba v souvislosti s prezidentskými volbami nebo teroristickými útoky z 11. září 2001 či sbírky kolem různých světových institucí. Zajímavou součástí archivu je také sbírka filmů z Prelingerova archivu. Jedná se o filmy natočené v letech 1903 – 1973 za reklamními, didaktickými a jinými účely, dokumentující každodenní

³⁶ <http://www.archive.org/>

³⁷ *Library of Congress* [online]. Last Updated: 08/22/2008 [cit. 2009-04-03]. Dostupný z WWW: <<http://www.loc.gov/about/>>.

³⁸ *Internet Archive* [online]. [cit. 2009-04-03]. Dostupný z WWW: <<http://www.archive.org/>>.

³⁹ Tamtéž

život v kultuře, ekonomice a politice ve Spojených státech.⁴⁰ V archivu lze nalézt i mnoho českých webů, např. můžeme zjistit, jak vypadaly stránky Státního ústředního archivu v Praze v roce 2003 (ačkoliv tyto webové stránky fungovaly už od roku 1999 na serveru ministerstva vnitra, nebyla zřejmě archivována verze starší než 2003).

Weby jsou z Internetu sklizeny a zpracovávány **softwarovými roboty**⁴¹, v současnosti používá Internet Archive ke sklizení (sběru) nástroj **Heritrix**. Samotná data jsou pak uložena na zálohovacích zařízeních **Petabox**, která byla navržena pracovníky archivu tak, aby byla schopna ukládat a zpracovávat jeden petabajt (milion gigabajtů) informací. Petaboxy jsou nenáročné na skladování a údržbu, lehce se přemísťují a mají malé nároky na spotřebu energie.⁴² Internet Archive sídlí v San Francisku, kde je také uložena celá jeho sbírka. Z důvodu ochrany dat před zničením byla vytvořena identická kopie archivu v Egyptě v Alexandrii (Bibliotheca Alexandrina).⁴³

Pro přístup uživatelů k archivu funguje webové rozhraní **Wayback Machine**,⁴⁴ které pomocí URL a volitelného nastavení času vyhledá požadovanou stránku. Pokud požadovaná stránka nebyla dosud archivována, dojde k jejímu sklizení za 24 hodin. Webová stránka je zobrazena tak, jak byla uložena do archivu, nedochází tedy k žádné úpravě ani migraci dat. Zatím nelze v archivu vyhledávat bez znalosti URL příslušné stránky. Doufejme, že v dohledné době bude možné vyhledávat web podle více parametrů, např. zadáním klíčových slov apod. Internet Archive tvoří svou sbírku prakticky bez jakéhokoliv souhlasu majitelů autorských práv webů či jejich obsahu.⁴⁵

Zájmy držitelů práv se archiv snaží chránit několika způsoby:

- Zpřístupňuje pouze stránky starší šest měsíců
- Vyřazením stránek z archivu na žádost majitele, buď vymazáním stávající kopie, nebo nepořizováním dalších kopií.
- Omezením přístupu ke stránkám

⁴⁰ VOJTÁŠEK, Filip. Archiv celosvětového webu zpřístupněn. *Ikaros* [online]. 2001, roč. 5, č. 12 [cit. 2009-04-03]. Dostupný na WWW: <<http://www.ikaros.cz/node/876>>. URN-NBN:cz-ik876. ISSN 1212-5075.

⁴¹ Softwarový robot je počítačový program vykonávající nějakou opakovanou činnost podle zadaných příkazů a kritérií. Dnes se nejčastěji setkáváme s internetovými roboty, kteří většinou sbírají ze sítě data nejrůznějšího typu. Roboty využívají např. internetové vyhledávače, ale existuje i mnoho škodlivých robotů, kteří sbírají emailové adresy a odesílají na ně nevyžádanou poštu (spam) apod.

⁴² *Internet Archive* [online]. [cit. 2009-04-03]. Dostupný z WWW: <<http://www.archive.org/>>.

⁴³ NOVOTNÝ, Luděk. *Internet Archive - celosvětový archiv webu* [online]. 2004 [cit. 2009-03-04]. Dostupný z WWW: <dSPACE.muni.cz/bitstream/ics_muni_cz/72/1/i-archiv.pdf>.

⁴⁴ <http://www.archive.org/web/web.php>

⁴⁵ NOVOTNÝ, Luděk. *Internet Archive - celosvětový archiv webu* [online]. 2004 [cit. 2009-03-04]. Dostupný z WWW: <dSPACE.muni.cz/bitstream/ics_muni_cz/72/1/i-archiv.pdf>.

- Zamezením robotovi sklidit stránku použitím SRE (Standard for Robot Exclusion)⁴⁶

Na závěr bychom se ještě měli zmínit o internetových vyhledávačích, které si také tvoří archivní kopie webů pro své vlastní účely. Takovým vyhledávačem je i populární **Google**.⁴⁷ Každý web, který **zaindexují**⁴⁸ jeho roboti je uložen jako záloha sloužící pro vnitřní hodnocení vyhledávače, zda stránka vyhovuje zadanému dotazu. Prakticky pokud zadáte nějaký dotaz do vyhledávače, Google zkoumá kritéria stránek, podle kterých vybírá a řadí výsledky dotazu, a to na webech uložených v jeho archivu a nikoliv přímo na aktuálních stránkách v Internetu. Archivní verze webových stránek je k dispozici i uživatelům. U většiny webů, které Google zobrazí jako výsledek dotazu, je odkaz **Archiv (Cache)**, který nás směřuje na poslední uloženou verzi stránky. V záhlaví okna prohlížeče je zobrazeno upozornění, že se nejedná o aktuální verzi a datum, kdy byla tato stránka uložena.

Google archiv najde uplatnění třeba v situaci, pokud nám nejde současná stránka zobrazit, např. v důsledku přetížení serveru nebo byly stránky nedávno vyjmuty z Internetu apod.⁴⁹

3.2 Knihovny

Ačkoliv jsou webové archivy jistě skvělým a velice významným projektem na poli archivace webů, nelze je považovat za jediné a dostačující řešení této problematiky. Spoléhání se na soukromou instituci jako je Internet Archive, že zajistí veškerý potřebný sběr a uchování webových stránek, je velmi neuvážené. Není zde žádná jistota, jak dlouho bude tato organizace fungovat, zda nezmění svou činnost nebo zda bude své archivy stále zpřístupňovat bezplatně, navíc v případě technických problémů může dojít k nenahraditelné ztrátě. Z předešlých důvodů je jasné, že archivaci webů v jednotlivých zemích by měly vzít do svých

⁴⁶ NOVOTNÝ, Luděk. *Internet Archive - celosvětový archiv webu* [online]. 2004 [cit. 2009-03-04]. Dostupný z WWW: <dspace.muni.cz/bitstream/ics_muni_cz/72/1/i-archiv.pdf>.

⁴⁷ <http://www.google.com/>

⁴⁸ Indexace je činnost, při které jsou hodnocena klíčová data a je vytvořen index sloužící pro zrychlení vyhledávacího procesu v databázích, seznamech, Internetu apod.

⁴⁹ *Google Cached Pages: What Are Cached Pages?* - *Google Guide* [online]. This page was last modified on: Tuesday March 13, 2007 [cit. 2009-04-05]. Dostupný z WWW: <http://www.googleguide.com/cached_pages.html>.

rukou státní instituce a to především knihovny⁵⁰, protože jejich pole působnosti je v oblasti uchovávání a zpřístupňování textových zdrojů určených k publikaci nejširší.

Problematikou dlouhodobého uchovávání elektronických zdrojů přístupných v síti Internet se zabývají národní knihovny mnoha států, mezi první patřily USA, Kanada, Austrálie a skandinávské země. Z jednotlivé i společné činnosti zemí zabývajících se uchováváním webových stránek vzniklo mnoho projektů jako třeba **BIBLINK**⁵¹, **CoBRA** nebo **NEDLIB**⁵², podporované Evropskou komisí a společný projekt národních knihoven severovýchodních zemí **The Nordic Web Archive (NWA)**⁵³.

V současnosti probíhají projekty: australský **PANDORA**⁵⁴, dánský **Netarchive.dk**⁵⁵, švédský **Kulturarw3**⁵⁶ a nizozemský **e-Depot**^{57, 58}. V roce 2003 vzniklo konsorcium **IIPC**⁵⁹ (**International Internet Preservation Consortium**), za účelem spolupráce knihoven na mezinárodní úrovni. V současné době je členem IIPC 37 knihoven a institucí z celého světa, včetně **Národní knihovny České republiky**.

Cíle konsorcia jsou:

- Umožnit sběr, uchovávání a neomezený přístup k bohatému internetovému obsahu z celého světa.
- Podporovat rozvoj a využití společných nástrojů, technik a standardů pro vytváření mezinárodních archivů.
- Prosazovat mezinárodní iniciativu a právní přepisy, které podporují sběr, uchovávání a přístup k internetovému obsahu.
- Pomáhat a podporovat knihovny, archivy, muzea a další instituce v shromažďování a uchovávání internetového obsahu jako kulturního dědictví.⁶⁰

Kromě obecných cílů se konsorcium nyní soustředí na vývoj komplexního souboru softwarových nástrojů (**IIPC Web Archiving Toolset**) a různých doporučení ohledně

⁵⁰ COUFAL, Libor. Archivace (nejen) českého webu z pohledu WebArchivu : Jak se na archivaci webu dívají autoři projektu WebArchiv?. *Čertoděj : čertovské věci o Web 2.0 a web trendech* [online]. 2007 [cit. 2009-04-07]. Dostupný z WWW: <<http://certodej.cz/S9314>>.

⁵¹ <http://hosted.ukoln.ac.uk/biblink/>

⁵² <http://nedlib.kb.nl/>

⁵³ <http://nwa.nb.no/>

⁵⁴ <http://pandora.nla.gov.au/>

⁵⁵ <http://netarchive.dk/>

⁵⁶ <http://www.kb.se/english/find/internet/websites/>

⁵⁷ <http://www.kb.nl/dnp/e-depot/e-depot-en.html>

⁵⁸ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 13 s. ISBN 978-80-7050-562-5.

⁵⁹ <http://www.netpreserve.org>

⁶⁰ *International Internet Preservation Consortium* [online]. c2004-2008 [cit. 2009-04-11]. Dostupný z WWW: <<http://www.netpreserve.org/about/index.php>>.

archivace webu na základě zkušeností členů konsorcia. Výsledky projektu by měly být volně k dispozici všem institucím, které se zabývají archivací webu, což umožní zapojení i těch zemí nebo institucí, jež nemají prostředky na vlastní vývoj.⁶¹

3.2.1 Legislativa

Knihovny se zabývají hlavně shromažďováním publikací jako součásti národního kulturního dědictví a budováním národní bibliografie. Tudíž jako na publikaci pohlížejí i na weby.⁶² Podporovány jsou přitom ve většině zemí zákonem o povinném výtisku.⁶³ U nás se jedná o **zákon č. 37/1995 Sb., o neperiodických publikacích** a dále o **zákon č. 46/2000 Sb., o právech a povinnostech při vydávání periodického tisku a o změně některých dalších zákonů (tiskový zákon)**. Za neperiodickou publikaci přitom zákon č. 37/1995 Sb. považuje: „*rozmnoženiny literárních, vědeckých a uměleckých děl určené k veřejnému šíření, které jsou vydávány jednorázově, popřípadě nejvýše jednou ročně anebo po částech i častěji, tvoří-li obsahově jeden celek.*“⁶⁴ Tato formulace by mohla být pro svou obecnost snadno aplikovatelná na elektronické dokumenty (publikace)⁶⁵ včetně dostupných online. Zákon přitom nepovažuje za neperiodické publikace mimo jiné také počítačové programy a audiovizuální díla⁶⁶, ty jsou částečně řešeny samostatným zákonem.⁶⁷ Veřejným šířením díla je myšleno zpřístupnění publikace blíže neurčenému okruhu osob.⁶⁸

Asi 95 % webových publikací jsou však zdroje periodické (průběžně vydávané), jimiž se zabývá další z již zmíněných dvou zákonů.⁶⁹ Tiskový zákon, pod který periodické publikace spadají, už situaci poněkud komplikuje, a to konkrétně vymezením pojmů: „*Pro účely tohoto zákona se rozumí: a) periodickým tiskem noviny, časopisy a jiné tiskoviny*

⁶¹ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 18 s. ISBN 978-80-7050-562-5.

⁶² MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchování webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 242.

⁶³ CELBOVÁ, Ludmila. Povinný výtisk elektronických publikací, zejména vzdálených elektronických zdrojů. *Ikaros* [online]. 2000, roč. 4, č. 10 [cit. 2009-03-28]. Dostupný na WWW: <<http://www.ikaros.cz/node/670>>. URN-NBN:cz-ik670. ISSN 1212-5075.

⁶⁴ *Zákon č. 37/1995 Sb., o neperiodických publikacích*. § 1, (1).

⁶⁵ Je dokument, v němž jsou informace digitálně zakódovány a který je přístupný v strojem čitelné podobě.

⁶⁶ CELBOVÁ, Ludmila. Povinný výtisk elektronických publikací, zejména vzdálených elektronických zdrojů. *Ikaros* [online]. 2000, roč. 4, č. 10 [cit. 2009-03-28]. Dostupný na WWW: <<http://www.ikaros.cz/node/670>>. URN-NBN:cz-ik670. ISSN 1212-5075.

⁶⁷ *Zákon č. 273/1993 Sb. o některých podmínkách výroby, šíření a archivování audiovizuálních děl*.

⁶⁸ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 10 s. ISBN 978-80-7050-562-5.

⁶⁹ Tamtéž

vydávané pod stejným názvem, se stejným obsahovým zaměřením a v jednotné grafické úpravě nejméně dvakrát v kalendářním roce,⁷⁰ Zde je jasné, že se jedná pouze o tiskoviny, tím pádem elektronická periodika nepodléhají povinnému výtisku (elektronické kopii).

Hlavním problémem těchto zákonů je, že se připravovaly v době, kdy bylo publikování elektronických dokumentů jen velmi okrajovou záležitostí. Málokdo počítal s tím, že nárůst počtu elektronických dokumentů bude tak rapidní. A nemusí se jednat pouze o dokumenty vydávané na nosičích všeho typu jako např. disky CD-ROM a DVD-ROM, ale v dnešní době stále více publikované prostřednictvím Internetu. Takové dokumenty v elektronické podobě mají sice svou specifickou formu, která si podmiňuje odlišný způsob uchovávání a zpřístupňování, a to především u online dokumentů, na druhou stranu se však z hlediska obsahu vůbec nemusejí lišit od klasických tištěných dokumentů, ať už se jedná o monografie či periodika (knihy, časopisy apod.).⁷¹

Povinným výtiskem elektronických publikací se zabývají mnohé země a v některých je tato problematika již ošetřena zákony. Pro srovnání si uveďme situace v jednotlivých zemích.

- Austrálie – V současnosti řeší novelu zákona z roku 1968 (Copyright Act), jehož součástí jsou ustanovení o povinném výtisku. Novela zahrnuje i povinný výtisk elektronických publikací včetně webových.
- Dánsko – Zde zákon o povinném výtisku platí od roku 2005 a týká se všech publikací bez ohledu na jejich formu. Vydavatelé webových stránek, jež odevzdávají povinný výtisk, musejí také sdělit knihovně přístupová data ke stránkám (pokud nějaká jsou). Uložené dokumenty jsou přístupné pouze z vymezených terminálů v určitých knihovnách a bez možnosti zřízení kopie dokumentu.
- Finsko – Archivací elektronických online publikací se zabývá Národní knihovna Finska. Zákon povinného výtisku vztahující se i na weby platí od roku 2008. Povinnosti vydavatelů jsou obdobné jako v Dánsku.
- Slovinsko – Zákon vztahující se k odevzdávání výtisku elektronických publikací přístupných na webu začal platit v roce 2006. Příjemcem publikací je Národní a univerzitní knihovna v Lublani a Univerzitní knihovna v Mariboru. Knihovny ke

⁷⁰ Zákon č. 46/2000 Sb., o právech a povinnostech při vydávání periodického tisku a o změně některých dalších zákonů (tiskový zákon). § 3.

⁷¹ CELBOVÁ, Ludmila. Povinný výtisk elektronických publikací, zejména vzdálených elektronických zdrojů. *Ikaros* [online]. 2000, roč. 4, č. 10 [cit. 2009-03-28]. Dostupný na WWW: <<http://www.ikaros.cz/node/670>>. URN-NBN:cz-ik670. ISSN 1212-5075.

staženým zdrojům poskytují veřejný online přístup. U webů, které mají omezený přístup, se knihovny snaží o uzavření dohody s jednotlivými vydavateli a k těmto webovým stránkám je možné přistupovat alespoň z určených terminálů knihoven.⁷²

Další překážka, kterou klade současná legislativa, se nachází v **autorském zákoně**.⁷³ Podle tohoto zákona je autorským dílem mimo jiné také dílo v elektronické podobě, vzniklé tvůrčí činností autora, vyjádřené v jakékoli objektivně vnímatelné podobě⁷⁴, tedy i webové stránky. Zákon umožňuje knihovnám vytvořit kopii webu pro uložení do tzv. **konzervačního fondu** elektronických publikací neboli elektronického (digitálního) archivu.⁷⁵ Pořizování kopií je definováno: „Rozmnožováním díla se rozumí zhotovování dočasných nebo trvalých, přímých nebo nepřímých rozmnoženin díla nebo jeho části, a to jakýmkoli prostředky a v jakékoli formě.“⁷⁶

Paragraf č. 37 o Knihovní licenci pak upravuje vztah samotných knihoven a dalších institucí k autorskému právu, a to takto:

„Do práva autorského nezasahuje knihovna, archiv, muzeum, galerie, škola, vysoká škola a jiné nevýdělečné školské a vzdělávací zařízení⁷⁷, a) zhotoví-li rozmnoženinu díla, která neslouží k přímému nebo nepřímému hospodářskému nebo obchodnímu účelu, pro své archivní a konzervační potřeby,“⁷⁸

Zde je vysloveno oprávnění knihovny k vytvoření fondu. Bohužel hned dále nacházíme v zákoně onen zmíněný problém:

„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, 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,“⁷⁹.

Z tohoto tvrzení lze vyvodit, že knihovna může zpřístupnit svůj konzervační fond pouze ve svých objektech, a to jednotlivým osobám, za podmínek, aby nebylo možno zhotovit

⁷² CELBOVÁ, Ludmila. Český web a povinný výtisk – jde to spolu dohromady? *Knihovna* [online]. 2008, roč. 19, č. 2, s. 59-75 [cit. 2009-03-28]. Dostupný z WWW: <<http://knihovna.nkp.cz/knihovna82/82005.htm>>. ISSN 1801-3252.

⁷³ Zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským.

⁷⁴ Tamtéž

⁷⁵ CELBOVÁ, Ludmila. Povinný výtisk elektronických publikací, zejména vzdálených elektronických zdrojů. *Ikaros* [online]. 2000, roč. 4, č. 10 [cit. 2009-03-28]. Dostupný na WWW: <<http://www.ikaros.cz/node/670>>. URN-NBN:cz-ik670. ISSN 1212-5075.

⁷⁶ Zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským. § 13, (1).

⁷⁷ 4) Zákon č. 257/2001 Sb., o knihovnách a podmínkách provozování veřejných knihovnických a informačních služeb (knihovní zákon), ve znění pozdějších předpisů.

⁷⁸ Zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským. § 37, (1).

⁷⁹ Zákon č. 216/2006 Sb., změna autorského zákona a některých dalších zákonů. § 37, (1).

kopii. Výše zmíněná citace zákona také hovoří o výjimce: „sdělování díla způsobem uvedeným v § 18 odst. 2“, která nám ukládá další omezení: „Sdělováním díla veřejnosti podle odstavce 1 je také zpřístupňování díla veřejnosti způsobem, že kdokoli může mít k němu přístup na místě a v čase podle své vlastní volby zejména počítačovou nebo obdobnou sítí.“. Knihovny tedy nemají možnost zpřístupňovat archivované zdroje prostřednictvím Internetu, což se jeví jako velké omezení.

S autorským právem z části souvisí licence **Creative Commons (CC)**. Tyto licence jsou výtvorem stejnojmenné americké neziskové organizace⁸⁰ a jsou nositeli myšlenky volného přístupu k informacím.⁸¹ Jedná se o to, že někteří majitelé nechtějí využívat všechna práva, která mají na své duševní vlastnictví, ale v mnoha státech tyto práva získávají automaticky bez možnosti jakkoliv je upravit podle své potřeby. Například takový začínající fotograf může mít zájem na šíření svých fotek pro nekomerční účely. Creative Commons slouží vlastně pro jakousi úpravu copyrightu.⁸²

Autor si sám vymezí práva ke svému dílu, přičemž má na výběr několik možností:

- Citování autora – uživatel má právo s dílem nakládat a vytvářet odvozená díla za podmínky, že autora původního díla bude náležitě citovat.
- Uživatel má právo s dílem nakládat a vytvářet odvozená díla za podmínky, že dílo nebude využito ke komerčním účelům.
- Uživatel je oprávněn nakládat pouze s původním a originálním dílem. Nesmí vytvářet další díla, která jsou od toho původního odvozena.
- Uživatel má právo nakládat a vytvářet odvozená díla za podmínky, že odvozenina bude distribuována za naprosto stejných podmínek jako dílo, ze kterého vznikla.⁸³

Všechny tyto základní podmínky je možné zkombinovat do šesti typů licencí:

⁸⁰ <http://creativecommons.org/>

⁸¹ GRUBER, Lukáš. Licence Creative Commons a perspektiva jejich zavedení do českého prostředí. *Ikaros* [online]. 2008, roč. 12, č. 3 [cit. 2009-04-05]. Dostupný na WWW: <<http://www.ikaros.cz/node/4612>>. URN-NBN:cz-ik4612. ISSN 1212-5075.

⁸² ZEMAN, Mirek. Creative Commons – budoucnost copyrightu?. *LUPA: SERVER O ČESKÉM INTERNETU* [online]. 2004 [cit. 2009-04-12]. Dostupný z WWW: <<http://www.lupa.cz/clanky/creative-commons-8211-budoucnost-copyrightu/>>. ISSN 1213-0702.

⁸³ GRUBER, Lukáš. Licence Creative Commons a perspektiva jejich zavedení do českého prostředí. *Ikaros* [online]. 2008, roč. 12, č. 3 [cit. 2009-04-05]. Dostupný na WWW: <<http://www.ikaros.cz/node/4612>>. URN-NBN:cz-ik4612. ISSN 1212-5075.

- by, uvedení autora
- by – sa, uvedení autora, zachování licence
- by – nd, uvedení autora, žádné odvozeniny
- by-nc, uvedení autora, nekomerční využití
- by-nc-sa, uvedení autora, nekomerční využití, zachování licence
- by-nc-nd , uvedení autora, nekomerční využití, žádné odvozeniny⁸⁴

Licence Creative Commons se nejvíce uplatňují na webové stránky či jiné elektronické zdroje na Internetu, ale může se jednat i o kterákoliv jiná díla ať už v elektronické či jiné podobě. V praxi to funguje tak, že autor navštíví stránky Creative Commons a vyplní zde příslušný dotazník, poté mu generátor na základě jeho odpovědi v dotazníku vyhotoví příslušný typ licence v podobě krátkého HTML kódu. Tento kód pak autor zkopíruje do zdrojového kódu svých stránek. Na stránkách se objeví logo CC se stručným vysvětlením pro uživatele těchto stránek, jaká práva a povinnosti se ke stránkám vztahují. Česká legislativa umožnila aplikaci CC licencí v ČR, novelou autorského zákona z roku 2006:

„(5) O podání návrhu na uzavření smlouvy jde i tehdy, směřuje-li projev vůle i vůči neurčitému okruhu osob.

(6) S přihlédnutím k obsahu návrhu nebo k praxi, kterou strany mezi sebou zavedly, nebo zvyklostem může osoba, které je návrh určen, vyjádřit souhlas s návrhem na uzavření smlouvy provedením určitého úkonu bez vyrozumění navrhovatele tím, že se podle ní zachová, zejména že poskytne nebo přijme plnění. V tomto případě je přijetí návrhu účinné v okamžiku, kdy byl tento úkon učiněn.“⁸⁵

Oba odstavce v podstatě upravují uzavření smlouvy v prostředí Internetu, kde jsou dokumenty publikované pro nejširší veřejnost a není dost dobře možné identifikovat koncového uživatele stránek, tedy uzavřít smlouvu s konkrétní osobou.⁸⁶ Za účelem zavedení CC licencí do českého prostředí vznikla pracovní skupina⁸⁷, jejímž hlavním cílem bylo: „Vytvoření české verze plného znění a zkráceného znění licence Creative Commons“⁸⁸, což se nakonec podařilo a 16. dubna 2009 byla na serveru organizace Creative Commons oficiálně

⁸⁴ GRUBER, Lukáš. Licence Creative Commons a perspektiva jejich zavedení do českého prostředí. *Ikaros* [online]. 2008, roč. 12, č. 3 [cit. 2009-04-05]. Dostupný na WWW: <<http://www.ikaros.cz/node/4612>>. URN-NBN:cz-ik4612. ISSN 1212-5075.

⁸⁵ *Zákon č. 216/2006 Sb., změna autorského zákona a některých dalších zákonů.* § 46.

⁸⁶ GRUBER, Lukáš. Licence Creative Commons a perspektiva jejich zavedení do českého prostředí. *Ikaros* [online]. 2008, roč. 12, č. 3 [cit. 2009-04-23]. Dostupný na WWW: <<http://www.ikaros.cz/node/4612>>. URN-NBN:cz-ik4612. ISSN 1212-5075

⁸⁷ <http://www.creativecommons.cz/>

⁸⁸ *Creative Commons Česká republika* [online]. 2008 [cit. 2009-04-23]. Dostupný z WWW: <<http://www.creativecommons.cz/o-projektu/>>.

spuštěna česká verze CC licencí. Autoři tedy mají možnost vystavovat svá díla pod licencí, která je v národním jazyce a plně v souladu s českým právem.⁸⁹ Creative Commons licence jsou velkým pokrokem na poli zpřístupňování a legálního využívání děl, zvláště v oblasti Internetu. Autorům poskytují možnost definovat si svá vlastní práva ke svým výtvorům a pro uživatelskou veřejnost představují způsob, jak snadno a srozumitelně vyložit autorské právo ke konkrétnímu dílu.

Legislativní překážky omezení archivace a následného zpřístupňování webů nejsou pouze problémem České republiky, ale i mnoha dalších zemí zejména v Evropě, Severní Americe a Austrálii. V těchto zemích se autorská práva vztahují na díla 70 let po smrti autora. Výjimkou je např. Jižní Korea, chránící dílo pět let od jeho vydání.⁹⁰ Náhradním řešením, ke kterému knihovny přistupují, je oslovení jednotlivých vydavatelů webových stránek, kteří disponují autorskými právy k uzavření smlouvy o poskytování elektronických online zdrojů. Toto řešení je však pouze provizorní, jelikož je velice časově i personálně náročné⁹¹ a neposkytuje téměř žádné záruky, že se podaří archivovat všechny žádané weby. Bez zlepšení přímé podpory legislativy, a zde se nemusí jednat pouze o úpravu tiskového a autorského zákona, ale např. i o otázky finanční, technické a problematiku sběru a zpřístupnění webů, nebude do budoucna možné plně realizovat vytvoření kvalitního a široké veřejnosti přístupného digitálního fondu v celé jeho požadované velikosti.

Nepostačující legislativy si všimla i organizace UNESCO⁹² a v roce 2003 přijala na své generální konferenci **Chartu o zachování digitálního dědictví**, zabývající se zákonodárstvím o zákonném uložení publikace a archivním zákonodárstvím, jako velmi důležitým prvkem z hlediska národní politiky pro zachování digitálního dědictví. Další nadnárodní institucí snažící se řešit problematiku uchovávání digitálních dokumentů a zejména s tím související legislativou je Rada Evropské unie⁹³. Ta vydává různá doporučení ohledně plnění prioritních úkolů vztahujících se k oblasti archivování digitálního obsahu

⁸⁹ *Creative Commons Česká republika* [online]. 2008 [cit. 2009-04-23]. Dostupný z WWW: <<http://www.creativecommons.cz/2009/03/24/cc-launch-oficialni-spusteni-ceskych-cc-licenci/>>.

⁹⁰ COUFAL, Libor. Archivace (nejen) českého webu z pohledu WebArchivu : Jak se na archivaci webu dívají autoři projektu WebArchiv?. *Čertoděj : čertovské věci o Web 2.0 a web trendech* [online]. 2007 [cit. 2009-04-02]. Dostupný z WWW: <<http://certodej.cz/S9314>>.

⁹¹ HUTAŘ, Jan, MATĚJKA, Lukáš, CELBOVÁ, Ludmila. Vyhledávání v archivu českých webových zdrojů. *INFORUM 2006: 12. konference o profesionálních informačních zdrojích* [online]. 2006 [cit. 2009-04-02]. Dostupný z WWW: <http://www.inforum.cz/pdf/2006/Hutar_Jan.pdf>.

⁹² United Nations Educational, Scientific and Cultural Organization. www.unesco.org/

⁹³ <http://ue.eu.int/>

světového dědictví, především pak posiluje koordinaci mezi jednotlivými státy ve vztahu k souvisejícím problémům.⁹⁴

3.2.2 Co a jak uchovávat

Co a jak uchovávat je další základní otázka (otázky) na poli uchovávání webů a můžeme si pod ní představit činnosti spojené s **výběrem** a **sběrem (sklizením)** webových stránek. Instrukce zabývající se dlouhodobým uchováváním webů přistupují k problematice výběru odlišně, a to především z hlediska jejich zaměření. Některé z nich shromažďují a zachovávají všechny webový obsah Internetu jako např. nám již známé webové archivy⁹⁵. Některé další instituce sbírají webové stránky tematického zaměření souvisejícího s jejich vlastním. Například americký Úřad pro letectví a vesmír (**NASA**)⁹⁶ shromažďuje všechny údaje o vesmíru. Drtivá většina knihoven však nemá prostředky na to, aby vytvářela archiv celosvětového webu a navíc to ani neumožňují legislativy některých států. Knihovny jednotlivých států, a zde se jedná hlavně o ty národní, se soustředí na sběr stránek z národní domény⁹⁷ nebo těch, které jsou přístupné v jazyce dané země.⁹⁸

Sběrem webových stránek je myšlena činnost, kdy je vytvořena digitální kopie díla (webu) a ta se později stává součástí sbírky digitálního fondu (archivu). V oblasti sběru se nejčastěji uplatňují dva přístupy: **ruční** a **automatizovaný**.⁹⁹ Strategie výběru webů se stanovuje na základě kritérií, jako jsou velikost knihovny, její účel a oborové zaměření, skladovací kapacita, rozpočet apod. Při procesu výběru se zohledňují i další faktory, např. koncová velikost uloženého materiálu, četnost sběru nebo podmínky zpřístupnění fondu z hlediska legislativy.¹⁰⁰ Ruční výběr shromažďuje pouze dokumenty, které vybrala pověřená osoba, v našem případě knihovník. Takový postup zvolila třeba **Australská národní knihovna**¹⁰¹. Do jejího digitálního fondu putují pouze dokumenty, jež prošly zhodnocením příslušného zaměstnance knihovny. Tento systém sebou nese jisté výhody i nevýhody.

⁹⁴ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 9 s. ISBN 978-80-7050-562-5.

⁹⁵ viz kapitola **3.1 Webové archivy**

⁹⁶ National Aeronautics and Space Administration. <http://www.nasa.gov/>

⁹⁷ viz kapitola **2.5.2 Domény**

⁹⁸ MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchovávání webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 243.

⁹⁹ Tamtéž

¹⁰⁰ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 22 s. ISBN 978-80-7050-562-5.

¹⁰¹ <http://www.nla.gov.au/>

Vybrané webové stránky mají většinou velkou informační hodnotu a jedná se o výběr toho nejdůležitějšího, co bylo na webu publikováno. Také knihovnický popis bývá velmi podrobný, často na úrovni jednotlivých dokumentů.

Na druhou stranu je výsledný obsah fondu poměrně malý (což může být ale i výhoda z hlediska menších nároků na velikost úložného prostoru). Některé důležité a hodnotné dokumenty mohou být opomenuty. Tento přístup však klade velké nároky na lidské kapacity.¹⁰²

Většina knihoven však používá systém plošného automatizovaného sběru pomocí speciálních softwarových nástrojů vyvinutých k tomuto účelu. Tyto nástroje zvané **sklízeče (harvesters)**¹⁰³ jsou vlastně jedním z typů softwarových robotů¹⁰⁴. Webový sklízeč stahuje a ukládá obsah webu podle předem nastavených parametrů. Sklízeč se naplní okruhem URL odkazů¹⁰⁵, které určují webové stránky k archivování (většinou vybrané podle národní domény).¹⁰⁶ První sklízeče byly naprogramovány již v roce 1994, avšak jejich funkce byly omezené, stažené dokumenty pouze zaindexovaly, ale nearchivovaly. Za tímto účelem byl v rámci projektů NEDLIB¹⁰⁷ a NWA¹⁰⁸ vyvinut **Finskou národní knihovnou**¹⁰⁹ nástroj **NEDLIB harvester**. Tento mezinárodně používaný sklízeč byl schopen pracovat s protokoly HTTP¹¹⁰ a FTP¹¹¹ a obsahoval dva režimy sběru: plný a doplňovací. Plným sběrem se rozumí uložení všech dokumentů z webu při každé návštěvě harvesteru. Doplňovací neboli inkrementální sběr ukládá pouze ty dokumenty, jež se od poslední prohlídky webu harvesterem změnily nebo se přestěhovaly na jinou adresu.¹¹²

Stahované dokumenty sklízeč ukládal do pracovního adresáře, kde proběhla jejich analýza, zpracování a archivace. Nejprve došlo ke stažení externích souborů z dokumentu (pokud nějaké obsahoval) jako video, obrázky, applety¹¹³ apod. a teprve potom sklízeč stáhl

¹⁰² MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchování webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 243.

¹⁰³ Slovo harvester označuje v anglickém jazyce kombajn.

¹⁰⁴ viz kapitola **3.1 Webové archivy**

¹⁰⁵ viz kapitola **2.2 Služba WWW**

¹⁰⁶ MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchování webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 245.

¹⁰⁷ Networked European Deposit Library (viz kapitola **3.2 Knihovny**)

¹⁰⁸ Nordic Web Archive (viz kapitola **3.2 Knihovny**)

¹⁰⁹ <http://www.nationallibrary.fi/>

¹¹⁰ viz kapitola **2.2 Služba WWW**

¹¹¹ viz kapitola **2.1 Internet**

¹¹² ŽABIČKA, Petr. NEDLIB Harvester. *Ikaros* [online]. 2000, roč. 4, č. 10 [cit. 2009-04-30]. Dostupný z WWW: <<http://www.ikaros.cz/node/672>>. URN-NBN:cz-ik672. ISSN 1212-5075.

¹¹³ viz kapitola **2.4.4 Flash animace a Java applety**

samotný HTML dokument¹¹⁴. Stažené dokumenty se uložily v balících asi po tisíci souborech, přičemž celý tento komplet souborů byl **zkomprimován**¹¹⁵ pomocí programů **tar** a **gzip**. Soubory pak byly ukládány na velkokapacitní magnetové pásky.¹¹⁶

Mezi nevýhody NEDLIB harvesteru patřila absence podpory javaskriptu¹¹⁷ a flash formátu^{118, 119}. V současné době se již sklízeč NEDLIB nepoužívá a na jeho místo nastoupily propracovanější nástroje, jako je např. již zmiňovaný Heritrix¹²⁰, které byly vytvořeny v reakci na neustále probíhající změny v technologiích a formátech vyskytujících se na webu a také podle rozrůstajících se potřeb a kritérií v oblasti dlouhodobého uchování webových stránek.

Dalším důležitým problémem archivace webů je otázka zajištění čitelnosti uložených souborů i ve vzdálenější budoucnosti (několik desetiletí či dokonce staletí). Tento problém odpadá u **otevřených formátů**¹²¹, u **uzavřených (proprietárních) formátů**¹²² je možná **konverze dat**¹²³ do některého z otevřených formátů. V budoucnu nelze vyloučit ani případnou **migraci**¹²⁴ uložených dat.¹²⁵

V další kapitole se podíváme na činnosti, které vyvíjí Národní knihovna České republiky ohledně archivace webů a právě na jejím příkladě si podrobněji rozebereme některé doposud uvedené problematiky spojené s dlouhodobým uchováváním webových stránek včetně jejich řešení.

¹¹⁴ viz kapitola **2.2 Služba WWW**

¹¹⁵ Komprimace či komprese dat je postup pro zmenšení velikosti datových souborů

¹¹⁶ ŽABIČKA, Petr. NEDLIB Harvester. *Ikaros* [online]. 2000, roč. 4, č. 10 [cit. 2009-04-30]. Dostupný z WWW: <<http://www.ikaros.cz/node/672>>. URN-NBN:cz-ik672. ISSN 1212-5075.

¹¹⁷ viz kapitola **2.4.2 Skripty**

¹¹⁸ viz kapitola **2.4.4 Flash animace a Java applety**

¹¹⁹ MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchování webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 246.

¹²⁰ viz kapitola **3.1 Webové archivy**

¹²¹ Otevřený formát (open format) sebou nenese právní omezení na jeho používání. Hlavním smyslem je zajistit dlouhodobou čitelnost a dále zamezit vytvoření monopolu proprietárního (uzavřeného) formátu kontrolovaného jedním subjektem. Příklady otevřených formátů: JPG, HTML, PDF...

¹²² viz otevřený formát

¹²³ znamená převod či změnu

¹²⁴ Migrace dat je, podobně jako konverze, převod dat z jednoho formátu do druhého. O migraci dat mluvíme většinou v případě, kdy je zastaralý formát nahrazován novým, např. při změně informačního systému nebo databáze. Základní rozdíl mezi migrací a konverzí spočívá v tom, že při migraci dochází nejenom ke změně formátu jako u konverze, ale mění se i místo, resp. data se přesouvají.

¹²⁵ MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchování webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 247.

3.2.3 Národní knihovna České republiky a dlouhodobé uchovávání webů

V České republice se archivací webových stránek zabývá **Národní knihovna**¹²⁶. Národní knihovna České republiky je státní příspěvkovou organizací a jejím zřizovatelem je Ministerstvo kultury ČR¹²⁷. Zákon definuje knihovnu takto: „*Národní knihovna je knihovnou s univerzálním knihovním fondem doplněným specializovanými fondy. Národní knihovna trvale uchovává konzervační fond a historický fond.*“¹²⁸ Mezi hlavní úkoly Národní knihovny patří trvalé uchovávání a doplňování fondu bohemikálních dokumentů (bohemika) vydaných v ČR i v zahraničí jako součásti národního kulturního dědictví. Takzvaná **bohemika**¹²⁹ tvoří **českou národní bibliografii (ČNB)**, kterou zpracovává a publikuje právě Národní knihovna.¹³⁰ ČNB představuje nejucelenější zdroj bibliografických informací o českém písemnictví a o produkci vydávané na území ČR vůbec. Aktuální vydání ČNB je realizováno čtvrtletně vždy v dubnu, červnu, říjnu a lednu. Databáze ČNB je vydávána na CD a DVD nosičích nebo je zdarma přístupná na Internetu na adrese <http://aip.nkp.cz/>¹³¹ a také v online katalogu Národní knihovny¹³². Snaha české národní bibliografie o zachycení nejenom tradičních tištěných, ale i elektronických dokumentů, podnítila zájem Národní knihovny o elektronické zdroje publikované v síti Internet. Knihovna přitom rozlišuje **elektronickou publikaci**¹³³ a **sít'ovou elektronickou publikaci**^{134 135}.

¹²⁶ <http://www.nkp.cz/>

¹²⁷ <http://www.mkcr.cz/>

¹²⁸ Zákon č. 257/2001 Sb., o knihovnách a podmínkách provozování veřejných knihovnických a informačních služeb (knihovní zákon). § 9, (1).

¹²⁹ Dokument vydaný na území dnešní České republiky (územní bohemikum), napsaný autory původem z Česka (autorské bohemikum), napsaný v českém jazyce (jazykové bohemikum) nebo pojednávající o České republice (obsahové bohemikum).

¹³⁰ CELBOVÁ, Ludmila. Registrace a zpřístupňování elektronických zdrojů publikovaných v síti Internet. *INFORUM : konference o profesionálních informačních zdrojích* [online]. 2000 [cit. 2009-05-07]. Dostupný z WWW: <<http://www.inforum.cz/archiv/inforum2000/prednasky/registracezp.htm>>.

¹³¹ ČNB: Česká národní bibliografie [online]. [2008] [cit. 2009-05-07]. Dostupný z WWW: <http://www.aipberoun.cz/Site/produkty_cnb_cz.htm>.

¹³² http://sigma.nkp.cz/F/?func=file&file_name=find-a&local_base=cnb

¹³³ Je publikovaný dokument, v němž jsou informace digitálně zakódovány a který je přístupný ve strojem čitelné podobě. Zahrnuje publikace off-line (fyzická forma) jako CD-ROMy a diskety a publikace on-line uchovávané použitím digitálních technologií.

¹³⁴ Je digitálně zakódovaný informační zdroj, který je zpřístupněný veřejnosti prostřednictvím komunikační sítě. Všechny dálkově přístupné (sít'ové) elektronické zdroje jsou z hlediska bibliografického popisu považovány za publikované.

¹³⁵ CELBOVÁ, Ludmila. Registrace a zpřístupňování elektronických zdrojů publikovaných v síti Internet. *INFORUM : konference o profesionálních informačních zdrojích* [online]. 2000 [cit. 2009-05-07]. Dostupný z WWW: <<http://www.inforum.cz/archiv/inforum2000/prednasky/registracezp.htm>>.

Síťové publikace však mají svá specifika a vyžadují odlišný přístup v získávání, evidenci, ukládání a zpřístupnění. Právě na tyto otázky měly za úkol odpovědět projekty, které vznikly na půdě Národní knihovny.¹³⁶

Knihovna a její projekty se také často zapojují do mezinárodní spolupráce, což jí umožňuje získat cenné zkušenosti a informace. Například mezinárodní projekt **Web Cultural Heritage** probíhající v letech 2005-2006, jehož iniciátorem a koordinátorem byla právě Národní knihovna ČR. Projektu se dále účastnily: Univerzitní knihovnice v Bratislavě, Národní a univerzitní knihovny Slovinska a Národní knihovny Estonska. Jejich cílem bylo vytvořit seznam kritérií výběru webových zdrojů určený pro kulturní paměťové instituce jako obecné doporučení. Projekt se skládal ze dvou okruhů činností, které probíhaly paralelně. Prvním z nich bylo provedení analýzy metod a politiky výběru v mnoha zemích z celého světa. Dále pak analýza používaných technologií a kritérií každé ze zúčastněných institucí. Druhý okruh porovnával software (harvestery) pro sklizení webů. Použity byly tři typy sklízečů (harvesterů) – **Heritrix**, **HTTrack**, **WebBird**. Výsledkem projektu byl návrh směrnice – **Obecná doporučení kritérií pro archivaci webu**^{137, 138}.

Od května roku 2007 se Národní knihovna ČR stala členem mezinárodního konsorcia **IIPC**¹³⁹. Konsorcium si vytyčilo jako hlavní okruhy vývoje pro léta 2007-2009 tyto čtyři oblasti: sklizení (**Harvesting**), zpřístupnění (**Access**), uchování (**Preservation**), standardizace (**Standards**). Zástupci z České republiky se stali členy pracovních skupin pro oblasti Access a Preservation.¹⁴⁰

Náplně jednotlivých oblastí jsou:

- Harvesting - vytvoření chytrého sklízeče, zlepšení kvality sklizní (především sklizení videa a hlubokého webu¹⁴¹).¹⁴²
- Access – vylepšení stávajících nástrojů **WERA**¹⁴³ a **WayBack Machine**¹⁴⁴, vyhledávání

¹³⁶ viz následující kapitola **3.2.3.1 Webarchiv**

¹³⁷ <http://www.webarchiv.cz/culture-2000/>

¹³⁸ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 17 s. ISBN 978-80-7050-562-5.

¹³⁹ viz kapitola **3.2 Knihovny**

¹⁴⁰ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 19 s. ISBN 978-80-7050-562-5.

¹⁴¹ Hluboký web (deep web) představuje všechnen obsah webu, který není zaindexovatelný vyhledávačem (vyhledávač ho nenajde). Pověštinou se jedná o stránky, které vyžadují autorizovaný přístup atd. Jeho opakem je povrchový web (surface web).

¹⁴² *International internet preservation consortium* [online]. c2004-2008 [cit. 2009-05-10]. Dostupný z WWW: <<http://www.netpreserve.org/about/workinggroups.php>>.

¹⁴³ WERA (Web Archiv Access) je nástroj vyvinutý v projektu NWA (viz kapitola **3.2 Knihovny**) a slouží pro přístup a vyhledávání v archivu organizace Internet Archive (viz kapitola **3.1 Webové archivy**)

¹⁴⁴ viz kapitola **3.1 Webové archivy**

v databázích a vývoj nástrojů pro **data mining**¹⁴⁵.

- Preservation – doporučení pro uchovávání webových stránek vycházející ze standardů pro digitální dokumenty¹⁴⁶

Mezinárodní spolupráci a svou účastí na mezinárodních projektech získala Národní knihovna také možnost využití různých softwarových nástrojů vyvinutých v zahraničí (**NEDLIB harvester**, **Heritrix**, **NWA Toolset**¹⁴⁷), které významně napomohly k rozvoji vlastních projektů.

3.2.3.1 WebArchiv

Národní knihovna ČR zpracovala návrh projektu zabývajícího se získáváním, uchováváním a zpřístupňováním síťových elektronických publikací.¹⁴⁸ V roce 1999 jí byl na tento projekt udělen grant Ministerstvem kultury ČR v rámci projektů vědy a kultury probíhajících v letech 2000-2001. Tento pilotní projekt nazvaný **Registrace, ochrana a zpřístupnění domácích elektronických zdrojů v síti Internet** byl realizován ve spolupráci s Moravskou zemskou knihovnou a Ústavem výpočetní techniky Masarykovy univerzity v Brně, jež projektu poskytoval informačně technologické služby. Spolupráce s těmito institucemi pokračuje doposud.¹⁴⁹

Hlavními problémy, na které se projekt zaměřil, byly:

- Určení kritérií pro výběr elektronických zdrojů pro registraci a archivaci vzhledem k politice budování fondů depozitních knihoven.
- Zajištění trvalého přístupu k uloženým publikacím.
- Způsob získávání elektronických zdrojů pro konzervační fond.
- Stanovení podmínek přístupu k archivovaným elektronickým zdrojům v souvislosti s autorským právem.
- Ověřování autentičnosti elektronických publikací.

¹⁴⁵ Data mining představuje získávání užitečných informací z dat. V oblasti informačních technologií jde zejména o informace z databází.

¹⁴⁶ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 19 s. ISBN 978-80-7050-562-5.

¹⁴⁷ NWA Toolset jedná se o nástroj sloužící pro zpřístupnění archivu. Jeho nástupcem je nástroj WERA (viz poznámka 143).

¹⁴⁸ CELBOVÁ, Ludmila. Elektronické zdroje publikované v síti Internet jako součást České národní bibliografie. *Ikaros* [online]. 2000, roč. 4, č. 6 [cit. 2009-05-10]. Dostupný na WWW: <<http://www.ikaros.cz/node/602>>. URN-NBN:cz-ik602. ISSN 1212-5075.

¹⁴⁹ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 14 s. ISBN 978-80-7050-562-5.

- Standardy a nástroje pro ukládání a vyhledávání dat o elektronických publikacích i pro ukládání a vyhledávání primárních zdrojů.¹⁵⁰

Pro tento dosud běžící projekt vznikl časem podprojekt nazývaný **WebArchiv**¹⁵¹ neboli také **archiv českého webu**. Posláním WebArchivu je v podstatě výběr, sběr, uchovávání a zpřístupnění webů, včetně veškeré problematiky s tím související.

Tu můžeme shrnout do celků:

- Legislativa (povinný výtisk – tiskový zákon, zpřístupnění archivu – autorský zákon)
- Strategie sklizení (výběr dokumentů k archivaci)
- Uložení sklizených webových stránek (datové formáty, popis uložených stránek)
- Technické podmínky zpřístupnění archivu

3.2.3.1.1 Legislativní překážky a jejich řešení

Otázky legislativy jsme zde již probírali v jedné z předchozích kapitol.¹⁵² Odevzdávání povinného výtisku je u webových publikací řešeno obráceným postupem než u tištěných a to tak, že si knihovna sama stáhne požadovaná data. Vydavatelé síťové publikace tak nevznikají žádné náklady ani starosti s odevzdáváním povinného výtisku. Novela autorského zákona z roku 2006 umožnila knihovnám zhotovit kopii (rozmnoženinu) síťové publikace pro uložení ve vlastním konzervačním fondu, což odstranilo problém s „nepovinností“ poskytnutí povinného výtisku periodických síťových publikací. Stále však zůstávají nevyřešeny licencované dokumenty nebo ty webové stránky, na jejichž přístup je zapotřebí registrace.¹⁵³ Co se týče podmínek zpřístupňování archivu je knihovna oprávněna poskytovat přístup pouze z terminálů umístěných v budovách knihovny, což je jediný způsob, jak se dostat ke všem datům obsaženým ve WebArchivu.¹⁵⁴

WebArchiv se snaží k vybraným archivovaným webům poskytovat online přístup a to způsobem uzavření smlouvy s jednotlivými vydavateli o poskytování elektronických online zdrojů. Seznam spolupracujících vydavatelů je k dispozici na stránkách WebArchivu¹⁵⁵.

¹⁵⁰ CELBOVÁ, Ludmila. Elektronické zdroje publikované v síti Internet jako součást České národní bibliografie. *Ikaros* [online]. 2000, roč. 4, č. 6 [cit. 2009-05-10]. Dostupný na WWW: <<http://www.ikaros.cz/node/602>>. URN-NBN:cz-ik602. ISSN 1212-5075.

¹⁵¹ <http://www.webarchiv.cz/>

¹⁵² viz kapitola **3.2.1 Legislativa**

¹⁵³ CELBOVÁ, Ludmila. Český web a povinný výtisk – jde to spolu dohromady? *Knihovna* [online]. 2008, roč. 19, č. 2, s. 59-75 [cit. 2009-05-12]. Dostupný z WWW: <<http://knihovna.nkp.cz/knihovna82/82005.htm>>. ISSN 1801-3252.

¹⁵⁴ *WebArchiv - archiv českého webu* [online]. Poslední aktualizace: 12.5.2009 [cit. 2009-05-12]. Dostupný z WWW: <<http://www.webarchiv.cz/vydavatele/>>.

¹⁵⁵ <http://www.webarchiv.cz/partneri/>

Velmi výhodnou alternativou k uzavření smlouvy je vystavení webových stránek pod licenci Creative Commons¹⁵⁶. Výsledek je přitom stejný jako u smlouvy a navíc odpadá zbytečná administrativa a časová náročnost oslovování vydavatelů.¹⁵⁷

Výše zmíněné problémy se řešitelé projektu WebArchiv snaží odstranit návrhem nové legislativy upravující povinné odevzdávání síťových publikací. Podklady pro návrh byly předloženy Ministerstvu kultury ČR v únoru 2008. Přípravovaná legislativa byla navržena ve variantě samostatného zákona o povinných síťových publikacích, který by pouze doplňoval oba stávající zákony o povinných výtiscích.¹⁵⁸ Zákon se také zabývá volným přístupem do archivu u publikací, které autor volně uveřejnil na Internetu bez omezení (přístup na heslo atd.). V současné době leží návrh zákona na Ministerstvu kultury, kde je podrobněji zkoumán z právního hlediska a tento rok by měla být provedena jeho analýza. Předpokládá se, že hotový zákon by se mohl dostat na půdu vlády ČR v druhé polovině roku 2010.¹⁵⁹

3.2.3.1.2 Strategie sklizení

V rámci výběru a sběru webových dokumentů používá WebArchiv dvě základní **metodiky sklizení: celoplošné a výběrové**. Celoplošné sklizeně probíhají pomocí harvesterů (sklížečů)¹⁶⁰ a jejich úkolem je zachytit veškeré dokumenty pod národní doménou (.cz) v okamžiku sklizení. Před samotnou sklizní je potřeba nastavit potřebné parametry, které ovlivní práci harvesteru.

Jedná se zejména o tyto parametry:

- Nastavení protokolu - ve kterém bude harvester sklízet zdroje. Celoplošné sklizeně WebArchivu probíhají v protokolech http a ftp.
- Velikost stahovaných souborů - nesmí překročit 100 MB.
- Maximální počet souborů sklizených z jedné webové stránky je 10 000.
- Hloubka zanoření – jedná se o to, kolik odkazů ve struktuře stránky harvester prohledá. V případě sklizení prováděných v rámci projektu WeArchiv je hloubka zanoření stanovena na 15 prokliků.

¹⁵⁶ viz kapitola 3.2.1 Legislativa

¹⁵⁷ *WebArchiv - archiv českého webu* [online]. Poslední aktualizace: 12.5.2009 [cit. 2009-05-12]. Dostupný z WWW: <<http://www.webarchiv.cz/vydavatele/>>.

¹⁵⁸ CELBOVÁ, Ludmila. Český web a povinný výtisk – jde to spolu dohromady? *Knihovna* [online]. 2008, roč. 19, č. 2, s. 59-75 [cit. 2009-05-12]. Dostupný z WWW: <<http://knihovna.nkp.cz/knihovna82/82005.htm>>. ISSN 1801-3252.

¹⁵⁹ Informace byly poskytnuty odpovědnou osobou z Národní knihovny ČR [cit. 2009-05-13].

¹⁶⁰ viz kapitola 3.2.2 Co a jak uchovávat

- Určení formátů stahovaných souborů - je závislé na možnostech harvesteru respektive na tom, jaké formáty podporuje. V případě potřeby je možné spektrum stahovaných formátů omezit.¹⁶¹

Tento proces sklizení je plně automatizovaný a nevyžaduje další intelektuální úsilí.¹⁶² Celoplošné sklizeně se realizují od roku 2001 a probíhají jednou ročně s výjimkou roku 2003. Poslední sklizeň byla provedena v listopadu 2008 a podařilo se stáhnout 3,9 TB dat.¹⁶³ Mezi nevýhody této metodiky sklizení patří omezení se pouze na národní doménu, mnoho vydavatelů však může publikovat též pod nadnárodními doménami (.com, .info, .eu, .org)¹⁶⁴. Rovněž frekvence sklizení není dostatečná vzhledem k tomu, jak často se některé weby aktualizují. Při množství dat, která jsou stáhnuta, není dost dobře možné provést kontrolu sklizeného obsahu. Některé ze sklizených webů tak nemusejí být plně funkční. U celoplošných sklizení se odhaduje, že až 40% sklizených dokumentů je neúplných či vykazuje určitý defekt.¹⁶⁵

Výběrové sklizeně vyžadují mnohem více individuálního přístupu než předchozí plošné. Výběr zdrojů k archivaci probíhá na základě okruhu kritérií, která určují které dokumenty archivovat a které ne. Samotný proces výběrové sklizeně se skládá z několika kroků:

- Vyhledání zdroje.
- Posouzení, zda zdroj splňuje požadovaná kritéria.
- Opatření licence pro zhotovení kopie a zpřístupnění.
- Stažení dokumentu (sklizení).
- Zpětná kontrola kvality zdroje¹⁶⁶

Celé spektrum zdrojů, v kterých je možné vyhledávat je pak ohraničeno těmito podmínkami:

- Všechny zdroje publikované na území České republiky.
- Veškeré dokumenty vydané v českém jazyce i mimo republiku.
- Všechny publikace od českých autorů (nezáleží na místě vydání).

¹⁶¹ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 23 s. ISBN 978-80-7050-562-5.

¹⁶² Tamtéž

¹⁶³ *WebArchív - archiv českého webu* [online]. Poslední aktualizace: 14.5.2009 [cit. 2009-05-14]. Dostupný z WWW: < <http://www.webarchiv.cz/celoplosne-sklizne/>>.

¹⁶⁴ viz kapitola **2.5.2 Domény**

¹⁶⁵ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 24 s. ISBN 978-80-7050-562-5.

¹⁶⁶ Tamtéž

- Díla, jejichž obsah se týká České republiky nebo českého národa (bez ohledu na místo vydání).¹⁶⁷

K vyhledání dalších nových zdrojů, které by mohly být uloženy do archivu, se využívají běžně dostupné vyhledávací utility¹⁶⁸ jako vyhledávače¹⁶⁹, **webové adresáře**¹⁷⁰ a různé portály¹⁷¹. Velice užitečným může v tomto ohledu být např. projekt **jednotné informační brány**^{172, 173} kde je možno vyhledávat všemožné zdroje podle určených kritérií. Kromě dokumentů z národní domény jsou vyhledávány i dokumenty z jiných domén. Z žánrových typů jsou preferovány především periodika, monografie, konferenční materiály, akademické práce a vládní dokumenty. Naopak stahovány nejsou počítačové hry, obsah intranetu¹⁷⁴, osobní blogy¹⁷⁵, databáze, vysílání rozhlasu a televize. Sklizené dokumenty jsou kontrolovány kurátory. Celá sbírka dokumentů WebArchivu vzešlých z výběrových sklizní je koncipována na základě metody **Konspekt**. Tato metoda byla vyvinuta v 70. letech skupinou amerických knihoven **Research Library Group**. Konspekt obsahuje 24 kategorií a sklizený dokument je přiřazen vždy do některé z nich. Každá z kategorií se dále člení do podkategorií podle desetinného třídění.

Jednotlivé kategorie jsou:

1. Antropologie, etnografie
2. Biologické vědy
3. Divadlo, film, tanec
4. Ekonomické vědy, obchod
5. Filozofie a náboženství
6. Fyzika a příbuzné vědy
7. Geografie. Geologie. Vědy o Zemi

¹⁶⁷ *WebArchiv - archiv českého webu* [online]. Poslední aktualizace: 14.5.2009 [cit. 2009-05-14]. Dostupný z WWW: <<http://www.webarchiv.cz/kriteria/>>.

¹⁶⁸ Utilita je program sloužící ke zjednodušení či zrychlení nějaké činnosti, která by šla jinak vykonat pouze s vynaložením většího úsilí.

¹⁶⁹ Například Google nebo český Seznam.

¹⁷⁰ Webové adresáře obsahují katalogy adres webových stránek, které jsou řazeny do tematických skupin a umožňují tak snadněji vyhledat webové stránky zaměřené na určitou oblast. Např. stránky o sportu, obchodu, vědě atd. Mezi webové adresáře patří Seznam, Google adresář a další.

¹⁷¹ Internetový portál funguje podobně jako webový adresář. Existují také oborové portály jako třeba portál veřejné správy - <http://portal.gov.cz/>

¹⁷² <http://www.jib.cz/>

¹⁷³ MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchování webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 244.

¹⁷⁴ Soukromá počítačová síť technologicky identická s Internetem.

¹⁷⁵ Blog je webová aplikace s příspěvky jednoho autora, jakýsi jeho zápisník. Blog tvoří také samostatnou webovou stránku.

8. Historie a pomocné historické vědy. Bibliografické studie
9. Hudba
10. Chemie. Krystalografie. Mineralogické vědy
11. Jazyk, lingvistika a literatura
12. Knihovnictví, informatika, všeobecné, referenční literatura
13. Matematika
14. Lékařství
15. Politické vědy (Politologie, politika, veřejná správa, vojenství)
16. Právo
17. Psychologie
18. Sociologie
19. Technika, technologie, inženýrství
20. Tělesná výchova a sport. Rekreační
21. Umění, architektura, muzeologie
22. Výchova a vzdělávání
23. Výpočetní technika
24. Zemědělství¹⁷⁶

Výhodou výběrových sklizní je funkčnost sklizených webů (zpětné ověření kompatibility) a rovněž libovolné nastavení frekvence sklizení. Mezi nevýhody patří větší personální nároky a s tím spojené ekonomické náklady. Projekt WebArchiv uskutečňuje ještě třetí způsob sklizení, kterým jsou tematické sbírky. Tvorba tematických sbírek je vlastně typem výběrových sklizní. Úkolem této sklizně je shromáždit informace k určité události, přičemž hlavním kritériem výběru zdroje je jeho relevance k události.¹⁷⁷ V současnosti obsahuje WebArchiv tyto tematické sbírky:

- Povodně (2002)
- Dalimilova kronika (2005)
- Vysočina (2005)
- Volby (2006)
- Nová budova Národní knihovny (2007)
- Praha olympijská (2007)

¹⁷⁶ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 31 s. ISBN 978-80-7050-562-5.

¹⁷⁷ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 28 s. ISBN 978-80-7050-562-5.

- Nová budova Národní technické knihovny (2007)
- Prezidentské volby (2008)¹⁷⁸

Národní knihovna používá ke sklizení webů softwarový nástroj **Heritrix**¹⁷⁹, který v roce 2005 nahradil dřívější NEDLIB harvester¹⁸⁰. Heritrix obsahuje dvě základní části: jádro a přípojný modul. Jádro slouží pro nastavení sklizně a pro správu běžících procesů. Moduly zajišťují jednotlivé kroky sklizně a mohou být navrženy k řešení určitých problémů, jako je duplicita stránek apod. Heritrix pracuje tak, že navštíví zadanou URL adresu¹⁸¹ a z té stáhne data, z těchto dat vybere další URL odkazy a proces opakuje, dokud nedojde nastaveného limitu hloubky zanoření. Tento harvester je využíván jak pro celoplošné tak pro výběrové sklizně a je dosud nejkvalitnějším nástrojem ke sklizení webu. Sklizené webové stránky jsou ukládány v datovém formátu **ARC**, který byl vyvinut speciálně pro účely archivace webu. Soubor ARC se skládá z metadatového záhlaví a řady URL odkazů. Samotné URL záznamy pak obsahují další **metadata**¹⁸², které obsahují informace jako: datum sklizení, URL adresu, IP adresu, typ serveru. Vedle metadat jsou tu uložena vlastní data webových stránek: soubory HTML, GIF, JPEG atd. Všechna data jsou spojována do balíků a ty následně procházejí kompresí. Velikost výsledných ARC souborů se pohybuje od 100 do 600 MB, ale může být i větší.¹⁸³

Jako formát metadat používá Národní knihovna **Dublin Core (DC)**^{184, 185}. Jedná se o soubor metadatových prvků, jehož vývoj započal v roce 1995 ve městě Dublin¹⁸⁶ v USA. Konference, kde se poprvé objevil pojem tzv. Dublinského jádra (Dublin Core), měla za úkol definovat univerzální systém popisu elektronických informačních zdrojů.¹⁸⁷ Dublin Core obsahuje 15 sémantických prvků:

1. Název (Title)
2. Tvůrce (Creator)

¹⁷⁸ *WebArchiv - archiv českého webu* [online]. Poslední aktualizace: 14.5.2009 [cit. 2009-05-14]. Dostupný z WWW: <http://www.webarchiv.cz/tematicke_sbirky/>.

¹⁷⁹ viz kapitola **3.1 Webové archivy**

¹⁸⁰ viz kapitola **3.2.2 Co a jak uchovávat**

¹⁸¹ viz kapitola **2.2 Služba WWW**

¹⁸² Metadata jsou data o datech a mají za úkol popsat určitý soubor informací a dat pro jejich snadnější vyhledávání. Pro představu by se metadata dala přirovnat ke katalogizačnímu lístku knihovní databáze.

¹⁸³ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 34-36 s. ISBN 978-80-7050-562-5.

¹⁸⁴ <http://dublincore.org/>

¹⁸⁵ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 36 s. ISBN 978-80-7050-562-5.

¹⁸⁶ Nejedná se o hlavní město Irské republiky.

¹⁸⁷ BARTOŠEK, Miroslav. Vyhledávání v Internetu a DUBLIN CORE. *Zpravodaj ÚVT MU*. ISSN 1212-0901, 1999, roč. IX, č. 4, s. 1-4.

3. Předmět (Subject)
4. Popis (Description)
5. Vydavatel (Publisher)
6. Příspěvatel (Contributor)
7. Datum (Date)
8. Typ (Type)
9. Formát (Format)
10. Identifikátor (Identifier)
11. Zdroj (Source)
12. Jazyk (Language)
13. Vztah (Relation)
14. Pokrytí (Coverage)
15. Práva (Rights)¹⁸⁸

Dublin Core metada lze vložit do záhlaví webové stránky a mohou být zapsané v jazyce HTML či XHTML¹⁸⁹, čímž se dá docílit lepší vyhledatelnosti stránky. Pro metada webů funguje speciálně navržená metoda zápisu **RDF**¹⁹⁰, pocházející od organizace **W3C**¹⁹¹. DC je možno rozšířit podle potřeby jednotlivých uživatelů o tzv. kvalifikátory, což jsou určité zpřesňující informace k libovolnému z prvků souboru DC metadat.¹⁹² Formátem Dublin Core se v ČR zabývá Knihovnicko-informační centrum Masarykovy university v Brně a provozuje oficiální český web o DC¹⁹³.¹⁹⁴ Na stránkách WebArchivu je pro uživatele k dispozici generátor metadat ve formátu DC¹⁹⁵. Na závěr nutno podotknout, že systém Dublin Core je vhodný jak pro odborné tak pro laické uživatele.¹⁹⁶

¹⁸⁸ *Dublin Core - Czech homepage* [online]. c1995-2005 , Poslední aktualizace: 20. listopad 2006 14:20:28 [cit. 2009-05-16]. Dostupný z WWW: <http://www.ics.muni.cz/dublin_core/elems.html>.

¹⁸⁹ viz kapitola **2.2 Služba WWW**

¹⁹⁰ Resource Description Framework

¹⁹¹ World Wide Web Consortium je mezinárodní konsorcium založené v roce 1994, které se zabývá vývojem a určováním nových standardů pro internetovou službu WWW. <http://www.w3.org/>

¹⁹² CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 37 s. ISBN 978-80-7050-562-5.

¹⁹³ http://www.ics.muni.cz/dublin_core/index.html

¹⁹⁴ *Dublin Core - Czech homepage* [online]. c1995-2005 , Poslední aktualizace: 20. listopad 2006 14:20:28 [cit. 2009-05-16]. Dostupný z WWW: <http://www.ics.muni.cz/dublin_core/index.html>.

¹⁹⁵ http://www.webarchiv.cz/generator/dc_generator.php

¹⁹⁶ *Dublin Core - Czech homepage* [online]. c1995-2005 , Poslední aktualizace: 20. listopad 2006 14:20:28 [cit. 2009-05-16]. Dostupný z WWW: <http://www.ics.muni.cz/dublin_core/charakteristika.html>.

Pro zpřístupnění sbírky na webu WebArchivu¹⁹⁷ je užíván nástroj **Wayback Machine**¹⁹⁸. Jeho nedostatkem je, že nepodporuje plnotextové (fulltextové) a jiné typy vyhledávání, což by mělo být v budoucnu vyřešeno¹⁹⁹, ale zatím hledanou webovou stránku nelze nalézt bez znalosti její URL adresy. Na webu jsou však k dispozici pouze zdroje, s jejichž vydavateli byla uzavřena smlouva o online zpřístupnění, anebo které jsou publikovány pod příslušnou CC licenci²⁰⁰. Zbytek archivovaných webů je pak přístupný z terminálu nacházejícího se v budově Národní knihovny v pražském Klementinu.

Význam projektu WebArchiv je nepochybný. V České Republice nefunguje dosud žádný jiný projekt zabývající se dlouhodobým uchováváním webů a tak je WebArchiv jediný, kdo může zachytit a zachycuje významnou část národního kulturního dědictví, jakým je webový obsah Internetu.

¹⁹⁷ <http://www.webarchiv.cz/vyhledej/>

¹⁹⁸ viz kapitola **3.1 Webové archivy**

¹⁹⁹ CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 40 s. ISBN 978-80-7050-562-5.

²⁰⁰ viz kapitola **3.2.1 Legislativa**

4 Webové stránky z pohledu spisové a archivní služby

Zatímco knihovny pohlíží na weby jako na publikace, v zájmu **archivů**²⁰¹ je uchovávat především **archiválie**. Stejně tak okruh webů, které by mohly být v popředí zájmu archivů, bude podstatně užší, než jak je tomu u knihoven. Tyto rozdílné pohledy obou institucí jsou dány zejména platnou legislativou, která určuje jejich činnost. Na začátek bychom si tedy měli říci, co archivní služba vůbec je, jaká je náplň jejich činností a podle čeho se řídí.

4.1 Úvod do archivnictví

Resort archivnictví spadá pod Ministerstvo vnitra ČR²⁰² a jeho podstata vyplývá ze **zákona č. 499/2004 Sb., o archivnictví a spisové službě a o změně některých zákonů (archivní zákon)**. Dále je činnost archivnictví upravena prováděcí vyhláškou k tomuto zákonu: **č. 645/2004 Sb., provádění ustanovení zákona o archivnictví a spisové službě**. Archivní zákon nám o archivnictví říká toto: „*Pro účely tohoto zákona se rozumí a) archivnictvím obor lidské činnosti zaměřený na péči o archiválie jako součásti národního kulturního dědictví a plnící funkce správní, informační, vědecké a kulturní,*“²⁰³. Archiválii pak definuje jako:

„záznam, který byl vzhledem k době vzniku, obsahu, původu, vnějším znakům a trvalé hodnotě dané politickým, hospodářským, právním, historickým, kulturním, vědeckým nebo informačním významem vybrán ve veřejném zájmu k trvalému uchování a byl vzat do evidence archiválií; archiváliemi jsou i pečeti, razítka a jiné hmotné předměty související s archivním fondem či s archivní sbírkou, které byly vzhledem k době vzniku, obsahu, původu, vnějším znakům a trvalé hodnotě dané politickým, hospodářským, právním, historickým, kulturním, vědeckým nebo informačním významem vybrány a vzaty do evidence,“²⁰⁴

Z výše uvedených citací lze vyčíst, že funkce archivu v principu není nepodobná funkci knihovny. Obě dvě se starají o určitou část kulturního dědictví. Avšak zatímco knihovnám jsou dodávány povinné výtisky publikací téměř všeho druhu, archivy provádějí

²⁰¹ Archiv je zařízení, sloužící k ukládání archiválií a péči o ně.

²⁰² <http://www.mvcr.cz/>

²⁰³ Zákon č. 499/2004 Sb., o archivnictví a spisové službě a o změně některých zákonů. § 2.

²⁰⁴ Tamtéž

tzv. **výběr archiválií**, kdy je posuzována hodnota určitých dokumentů a některé z nich jsou prohlášeny za archiválie a následně uloženy do archivu a zaevidovány.

Dokumentem se rozumí: „každý písemný, obrazový, zvukový, elektronický nebo jiný záznam, ať již v podobě analogové či digitální, který vznikl z činnosti původce,“²⁰⁵. Původce je právě ten, kdo má za povinnost dokumenty uchovávat a umožnit z nich výběr archiválií. Dle zákona existují dvě skupiny původců: původci veřejnoprávní a soukromoprávní. Pro naši práci mají význam předně původci veřejnoprávní, kam se řadí:

- organizační složky státu
- státní příspěvkové organizace
- státní podniky
- územní samosprávné celky
- organizační složky a právnické osoby založené nebo zřízené územními samosprávnými celky, pokud vykonávají veřejnou správu nebo zaměstnávají více než 25 zaměstnanců
- školy a vysoké školy
- právnické osoby zřízené zákonem
- zdravotnická zařízení²⁰⁶

Tito jmenovaní původci mají za povinnost vykonávat **spisovou službu**, kterou se v této kapitole budeme rovněž zabývat. Spisová služba resp. její výkon označuje:

*„zajištění odborné správy dokumentů došlých a vzešlých z činnosti původce, popřípadě z činnosti jeho právních předchůdců, zahrnující jejich řádný příjem, evidenci, rozdělování, oběh, vyřizování, vyhotovování, podepisování, odesílání, ukládání a vyřazování ve skartačním řízení, a to včetně kontroly těchto činností,“*²⁰⁷

Spisová služba je dále řízena **vyhláškou č. 646/2004 Sb., o podrobnostech výkonu spisové služby**. Výběr archiválií u původců, kteří vykonávají spisovou službu, se provádí v tzv. **skartačním řízení**, což je postup, při kterém jsou vyřazovány dokumenty nadále nepotřebné pro práci původce a při kterém se provádí výběr archiválií příslušným archivem. Každý dokument uložený do **spisovny**²⁰⁸ je mimo jiné opatřen **skartační lhůtou** a **znakem**. Skartační lhůta je doba, po kterou je dokument uložen v organizaci. Skartační znak vyjadřuje

²⁰⁵ Zákon č. 499/2004 Sb., o archivnictví a spisové službě a o změně některých zákonů. § 2, d).

²⁰⁶ Zákon č. 499/2004 Sb., o archivnictví a spisové službě a o změně některých zákonů. § 3, (1).

²⁰⁷ Zákon č. 499/2004 Sb., o archivnictví a spisové službě a o změně některých zákonů. § 2, k).

²⁰⁸ Spisovna je místo vyhrazené k uložení a vyhledávání dokumentů. Dokumenty se do spisovny ukládají zpravidla ihned po svém vyřízení.

hodnotu dokumentu podle obsahu a označuje způsob, jakým se s dokumentem naloží ve skartačním řízení:

- znak **A** – dokument bude vybrán k uložení do archivu
- znak **S** – dokument bude ve skartačním řízení navržen ke zničení
- znak **V** – dokument bude posouzen ve skartačním řízení a označen jako A nebo S

Skartační znaky a lhůty jsou uvedeny ve **skartačním plánu**, který je součástí **spisového a skartačního řádu**. Tento řád upravuje vedení spisové služby a je povinen vydat ho každý veřejnoprávní původce. Součástí řádu je také **spisový plán**, který obsahuje seznam **spisových znaků** jednotlivých věcných skupin dokumentů. Podle těchto znaků jsou dokumenty ukládány ve spisovně.

V současné době je již připravena novela archivního zákona s navrhovanou platností od 1.7. 2009.²⁰⁹ Tato novela reaguje především na zvýšení nárůstu elektronických dokumentů ve veřejné správě a justici. Celková elektronizace veřejné správy souvisí zejména s projektem **eGovernment** a tzv. **zákonem o eGovernmentu**²¹⁰, který nově zavádí komunikaci mezi úřady či s právníckými nebo fyzickými osobami prostřednictvím datových schránek.

Celý proces zavádění jednotlivých a navzájem propojených služeb eGovernmentu má za úkol zefektivnění služeb a odlehčení byrokracie orgánů veřejné správy. V souvislosti s datovými schránkami lze očekávat další navýšení počtu dokumentů, jež budou existovat pouze v elektronické podobě. Tento jev ovlivní zákonitě i požadavky na výkon spisové a archivní služby. Nová legislativa proto bude zahrnovat i novelu vyhlášky č. 646/2004 Sb., o podrobnostech výkonu spisové služby, která bude podrobněji upravovat požadavky na vedení elektronické spisové služby. Základní princip legislativní úpravy spočívá v preferenci elektronické spisové služby před spisovou službou vedenou v listinné podobě.²¹¹

4.2 Archivy a dlouhodobé uchovávání webových stránek

Archivy jsou v oblasti dlouhodobého uchovávání webů úplnými nováčky a je nutno podotknout, že k žádné řízené archivaci webů v archivech zatím nedochází. Rovněž v oboru archivní služby dosud nevznikly žádné projekty zabývající problematikou dlouhodobého

²⁰⁹ MALÍKOVÁ, Jana. *Novela zákona o archivnictví a spisové službě* [online]. c2008 [cit. 2009-06-01]. Dostupný z WWW: <<http://www.mvcr.cz/clanek/novela-zakona-o-archivnictvi-a-spisove-sluzbe.aspx>>.

²¹⁰ *Zákon č. 300/2008 Sb., o elektronických úkonech a autorizované konverzi dokumentů.*

²¹¹ MALÍKOVÁ, Jana. *Novela zákona o archivnictví a spisové službě* [online]. c2008 [cit. 2009-06-01]. Dostupný z WWW: <<http://www.mvcr.cz/clanek/novela-zakona-o-archivnictvi-a-spisove-sluzbe.aspx>>.

uchovávání webových stránek, až na několik odborných článků a diskuzí na toto téma. Dosud ještě nebyla vyřešena hlavní otázka, zda mají být weby uchovávány v archivech, respektive zda mají být některé webové stránky či jejich části považovány za archiválie (viz níže).

Názor na tuto problematiku u samotných archivářů je různý²¹² a proto dosud nepadlo jasné stanovisko, jak se k webům stavět, což brání započetí potencionálního projektu, který by se zabýval dlouhodobým uchováváním webů v archivech a mohl řešit dílčí problematiku s tím spojenou. Navíc dosud není v provozu a plně dořešena archivace digitálních dokumentů obecně.

Nicméně my budeme dále uvažovat v naší práci o webových stránkách jako o možných archiváliích. Existují k tomu pádné důvody podobně jako u knihoven a nelze je přehlížet. Podle článku Oskara Macka a Michala Wannera²¹³ jsou webové stránky:

*„stejným nástrojem ke komunikaci s veřejností jako jakékoli jiné médium. Odráží tedy nejen jeho názory a přístup k problémům, ale i komunikační strategii, reflexi dobového vkusu a nálad ve veřejnosti. Weby jsou obrazem toho, jak příslušný úřad, organizace či jiný subjekt viděly samy sebe a svou roli ve společnosti.“*²¹⁴

Posouzení hodnoty webové stránky bude záviset na archiváři, který bude provádět výběr archiválií u konkrétního původce disponujícího vlastním webem. Hlavním kritériem při výběru bude nejspíše trvalá hodnota obsahu webu, jako je tomu u jiných dokumentů. Další pomůckou může být archiváři jeho vlastní zkušenost a praxe nebo i případná směrnice, vydaná za tímto účelem. Z posuzování hodnoty webových stránek nemůžeme ale vyloučit ani původce. Z důvodu častých změn obsahu webů²¹⁵ se může stát, že jeho vyhodnocení bude záviset primárně na původci.²¹⁶

Obecně lze však předpokládat, že počet webů vybraných k archivaci nebude nijak závratný, jako je tomu u knihoven. Na rozdíl od knihoven, které se snaží uchovat všechn web, ať už světový nebo národní, budou archivy vybírat pouze z webů původců a uchovány budou pouze ty, jež splní kritéria výběru. Navíc nemusí být vybrán k archivaci celý web, ale pouze jeho určitá část. Předně můžeme z archivace vyloučit např. ten obsah webu, který má svou obdobu v papírové, či jiné digitální podobě. Navzdory tomu, že převážná část webů nebude pravděpodobně shledána trvalé hodnoty a nebude tedy účelné je dlouhodobě

²¹² viz kapitola **4.2.1 Průzkum**

²¹³ viz kapitola **1 Úvod**

²¹⁴ MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchovávání webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 249.

²¹⁵ viz kapitola **4.2.2.1 Webové stránky jako dokument**

²¹⁶ MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchovávání webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 250.

uchovávat, přesto se webové stránky původců stále vyvíjí a je potřeba je podrobit bližšímu prozkoumání z hlediska jejich hodnoty.

Původci často využívají svůj web jako jakousi nástěnku, kde jsou zveřejňovány právě dokumenty a informace existující i v jiné podobě, přenášejí na web některé úkony, např. komunikaci s občany prostřednictvím formulářů apod. Zajímavý je i případ britského Národního archivu, jehož stránky poskytují přístup k množství naskenovaných archiválií, čímž se archiv snaží řešit nedostatek místa a celkové přetížení badatelny.²¹⁷ K tomu jak původci i archivy pohlíží na weby a na problematiku jejich archivace, byl za tímto účelem proveden průzkum.

4.2.1 Průzkum

Cílem průzkumu bylo pokusit se zjistit postoje jednotlivých archivů a původců a jejich názor na danou problematiku. Z výsledků průzkumu pak můžeme v této práci částečně vycházet. Samotný průzkum byl proveden pomocí dotazníku, který byl prostřednictvím elektronické pošty rozeslán vybraným archivům a původcům. Pro každou z těchto dvou skupin byl vytvořen odlišný dotazník. Z archivů byly osloveny všechny oblastní, zemské, okresní a také Národní archiv, dohromady tedy 80 archivů. Z poměrně širokého okruhu veřejnoprávních původců byl dotazník rozeslán všem krajským úřadům, magistrátům okresních měst, některým státním podnikům a státním příspěvkovým organizacím. Celkem bylo osloveno 111 původců. Celé vyhodnocení průzkumu je anonymní a neobjeví se zde název žádné konkrétní instituce, ať už se průzkumu zúčastnila či nikoliv. Odpověď na uvedené otázky zaslalo 49 archivů z celkového počtu 80 dotázaných a 39 ze 111 původců, což lze pro náš účel považovat za dostatečné, abychom mohli vyvodit nějaké závěry. Co se samotných odpovědí týče, dotazovaní odpovídali povětšinou stručně ano nebo ne. Některé z odpovědí však byly i velmi přínosné. Jednotlivé dotazy byly voleny stručně a jasně.

Dotazník pro archivy se skládal z těchto pěti otázek:

1) *Myslíte si, že weby některých původců mají potenciál stát se archiváliemi?*

Otázka číslo 1 měla za úkol prozkoumat názor na hodnotu webových stránek z hlediska archivního, což je jeden ze základních smyslů této práce, a dlouhodobého uchovávání webů vůbec. Na tuto otázku odpovědělo 34 archivů kladně, tedy okolo 70%

²¹⁷ MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchovávání webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 250.

zúčastněných. Většina si tedy myslí, že některé webové stránky nebo jejich části by mohly být archivovány, či minimálně stojí za to se jimi zabývat.

2) *Zabývali jste se někdy už problematikou dlouhodobého uchovávání webových stránek jako archivních dokumentů?*

Druhá otázka je spíše doplňující k té první a zjišťovala povědomí i zájem archivářů o danou problematiku. Zde již ohlas nebyl takový a pouze 12 respondentů odpovědělo ano. Část z nich přitom uvedla, že se uchováváním webů zabývala na konferenci **České archivní společnosti**²¹⁸.

3) *Oslovil vás někdy některý původce ohledně návrhu webových stránek do skartačního řízení či možnosti uložení stránek do archivu?*

Doposud žádný z archivů se nesetkal s návrhem webových stránek svého původce do skartačního řízení ani o uložení do archivu. Taková je jednoznačná odpověď na otázku číslo tři.

4) *Je ve vašich možnostech zabezpečit uložení a následné zpřístupnění webu badatelům, pokud by prošel skartačním řízením?*

Možností technicky uložit a zabezpečit web disponuje pouze 6 archivů a mnoho z nich dodává, že pouze za určitých podmínek a spíše ojediněle.

5) *Uchováváte starší verze vašeho webu?*

Poslední pátý dotaz se zajímal o webové stránky archivů a o to zdali mají zájem o jejich starší verze a uchovávají je. Takto činí pouze 5 archivů. Nutno podotknout, že 11 dotázaných si nezajišťuje své stránky samo, ale o jejich web se stará příslušný oblastní či zemský archiv. Další z archivů uvedly, že žádná starší verze jejich stránek zatím neexistuje.

Pokud bychom tedy měli vyvodit z výše uvedeného průzkumu nějaké výsledky, můžeme konstatovat, že většina archivů si připouští možnost, že by se webová stránka mohla stát archiválií. Navzdory tomu zájem, či informovanost o této problematice nejsou nijak velké. Zde je ale také potřeba si uvědomit, že průzkum oslovoval jednotlivé archivy jako instituce a ne jednotlivé archiváře respektive odbornou archivní veřejnost. Drtivá většina archivů nemá vybavení, prostředky a ani kompetence jednak na vlastní výzkum v oblasti dlouhodobého uchovávání webů a dále na uložení stránek v archivu, což souvisí s otázkou číslo čtyři. Nelze předpokládat, že by webové stránky původců byly ukládány v archivu, který

²¹⁸ <http://www.cesarch.cz/>

spravuje fond²¹⁹ onoho původce. Pro uchovávání dokumentů v digitální podobě se v současnosti pracuje na projektu národního digitálního archivu²²⁰, který se pravděpodobně bude také zabývat archivací webů ve spolupráci s Odborem archivní správy Ministerstva vnitra.

Jednotlivé otázky pro původce:

1) *Myslíte si, že informační hodnota vašeho webu je trvalá či je pouze poplatná době, kdy jsou tyto informace aktuální?*

První z otázek se zabývá opět hodnotou webu, zda je pomíjivá či je trvalého charakteru. Z odpovědí vyplynulo, že 29 původců pokládá svůj web za trvalý, avšak téměř všichni uvedli, že trvalé hodnoty jsou pouze některé jeho části.

2) *Navrhli by jste vaše webové stránky jako jeden z typů el. dokumentů do skartačního řízení pro pozdější uložení do archivu jako archiválie?*

Ohledně možnosti navrhnout svoje webové stránky do skartačního řízení se vyjádřilo kladně 15 institucí. Tato otázka byla myšlena spíše teoreticky, protože v současné době není účast webů ve skartačních řízeních příliš reálná, z důvodů chybějících postupů neboli know how a také legislativní podpory.

3) *Vztahují se na vaše webové stránky nějaké vnitřní normy?*

Otázka číslo tři zjišťovala, jestli si původce ošetřuje svoje stránky nějakými předpisy či normami, z čehož se dá částečně vyčíst, jak moc je považuje za důležité ať už z reprezentativního nebo jiného hlediska. Vnitřní normu pro svůj web má vydáno 26 původců a většinou se týkají způsobu aktualizace stránek, dále způsobu publikování na nich.

4) *Jak často se v průměru vaše stránky aktualizují?*

Na další otázku 27 dotázaných odpovědělo, že jejich weby se aktualizují denně nebo dokonce několikrát za den, u ostatních je to pak v průměru týdně až měsíčně.

5) *Používáte služeb placeného webhostingu?*

Co se týče dotazu na používání webhostingu, ten měl za úkol, zjistit kolik původců si provozuje stránky samo a tím pádem k nim má i neomezený přístup. Výsledek je pouze 17 původců, kteří využívají služeb webhostingu.

6) *Uchováváte starší verze vašeho webu?*

Poslední dotaz je stejný jako u archivů a 17 respondentů odpovědělo ano.

²¹⁹ Archivní fond je soubor archiválií, který vznikl výběrem z dokumentů vzniklých z činnosti určitého původce (původců), který obvykle odráží organizaci, funkce a činnost původce a velmi často i strukturu původního uspořádání písemností u původce.

²²⁰ viz kapitola 4.3 Webové stránky v archivní péči

Cílem průzkumu u původců bylo zjistit, co si o hodnotě svých webů myslí či jaká je situace jejich webů v současnosti. Tento průzkum byl proveden spíše ze zajímavosti, protože převážná část původců zřejmě nebude patřit mezi iniciátory dlouhodobého uchovávání webů, které by pro ně znamenalo další povinnosti, případně náklady s tím spojené. Rovněž odpovědi na některé otázky z dotazníku musejí být brány s rezervou, vzhledem k tomu, že za určitou instituci mohl odpovídat zaměstnanec nepříliš seznámený s oblastí archivnictví, což se týká především prvních dvou otázek, které nemusely být vždy úplně pochopeny. Na druhou stranu byly zkoumány i některé aspekty, jež by mohly hrát při řešení archivace webů důležitou roli. Význam dotazů okolo webhostingu, aktualizace a uchovávání stránek bude zřetelnější dále, až budeme narážet na jednotlivá úskalí spojená s dlouhodobým uchováváním webových stránek. Na závěr si zde dovoluji uveřejnit dva zajímavé názory od různých původců:

„Všechny dokumenty uveřejněné na webu, které nemají vysloveně pomíjivý charakter, jsou výsledkem nějakého procesu, nějaké akce nebo úředního úkonu, který je pojímán z hlediska archivního zákona podle svého charakteru. Na webu už je jen zobrazení - vnější prezentace - zveřejnění těchto dokumentů.“

„Z archivního hlediska by jistě bylo zajímavé mít možnost zpětně vidět webové stránky, které jednou budou nepopíratelně vypovídat o společnosti v určité době, bohužel v současné podobě archivnictví, tak jak ji všichni známe, je tato vize pro nás téměř nepředstavitelná.“

4.2.2 Webové stránky v kontextu spisové služby

Pokud budeme uvažovat o webové stránce jako o potenciální archiválii, můžeme se na weby původců dívat jako na jeden z typů elektronických dokumentů, které má původce ve své registratuře²²¹ a jako takové je potřeba zohlednit je v systému spisové služby.

V současné době je již na spadnutí novela archivního zákona²²², která upravuje i požadavky na správu digitálních dokumentů a zavádí tzv. **Národní standard pro vedení spisové služby v elektronické podobě**.²²³ Ačkoliv není v době psaní této práce k dispozici konečná verze zmiňované novely, lze s největší pravděpodobností předpokládat, že v tomto směru se oficiální verze nezmění. Můžeme se tedy pokusit začlenit webové stránky do okruhu elektronických dokumentů celkově, se zřetelem na jejich specifika. Bohužel nebude možné detailněji upřesnit některé věci z již zmiňovaného důvodu absence novely archivního zákona.

²²¹ Za registraturu můžeme v tomto případě považovat část agendy, kterou původce stále využívá ve svých činnostech.

²²² viz kapitola 4.1 Úvod do archivnictví

²²³ Pracovní návrh textu vyhlášky – náhrady vyhlášky č.646/2004 Sb., o podrobnostech výkonu spisové služby – varianta 8. 10. 2008.

Na úvod si však objasníme několik obecných pojmů z hlediska Národního standardu pro vedení spisové služby v elektronické podobě:

- **Analogový dokument** - Dokument obsahující informace ve spojitě, nikoliv digitální formě, nelze jej proto duplikovat bez ztráty kvality (např. zvukový záznam na gramofonové desce nebo magnetofonové pásce, filmový záznam, videozáznam systému VHS apod.).²²⁴
- **Digitální dokument** – Dokument vytvořený způsobem zpracování entity představované numerickým řetězcem tvořeným čísly „1“ a „0“(proud bitů) interpretovatelný pomocí výpočetní techniky. **Pojem „elektronický“ se pro účely národního standardu užívá obdobně.**²²⁵
- **Elektronické dokumenty** jsou dokumenty pořízené prostřednictvím výpočetní techniky, jejichž čitelnost je na výpočetní technice závislá.
- Elektronické dokumenty jsou záznamy (data) na datových nosičích nebo v permanentní paměti počítače (harddisk).²²⁶ Za elektronický dokument nelze považovat v případě Národního standardu analogový dokument, protože jako takový nemůže být spravován prostředky výpočetní techniky.²²⁷
- **Elektronický systém spisové služby ERMS** (Electronic Record Management System) - informační systém určený ke správě dokumentů ve smyslu ustanovení § 2 písm. k) zákona č. 499/2004 Sb., o archivnictví, spisové službě a o změně některých zákonů.²²⁸

Národní standard pro vedení spisové služby v elektronické podobě je položen na základech mezinárodního standardu **MoReq2**²²⁹. Zkratka MoReq znamená v českém překladu „Modelové požadavky pro správu elektronických dokumentů“ a je souhrnem specifikačních požadavků na vedení elektronické spisové služby jak veřejného, tak soukromého sektoru. Jedná se v podstatě o přehled minimálních požadavků na systémy elektronické správy dokumentů, jež je možné aplikovat v různém prostředí s různými podmínkami. MoReq

²²⁴ *Vydavatelství VŠCHT Praha* [online]. 2000 [cit. 2009-06-21]. Dostupný z WWW: <http://vydavatelstvi.vscht.cz/knihy/uid_es-005/hesla/trida.OAAA.html>.

²²⁵ *Pracovní verze národního standardu pro spisovou službu vykonávanou v elektronické podobě.* [online]. Ministerstvo vnitra České republiky [cit. 2009-06-21]. Dostupný z WWW: <www.mvcr.cz/soubor/moreq-pracovni-pdf.aspx>.

²²⁶ KUNT, Miroslav. *Elektronické dokumenty z hlediska požadavku zákona č. 499/2004 Sb. a potřeb jejich dlouhodobého ukládání* [online]. 2005, 29.5. 2005 [cit. 2009-06-21]. Dostupný z WWW: <http://www.nacr.cz/Z-Files/elektronicke_doku.pdf>.

²²⁷ *Pracovní verze národního standardu pro spisovou službu vykonávanou v elektronické podobě.* [online]. Ministerstvo vnitra České republiky [cit. 2009-06-21]. Dostupný z WWW: <www.mvcr.cz/soubor/moreq-pracovni-pdf.aspx>.

²²⁸ Tamtéž

²²⁹ Model Requirements Specification for the Management of Electronic Records. <http://www.moreq2.eu/>

nespecifikuje žádný konkrétní model ERMS a je určen k obecnému použití, což je jeden z jeho hlavních cílů. Vznik standardu je úzce spojen s fórem DLM²³⁰, což je mezinárodní organizace, která se zabývá všemi aspekty nakládání s dokumenty v digitální podobě. V současné době existuje nástupce MoRequ, kterým je již zmiňovaný MoReq2.²³¹

4.2.2.1 Webové stránky jako dokument

Jak už zde bylo naznačeno, webovou stránku můžeme chápat jako dokument z hlediska archivního. Protože se jedná o digitální dokument, měl by jako takový podléhat nárokům standardu pro elektronickou spisovou službu²³². Avšak webové stránky mají narozdíl od ostatních digitálních dokumentů svá specifika, která si budou žádat nové úpravy v legislativě, standardech či nové metodické pokyny apod.

Předně můžeme web označit jako dokument vzešlý z činnosti původce. Tento dokument „nežije“ klasickým způsobem jako je tomu u ostatních dokumentů, které dříve či později opustí agendu původce a putují buď do archivu, nebo jsou skartovány. Lze předpokládat, že webová stránka bude přítomna v registratuře po celou dobu existence původce. Do skartačního řízení a případně uložena do archivu může být tedy pouze její kopie. Další odlišností webu je jeho proměnlivost. Zde narážíme na výsledky našeho průzkumu, kdy více než dvě třetiny původců uvedly, že jejich web se aktualizuje denně nebo i častěji. Stránky často obsahují množství informací, které se neustále mění. Rovněž se web proměňuje v důsledku dynamických technologií²³³, které ho oživují. Uchovávání webu po každé změně by zřejmě bylo pro původce velice náročné a ani není účelné zachovávat veškeré informace z webu či jednotlivé stránky vytvořené činností skriptů²³⁴. Přijatelnější variantou se zdá být určení těch částí webu, jejichž hodnota je trvalá a ty pak při každé změně uchovávat. Posouzení trvalé hodnoty stránek se může jevit jako dosti nelehký úkol, který bude spočívat na bedrech příslušného archiváře. Archivář by měl mít navíc alespoň základní znalosti o fungování webu a webových technologiích, aby byl schopen orientovat se i v datové skladbě stránek. Například určit, zda se zachová pouze samotný HTML kód nebo třeba i skripty pro

²³⁰ <http://dlimforum.typepad.com/>

²³¹ MoReq2 - spolupráce českých archivářů na tvorbě standardu pro elektronickou spisovou službu. *Ministerstvo vnitra ČR* [online]. 2007 [cit. 2009-06-18]. Dostupný z WWW: <<http://www.mvcr.cz/clanek/moreq2-spoluprace-ceskych-archivaru-na-tvorbe-standardu-pro-elektronickou-spisovou-sluzbu.aspx>>.

²³² viz předchozí kapitola

²³³ viz kapitola **2.4 Webové technologie**

²³⁴ viz kapitola **2.4.2 Skripty**

zachování původního významu, pro který bude stránka archivována apod. Web může navíc obsahovat i obrázky a video soubory, které mohou být rovněž vybrány k archivaci.

Problém periodicity ukládání stránek souvisí i se stanovením skartačních znaků a lhůt. Pokud budeme teoreticky uvažovat o vzniku povinnosti původců navrhovat webové stránky do skartačního řízení, např. v důsledku legislativního zahrnutí webu mezi dokumenty, mohla by situace proběhnout takto:

Původce disponující vlastním webem, musí své stránky zaevidovat, zařadit do spisového a skartačního plánu, což znamená opatřit je spisový znakem a skartačním znakem a lhůtou. Protože by weby původců procházely svým prvním skartačním řízením, nabízí se zde skartační znak V. Větším úskalím bude stanovení skartační lhůty, která by měla být pokud možno co nejkratší. Otázkou je také, zda vyvolávat samostatná skartační řízení zaměřená pouze na weby. Možností je zavedení speciální skartační lhůty 0 určené přímo pro weby, případně i jiné digitální dokumenty. Původce by buď mohl vyvolat řízení téměř ihned po zapsání webu do evidence v případě samostatných skartačních řízení, nebo navrhnout web v nejbližším možném skartačním řízení.

Při prvním skartačním řízení by tedy došlo k posouzení trvalé hodnoty stránek, případně k vybrání těch částí webu vhodných k archivaci. Výběr archiválií u webů se bude provádět z aktuální verze webu a archivář bude zkoumat výstup stránky v prohlížeči, v podstatě jako když by si stránku prohlížel kdekoliv na Internetu. Vybraný web nebo jeho částí je potřeba dále zaevidovat s přesným určením toho, co bylo ze stránek určeno k uložení do archivu. Teoreticky je možný vznik samostatné dokumentace (evidence) webových stránek či minimálně uzpůsobení té stávající pro tyto účely. Původce pak předá vybraný web do archivu za předem určených podmínek. Pokud dojde ke změně ve vybraných částech webové stránky, zhotoví se jejich kopie a ty budou evidovány jako samostatné dokumenty. Tyto nově vzniklé dokumenty se opatří skartačním znakem A a uloží do spisovny. Nyní se opět dostáváme k problému definice skartační lhůty. Dalším možným řešením by bylo předávat weby, které prošly svým prvním skartačním řízením, rovnou do archivu bez celého procesu výběru archiválií ve skartačních řízeních, které by se nejspíše zmnožily²³⁵, což není příliš žádoucí ani pro původce ani pro archiváře. Pokud původce přejde na nový web (novou verzi) celý naznačený postup se zopakuje.

V oblasti dlouhodobého uchovávání webových stránek se v podstatě jeví dva základní

²³⁵ MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchovávání webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 250.

principy: postavit web do systému spisové služby nebo mimo něj. Obě dvě možnosti budou vyžadovat nejen změny v archivní legislativě, ale i nové směrnice a metodické pokyny jak pro původce, jenž bude zahrnovat weby do svých spisových norem, tak i pro archiváře, který je bude posuzovat. Další problematika, kterou bude potřeba vzít v úvahu, jsou technické možnosti původce, s čímž je spojena otázka webhostingu. U původců, kteří si spravují své stránky sami, by v celku neměl být větší problém. Rovněž původci provozující elektronickou spisovou službu disponují většinou dostatečným technickým i odborným zázemím. Avšak někteří původci využívající webhosting nemají přímý přístup ke svým stránkám, navíc by ani často nebyli schopni zabezpečit kvalitní technickou a odbornou podporu při úkonech spojených se správou webových stránek jako dokumentu. Tito původci by pak byli nuceni na vlastní náklady pověřit těmito úkony nějaký specializovaný subjekt, což můžeme označit za formu tzv. outsourcingu.

Outsourcing znamená vyčlenění různých vedlejších a podpůrných činností určité instituce, které pak na základě smlouvy provádí jiná společnost. V dnešní době je outsourcing využíván právě v oblasti informačních technologií a nabízí tak vhodnou možnost pro původce, kteří by jinak museli budovat vlastní odborné zázemí pro tyto činnosti. I tak ale přinese archivace webů nároky na jejich původce ať už finanční nebo v podobě zátěže dalšími povinnostmi a úkony.

4.3 Webové stránky v archivní péči

Nyní se podíváme trochu blíže na projekt Národního digitálního archivu, který by se vší pravděpodobností mohl stát řešitelem dlouhodobého uchování webů, a v něm by také byly weby archivovány.

Projekt Národního digitálního archivu byl deklarován usnesením vlády č. 11/2004. Je budován jako součást Národního archivu, který je garantem řešení tohoto projektu.²³⁶ Digitální archiv je také zakotven novele archivního zákona a jeho základními činnostmi jsou:

a) ukládá archiválie v digitální podobě vybrané k trvalému uložení veřejnými archivy s výjimkou archiválií vybraných bezpečnostními archivy zpravodajských služeb České republiky, udržuje jejich obsah a zajišťuje jejich trvalou čitelnost,

²³⁶ BERNAS, Jiří, KUNT, Miroslav. Národní digitální archiv : Představení projektu. *Internet ve státní správě a samosprávě* [online]. 2009 [cit. 2009-06-22]. Dostupný z WWW: <www.issc.cz/archiv/2008/download/prezentace/bernas_nap.ppt>.

b) zpracovává a zpřístupňuje archiválie v digitální podobě převzaté od původců uvedených v odstavci 1 písm. a) a c) a archiválie v digitální podobě získané podle odstavce 1 písm. d),

c) plní metodickou a poradenskou funkci v oblasti předarchivní péče o dokumenty v digitální podobě, digitalizace dokumentů a archiválií.²³⁷

Předpokládá se, že archiv by mohl vzniknout k 31. 12. 2011.²³⁸ Existovat by měla dvě fyzická pracoviště digitálního archivu. Na obou těchto pracovištích budou uloženy stejné dokumenty, což bude fungovat jako záloha. Toto řešení má za cíl předejít poškození či úplnému zničení uložených dat v důsledku živelné nebo jiné pohromy. Jedno úložiště dat se bude nacházet při Národním archivu a druhé v Jihočeském kraji. Jako fyzikální metody uložení dat byly zvoleny disková pole a UDO disky.²³⁹

Systém přebírání a správy digitálních dokumentů je primárně postaven na provázanosti s elektronickým systémem spisové služby (ERMS) a s Národním standardem. Archiv má určené formáty, které bude akceptovat.²⁴⁰ Důraz je kladen zejména na otevřenost, rozšířenost, transparentnost a sebedokumentaci formátů. Hlavním cílem je zajistit čitelnost dokumentu i po více jak 50 letech.²⁴¹ Při dnešním tempu vývoje v oblasti informačních technologií není jisté, zda některé dokumenty budou stále čitelné i novými technologiemi. I přes vhodný výběr formátů se archiv zřejmě do budoucna nevyhne migraci²⁴², která je spolu s emulací jedním z principů zachování čitelnosti dokumentu. Emulace je vytváření prostředí, ve kterém je aplikace spustitelná, forma dokumentu se přitom vůbec nemění.²⁴³

Určené formáty budou závazné i pro původce, který musí dokumenty předávat do archivu v předepsaném formátu a opatřené metadaty. Elektronický systém původce by měl zvládat tvorbu metadat jednotlivých dokumentů, protože dodatečné doplnění se jeví jako příliš pracné. Metadata, což jsou vlastně informace o dokumentu (jeho popis) se samozřejmě vztahují i na webové stránky a již jsme zde o nich mluvili²⁴⁴ v souvislosti s Národní

²³⁷ BERNAS, Jiří. Národní digitální archiv. *Co po nás zbude* [online]. 2009 [cit. 2009-06-22]. Dostupný z WWW: <<http://www.cnz.cz/ke-stazeni/2009/seminar-ds-mza-brno/prezentace/NDA-20090224.pdf>>.

²³⁸ BERNAS, Jiří, KUNT, Miroslav. Národní digitální archiv : Představení projektu. *Internet ve státní správě a samosprávě* [online]. 2009 [cit. 2009-06-22]. Dostupný z WWW: <www.issc.cz/archiv/2008/download/prezentace/bernas_nap.ppt>.

²³⁹ BERNAS, Jiří. Národní digitální archiv. *Co po nás zbude* [online]. 2009 [cit. 2009-06-22]. Dostupný z WWW: <<http://www.cnz.cz/ke-stazeni/2009/seminar-ds-mza-brno/prezentace/NDA-20090224.pdf>>.

²⁴⁰ Usnesení vlády České republiky č. 1338/2008, kterým se stanoví výstupní datové formáty statických dokumentů textové, obrazové a kombinované povahy v digitální podobě.

²⁴¹ BERNAS, Jiří, KUNT, Miroslav. Národní digitální archiv: Představení projektu. *Internet ve státní správě a samosprávě* [online]. 2009 [cit. 2009-06-22]. Dostupný z WWW: <www.issc.cz/archiv/2008/download/prezentace/bernas_nap.ppt>.

²⁴² viz kapitola 3.2.2 Co a jak uchovávat

²⁴³ BERNAS, Jiří. Principy digitální archivace. *Internet ve státní správě a samosprávě* [online]. 2009 [cit. 2009-06-22]. Dostupný z WWW: <http://www.issc.cz/archiv/2009/download/prezentace/bernas_na.pdf>.

²⁴⁴ viz kapitola 3.2.3.1.2 Strategie sklizení

knihovnou, která používá Dublin Core formát metadat. Systém metadat v oblasti spisové a archivní služby se zakládá na normě MoReq2²⁴⁵, která definuje svůj vlastní model metadat.²⁴⁶

Základní typy metadat pro digitální archiv jsou tyto:

- Popisná – tvořena primárně v ERMS
- Uchovávací – tvořena archivem
- Strukturální

Dokud však nebude v platnosti nová legislativa nelze s jistotou říci, jak budou používaná metadata vypadat.

V případě ukládání webových stránek bude nutné stanovit podmínky pro předávání webů do archivu stejně jako u ostatních digitálních dokumentů. Předně stanovit formáty a složení, ve kterých bude archiv určené weby přijímat. Například obsah webu bude zřejmě limitován přístupem přes protokol HTTP²⁴⁷. Základním formátem webu bude HTML nebo XHTML²⁴⁸. Archiv může požadovat převedení některých dynamických částí webu do statického HTML a určit, které skripty bude akceptovat.²⁴⁹ Hlavní podmínkou bude, aby po předání do archivu byla zachována čitelnost a funkčnost webu v standardním prohlížeči.

Webové stránky předávané do archivu by rovněž měla provázet dokumentace obsahující například:

- Jméno webové stránky
- Webovou a serverovou platformu a také specifikaci softwaru potřebného k správné funkci stránek
- Název a verzi použitelného prohlížeče
- Datum sejmutí webu (zhotovení kopie)
- Další potřebné údaje²⁵⁰

Posledním tématem je nahlížení do archiválií, čili zabezpečení přístupu badatelů k archivovaným webům. Badatel bude k digitálním dokumentům přistupovat nepřímo prostřednictvím nástrojů výpočetní techniky (počítače) – tzv. nepřímý přístup. Takový přístup vznesne jistě i nároky na vybavení a provoz badatelny. Navíc by mohl vzniknout problém

²⁴⁵ viz kapitola 4.2.2 **Webové stránky v kontextu spisové služby**

²⁴⁶ BERNAS, Jiří, KUNT, Miroslav. Národní digitální archiv: Představení projektu. *Internet ve státní správě a samosprávě* [online]. 2009 [cit. 2009-06-22]. Dostupný z WWW:

²⁴⁷ viz kapitola 2.2 **Služba WWW**

²⁴⁸ Tamtéž

²⁴⁹ MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchování webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 252.

²⁵⁰ MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchování webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 253.

s nedostatkem volného místa či prostředků k prohlížení digitálních dokumentů v badatelně. Ve světě je často řešen tento problém zpřístupňováním některých archiválií online.²⁵¹ U nás stanovuje podmínky nahlížení do archiválií zákon²⁵² a badatelský řád příslušného archivu. Do archiválií je možné nahlížet pouze v prostorách badatelny s výjimkou výstav. Badatel je vždy povinen vyplnit badatelský list a až pak je mu umožněno prohlížení archiválií, což v prostředí Internetu není dost dobře možné. Navíc běžně přístupné jsou archiválie, od jejichž vzniku uplynulo minimálně třicet let. Nahlížení do archiválií prostřednictvím Internetu se do budoucna jeví jako velmi perspektivní. Bude však zapotřebí úprav v archivním zákoně a stanovení podmínek, jak a které archiválie budou moci být zpřístupněny online.

5 Závěr

Problematika dlouhodobého uchovávání webových stránek v oblasti českého archivnictví je v současné době na úplném počátku. Účelem této práce nebylo navrhnout jasná řešení a postupy nebo určit správné technologické provedení. Toto není možné z několika důvodů: Otázka archivace webů vyžaduje dlouhodobý vývoj a odborný přístup specializovaných pracovišť (Národní digitální archiv, Odbor archivní správy MV). Je potřeba řešit jak otázky legislativy a nové postupy např. nové skartační lhůty, tak otázky technologické. V neposlední řadě nemám potřebné hlubší znalosti v oboru informačních technologií a také dlouholetou praxi ve spisové a archivní službě, která je neméně důležitá.

Ve své práci jsem se pokusil především shrnout i veškerou okolní problematiku, jako pohled na weby a Internet z technické stránky a archivaci webů u institucí (knihovny), které již mají v této oblasti zkušenosti a stojí za nimi několikaletý výzkum. Co se týče hlavního tématu práce, snažil jsem se z mého pohledu nastínit základní problémy a specifika stránek a současně nastínit některé možné směry či teoretická řešení. Mimo to spatřuji vlastní přínos práce v provedeném průzkumu, jehož výsledky jsou zajímavé a podnětné, ikdyž nemohou být považovány za směrodatné v širším významu a nepřesahují rámec bakalářské práce. Přesto doufám, že položené otázky někoho z oslovených inspirovaly nebo alespoň podnítily k zamyšlení. Tímto chci poděkovat všem, kteří se průzkumu zúčastnili za jejich příspěvek k věci.

²⁵¹ viz kapitola 4.2 **Archivy a dlouhodobé uchovávání webových stránek**

²⁵² *Zákon č. 499/2004 Sb., o archivnictví a spisové službě a o změně některých zákonů.*

A jak bych tedy odpověděl na otázku, která je zároveň názvem mé bakalářské práce: „*Webové stránky jako archiválie?*“ Určitě ANO, avšak cesta, kterou bude potřeba projít než se tak stane, bude ještě dlouhá a trnitá. Výhledově do budoucna se s archivací webu nezačne pravděpodobně dříve jak po dokončení Národního digitálního archivu a dokud nebude plně zaběhnuta archivace digitálních dokumentů. Jako nejschůdnější vidím řešení v začlenění webů mezi ostatní digitální dokumenty, samozřejmě s ohledem na jejich specifika. Jak už jsem říkal, toto všechno si bude žádat úpravy v legislativě a nové instrukce pro původce i archiváře, což je samo o sobě běh na dlouhou trať. Zatím se musíme spolehnout na Národní knihovnu a její projekt WebArchiv, že se mu podaří uchovat to nejdůležitější z českého webu. Možná je i případná spolupráce Národní knihovny a archivu, protože se může stát, že obě instituce budou archivovat stejný web.

Na závěr si dovolím jednu svou myšlenku ne úplně související s tématem. Už samotné zaměření této práce a některé v ní zmíněné problematiky naznačují, kam by se mohlo povolání archiváře v budoucnu posouvat a jaké znalosti a dovednosti by od něj mohly být očekávány. Je otázkou i přizpůsobení struktury stávajících vysokoškolských oborů nebo vznik oborů nových, tak aby vyhovovaly i nárokům možná už velmi blízké budoucnosti. Rozhodně to stojí za hlubší zamyšlení.

6 Seznam pramenů a literatury

I. Monografie a akademické práce

BROŽA, Petr. *Tvorba WWW stránek pro úplné začátečníky*. 5. aktualiz. vyd. Brno: Computer Press, a.s., 2006. 149 s. ISBN 80-251-1300-0.

BROŽA, Petr, BURANSKÝ, Imrich. *Programování WWW stránek pro úplné začátečníky*. 3. aktualiz. vyd. Brno: Computer Press, 2003. 202 s. ISBN 80-7226-818-X.

CELBOVÁ, Ludmila, et al. *Archivace webu*. 1. vyd. Praha: Národní knihovna ČR, 2008. 40 s. ISBN 978-80-7050-562-5.

KOUDELA, Vlastimil. *Analýza WWW prezentací veřejné správy*. Pardubice, 2005. 85 s. Univerzita Pardubice, Ústav systémového inženýrství a informatiky. Vedoucí bakalářské práce Miloslava Bartoníčková.

II. Periodika a články ve sbornících

BARTOŠEK, Miroslav. Vyhledávání v Internetu a DUBLIN CORE. *Zpravodaj ÚVT MU*. ISSN 1212-0901, 1999, roč. IX, č. 4, s. 1-4.

MACEK, Oskar, WANNER, Michal. *Skartační řízení nebo automatizovaná sklizeň? : K problematice dlouhodobého uchování webů*. Archivní časopis. 2005, roč. 55, č. 4, s. 241-254.

III. Elektronické zdroje

BERNAS, Jiří, KUNT, Miroslav. Národní digitální archiv : Představení projektu. *Internet ve státní správě a samosprávě* [online]. 2009 [cit. 2009-06-22]. Dostupný z WWW: <www.iss.cz/archiv/2008/download/prezentace/bernas_nap.ppt >

BERNAS, Jiří. Národní digitální archiv. *Co po nás zbude* [online]. 2009 [cit. 2009-06-22]. Dostupný z WWW: <<http://www.cnz.cz/ke-stazeni/2009/seminar-ds-mza-brno/prezentace/NDA-20090224.pdf>>.

BERNAS, Jiří. Principy digitální archivace. *Internet ve státní správě a samosprávě* [online]. 2009 [cit. 2009-06-22]. Dostupný z WWW: <http://www.issz.cz/archiv/2009/download/prezentace/bernas_na.pdf>.

CELBOVÁ, Ludmila. Český web a povinný výtisk – jde to spolu dohromady? *Knihovna* [online]. 2008, roč. 19, č. 2, s. 59-75 [cit. 2009-03-28]. Dostupný z WWW: <<http://knihovna.nkp.cz/knihovna82/82005.htm>>. ISSN 1801-3252.

CELBOVÁ, Ludmila. Elektronické zdroje publikované v síti Internet jako součást České národní bibliografie. *Ikaros* [online]. 2000, roč. 4, č. 6 [cit. 2009-05-10]. Dostupný na WWW: <<http://www.ikaros.cz/node/602>>. URN-NBN:cz-ik602. ISSN 1212-5075.

CELBOVÁ, Ludmila. Povinný výtisk elektronických publikací, zejména vzdálených elektronických zdrojů. *Ikaros* [online]. 2000, roč. 4, č. 10 [cit. 2009-03-28]. Dostupný na WWW: <<http://www.ikaros.cz/node/670>>. URN-NBN:cz-ik670. ISSN 1212-5075.

CELBOVÁ, Ludmila. Registrace a zpřístupňování elektronických zdrojů publikovaných v síti Internet. *INFORUM : konference o profesionálních informačních zdrojích* [online]. 2000 [cit. 2009-05-07]. Dostupný z WWW: <<http://www.inforum.cz/archiv/inforum2000/prednasky/registraceazp.htm>>.

COUFAL, Libor. Archivace (nejen) českého webu z pohledu WebArchivu : Jak se na archivaci webu dívají autoři projektu WebArchiv?. *Čertoděj : čertovské věci o Web 2.0 a web trendech* [online]. 2007 [cit. 2009-04-07]. Dostupný z WWW: <<http://certodej.cz/S9314>>.

Creative Commons Česká republika [online]. 2008 [cit. 2009-04-23]. Dostupný z WWW: <<http://www.creativecommons.cz/o-projektu/>>.

ČNB: *Česká národní bibliografie* [online]. [2008] [cit. 2009-05-07]. Dostupný z WWW: <http://www.aipberoun.cz/Site/produkty_cnb_cz.htm>.

Dublin Core - Czech homepage [online]. c1995-2005 , Poslední aktualizace: 20. listopad 2006 14:20:28 [cit. 2009-05-16]. Dostupný z WWW: <http://www.ics.muni.cz/dublin_core/elems.html>.

Google Cached Pages: What Are Cached Pages? - Google Guide [online]. This page was last modified on: Tuesday March 13, 2007 [cit. 2009-04-05]. Dostupný z WWW: <http://www.googleguide.com/cached_pages.html>.

GRUBER, Lukáš. Licence Creative Commons a perspektiva jejich zavedení do českého prostředí. *Ikaros* [online]. 2008, roč. 12, č. 3 [cit. 2009-04-05]. Dostupný na WWW: <<http://www.ikaros.cz/node/4612>>. URN-NBN:cz-ik4612. ISSN 1212-5075.

Historie Internetu | i-extra.net [online]. c2007-2008 [cit. 2009-02-06]. Dostupný z WWW: <<http://i-extra.net/internet-a-site/historie-internetu>>. ISSN 1803-134X.

HUTAŘ, Jan, MATĚJKA, Lukáš, CELBOVÁ, Ludmila. Vyhledávání v archivu českých webových zdrojů. *INFORUM 2006: 12. konference o profesionálních informačních zdrojích* [online]. 2006 [cit. 2009-04-02]. Dostupný z WWW: <http://www.inforum.cz/pdf/2006/Hutar_Jan.pdf>.

International Internet Preservation Consortium [online]. c2004-2008 [cit. 2009-04-11]. Dostupný z WWW: <<http://www.netpreserve.org/about/index.php>>.

Internet Archive [online]. [cit. 2009-04-03]. Dostupný z WWW: <<http://www.archive.org/>>.

JANOVSKÝ, Dušan. *Jak psát web, návod na html stránky* [online]. [2008], Poslední aktualizace 27. prosince 2008 [cit. 2009-01-29]. Dostupný z WWW: <<http://www.jakpsatweb.cz/>>. ISSN 1801-0458.

KOSEK, Jirí. *Dynamické HTML: Úvod* [online]. c1998, Poslední modifikace: 24-01-98 [cit. 2009-02-15]. Dostupný z WWW: <<http://www.kosek.cz/clanky/dhtml/uvod.html>>.

KUNT, Miroslav. *Elektronické dokumenty z hlediska požadavku zákona č. 499/2004 Sb. a potřeb jejich dlouhodobého ukládání* [online]. 2005, 29.5. 2005 [cit. 2009-06-21]. Dostupný z WWW: <http://www.nacr.cz/Z-Files/elektronicke_doku.pdf>.

Library of Congress [online]. Last Updated: 08/22/2008 [cit. 2009-04-03]. Dostupný z WWW: <<http://www.loc.gov/about/>>.

MALÍKOVÁ, Jana. *Novela zákona o archivnictví a spisové službě* [online]. c2008 [cit. 2009-06-01]. Dostupný z WWW: <<http://www.mvcr.cz/clanek/novela-zakona-o-archivnictvi-a-spisove-sluzbe.aspx>>.

NOVOTNÝ, Luděk. *Internet Archive - celosvětový archiv webu* [online]. 2004 [cit. 2009-03-04]. Dostupný z WWW: <dspace.muni.cz/bitstream/ics_muni_cz/72/1/i-archiv.pdf>.

Pracovní verze národního standardu pro spisovou službu vykonávanou v elektronické podobě. [online]. Ministerstvo vnitra České republiky [cit. 2009-06-21]. Dostupný z WWW: <www.mvcr.cz/soubor/moreq-pracovni-pdf.aspx>.

Vydavatelství VŠCHT Praha [online]. 2000 [cit. 2009-06-21]. Dostupný z WWW: <http://vydavatelstvi.vscht.cz/knihy/uid_es-005/hesla/trida.OAAA.html>.

WebArchiv - archiv českého webu [online]. Poslední aktualizace: 12.5.2009 [cit. 2009-05-12]. Dostupný z WWW: <<http://www.webarchiv.cz/vydavatele/>>.

Wikipedie, otevřená encyklopedie [online]. 2002, 29. ledna 2009 [cit. 2009-01-29]. Dostupný z WWW: <http://cs.wikipedia.org/wiki/Hlavn%C3%AD_strana>.

ZEMAN, Mirek. Creative Commons – budoucnost copyrightu?. *LUPA: SERVER O ČESKÉM INTERNETU* [online]. 2004 [cit. 2009-04-12]. Dostupný z WWW: <<http://www.lupa.cz/clanky/creative-commons-8211-budoucnost-copyrightu/>>. ISSN 1213-0702.

ŽABIČKA, Petr. NEDLIB Harvester. *Ikaros* [online]. 2000, roč. 4, č. 10 [cit. 2009-04-30]. Dostupný z WWW: <<http://www.ikaros.cz/node/672>>. URN-NBN:cz-ik672. ISSN 1212-5075.

IV. Legislativa

Zákon č. 37/1995 Sb., o neperiodických publikacích.

Zákon č. 46/2000 Sb., o právech a povinnostech při vydávání periodického tisku a o změně některých dalších zákonů (tiskový zákon).

Zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským.

Zákon č. 216/2006 Sb., změna autorského zákona a některých dalších zákonů.

Zákon č. 257/2001 Sb., o knihovnách a podmínkách provozování veřejných knihovnických a informačních služeb (knihovní zákon), ve znění pozdějších předpisů

Zákon č. 273/1993 Sb. o některých podmínkách výroby, šíření a archivování audiovizuálních děl.

Zákon č. 499/2004 Sb., o archivnictví a spisové službě a o změně některých zákonů.

Pracovní návrh textu vyhlášky – náhrady vyhlášky č.646/2004 Sb., o podrobnostech výkonu spisové služby – varianta 8. 10. 2008.

7 Přílohy

I. Výsledek průzkumu

Archivy

Otázky:

- 1) Myslíte si, že weby některých původců mají potenciál stát se archiváliemi?
- 2) Zabývali jste se někdy už problematikou dlouhodobého uchovávání webových stránek jako archivních dokumentů?
- 3) Oslovil vás někdy některý původce ohledně návrhu webových stránek do skartačního řízení či možnosti uložení stránek do archivu?
- 4) Je ve vašich možnostech zabezpečit uložení a následné zpřístupnění webu badatelům, pokud by prošel skartačním řízením?
- 5) Uchovávejte starší verze vašeho webu?

Celkem odpovědělo: 49

Výsledky:

Otázka	Odpověď: ANO	NE
Č. 1	69%	31%
Č. 2	24%	76%
Č. 3	0%	100%
Č. 4	12%	88%
Č. 5	10%	90%

Původci

Otázky:

- 1) Myslíte si, že informační hodnota vašeho webu je trvalá či je pouze poplatná době, kdy jsou tyto informace aktuální?
- 2) Navrhli by jste vaše webové stránky jako jeden z typů el. dokumentů do skartačního řízení pro pozdější uložení do archivu jako archiválie?
- 3) Vztahují se na vaše webové stránky nějaké vnitřní normy?
- 4) Jak často se v průměru vaše stránky aktualizují?
- 5) Používáte služeb placeného webhostingu?
- 6) Uchováváte starší verze vašeho webu?

Celkem odpovědělo: 39

Výsledky:

Otázka	Odpověď: ANO	NE
Č. 1	74%	26%
Č. 2	38%	62%
Č. 3	67%	33%
Č. 4	Odpověď:denně – 69%	
Č. 5	44%	56%
Č. 6	44%	56%

8 Resumé

This bachelor paper deals with Web items as archival documents. This work is based on the article about long-term preservation of Web items published in Archival magazine. As other resources are used specialised websites, professional literature, valid legislation and own research.

The first part of this work describes Internet basics and particularly its utility of World Wide Web. I mainly focused on problematics of websites technologies, like HTML (Hyper Text Markup Language), cascades styles and individual scripts. This work also contains chapter about webhosting and domains.

Next part dwell on non-archive institutions which have been already working on long-term preservation of Web items for a long time. These organizations are Internet Archive and national libraries of many countries. All problematics is described on the example of National library in the Czech republic and its project of WebArchiv.

In the final part I tried to clarify to explain the websites role in the system of records management and records keeping with taking into account the law 499/2004 Sb. „o archivnictví a spisové službě a o změně některých zákonů a prováděcích vyhlášek“ etc. First of all I outlined some possible methods and solution of long-term preservation of Web items. There is also mentioned the question of permanent value of Web items and prospects for future.