

UNIVERZITA KARLOVA V PRAZE
FILOZOFICKÁ FAKULTA
ÚSTAV INFORMAČNÍCH STUDIÍ A KNIHOVNICTVÍ

Studijní program: Informační studia a knihovnictví

Studijní obor: Informační studia a knihovnictví

BAKALÁŘSKÁ PRÁCE

Adam Kolín

Projekty masové digitalizace tištěných dokumentů

Mass digitization projects of printed documents

Praha, 2012

Vedoucí: Mgr. Blanka Vorlíčková

Rád bych tímto poděkoval Mgr. Blance Vorlíčková, která mi při konzultacích poskytla potřebné rady a také připomínky ohledně náplně, obsahu a struktury této bakalářské práce.

Prohlašuji, že jsem bakalářskou práci vypracoval samostatně, že jsem řádně citoval všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 30. 7. 2012

.....

podpis

Abstrakt

Bakalářská práce představuje problematiku masové digitalizace a jejích projektů, které v současnosti ovlivňují mnoho digitalizačních aktivit po celém světě a které mají dopad na různé oblasti knihovnictví, jako jsou přístup k informacím, role knihoven a knihy nebo autorsko-právní vztahy.

Úvodní část je věnována vymezení pojmu masová digitalizace a jejímu zasazení do kontextu digitalizačních snah. Ve druhé části jsou stručně popsány výhody a nevýhody projektů masové digitalizace a její jednotlivé aspekty, např. autorská práva, OCR technologie, metadata, finance, dlouhodobá ochrana dokumentů atd., které do značné míry tyto projekty ovlivňují. Další část rozděluje projekty masové digitalizace podle způsobu financování. Jednotlivé projekty jsou zde charakterizovány v kontextu se situací, která panuje v Asii a Evropě s uvedenými příklady z jednotlivých zemí a se stavem v České republice.

Závěrečná část pak srovnává projekty Google Book Search a Open Content Alliance s Norskou digitální knihovnou, jako zástupcem masové digitalizace mezi národními digitálními knihovnami. Hodnocení výsledků srovnání a vlivy masové digitalizace jsou uvedeny v závěru bakalářské práce.

Klíčová slova:

digitalizace, masová digitalizace, digitalizační projekty, Google Book Search, Open Content Alliance

Abstract

The bachelor thesis introduces the issue of mass digitization and its projects, which currently affect many digitizing activities all over the world and which have an impact on various areas of librarianship, such as access to information, roles of libraries and books or copyright relations.

The introductory part is devoted to the definition of mass digitization and its placement in the context of digitization efforts. The second part briefly describes the advantages and disadvantages of mass digitization projects and its particular aspects, such as copyrights, OCR technology, metadata, finance, long-term preservation of documents etc., which largely affect those projects. Mass digitization projects are divided by types of financing in the next part. Particular projects are characterized in the context with the situation that prevails in Asia and Europe with mentioned examples from particular countries, and the situation in the Czech Republic.

The final part compares the projects Google Book Search and Open Content Alliance with the Norwegian National Digital Library, as a representative of mass digitization among national digital libraries. Evaluation of results of comparison and the effects of mass digitization are stated at the end of the bachelor thesis.

Key words:

digitization, mass digitization, digitization projects, Google Book Search, Open Content Alliance

OBSAH

1	Úvod.....	8
2	Definice masové digitalizace	10
3	Přínosy a problémy masové digitalizace tištěných dokumentů	14
3.1	Problémové aspekty masové digitalizace	14
3.1.1	Finance	14
3.1.2	Autorská práva	15
3.1.3	Výběr materiálů k digitalizaci	16
3.1.4	Technologie	16
3.1.5	OCR technologie.....	16
3.1.6	Metadata.....	17
3.1.7	Standardy souborů.....	18
3.1.8	Dlouhodobá ochrana digitálních dat.....	18
3.1.9	Přístup k digitalizovaným dokumentům	19
3.1.10	Uživatelské rozhraní	20
3.2	Výhody a nevýhody masové digitalizace.....	20
4	Projekty masové digitalizace	22
4.1	Rozdělení projektů masová digitalizace	22
4.2	Projekty financované soukromými subjekty	22
4.2.1	Google Book Search	22
4.2.2	Windows Live Search Books Program	23
4.3	Projekty konsorcií.....	25
4.3.1	Open Content Alliance	25
4.3.2	The Million Book Project	27
4.4	Situace mimo Evropu	28
4.4.1	Japonsko.....	29
4.4.2	Jižní Korea.....	30
4.4.3	Čína.....	31
4.5	Situace v Evropě	32
4.5.1	Francie	33
4.5.2	Norsko	35
4.5.3	Situace v České republice.....	37
4.6	Další projekty	38
5	Srovnání vybraných projektů masové digitalizace	39

5.1	Kritéria výběru projektů masové digitalizace.....	39
5.2	Srovnání celosvětových projektů masové digitalizace.....	43
5.2.1	Cíle projektů	43
5.2.2	Počet partnerů	43
5.2.3	Finance	43
5.2.4	Autorská práva	44
5.2.5	Výběr materiálů k digitalizaci	45
5.2.6	Metadata	45
5.2.7	Technologie – skenery atd.	45
5.2.8	Standardy souborů	46
5.2.9	Přístup k digitalizovaným dokumentům	46
5.2.10	Počet zpřístupněných dokumentů	46
5.2.11	Obsah	46
5.2.12	Uživatelské rozhraní	47
5.3	Výsledky srovnání.....	49
5.4	Rámcové srovnání projektů masové digitalizace s vybranou národní digitální knihovnou ...	49
5.4.1	Srovnání projektů masové digitalizace s digitalizačním projektem Norské národní knihovny	50
6	Závěr	51
7	Seznam literatury	53
	Přílohy	60

Seznam použitých zkratek

Bibliothèque National de France	BNF
Google Book Search	GBS
národní digitální knihovna	NDK
Open Archival Information System	OAIS
Open Content Alliance	OCA

1 Úvod

Tématem bakalářské práce je masová digitalizace a její projekty. Digitalizace v jakékoli formě je jednou z mnoha náplní současného knihovnictví, které do značné míry ovlivňují i další témata s ním spojená, např. otázka autorských práv.

K výběru tohoto tématu jsem se rozhodl, protože se v současné době jedná o jednu z možných cest, jak alespoň z části získat zpět pozornost potencionálních i stávajících uživatelů a rozšířit služby knihoven v internetovém prostředí. Knihovny totiž mají stále co nabídnout po obsahové stránce dnešním uživatelům, kteří si zvykli na jednoduchou možnost získání informací přes různé druhy vyhledávačů. Ty z velké části nabízející informace, jejichž relevance nemusí být příliš vysoká. Knihovny a jejich fondy však po obsahové stránce zaručují vysokou kvalitu.

Cílem bakalářské práce je popsat definice masové digitalizace, analyzovat a porovnat vybrané projekty, které jsou zaměřeny na masovou digitalizaci tištěných dokumentů, tzn. především knih a periodik.

U projektů se práce zaměří především na zjištění, jaké poskytují výstupy a služby, v čem zásadním se tyto projekty liší a co je naopak spojuje, jakým způsobem prezentují své výstupy a do jaké míry ovlivňují projekty masové digitalizace a digitalizaci v evropských národních knihovnách. Budou zde popsány přínosy a nevýhody masové digitalizace, které s tímto tématem přímo souvisí.

V úvodní kapitole je představena masová digitalizace jako pojem, který je potřeba vymezit, protože je chápán několika způsoby. Existuje několik projektů masové digitalizace, které vykazují společné rysy, přesto termín masová digitalizace zatím stále není jednotný.

V následující kapitole jsou stručně představeny jednotlivé aspekty masové digitalizace, jako jsou otázky autorských práv či metadat. Podle toho jsou pak vyvozeny výhody a nevýhody masové digitalizace v obecné rovině.

Představení vybraných projektů je náplní 4. kapitoly, zaměřující se na popis samotných projektů masové digitalizace. Projekty jsou představeny v následujících oblastech: je zde pojednáno o jejich historii, o jejich obsahové stránce a o institucích do nich zapojených. V podkapitole budou také zmíněna základní fakta o situaci v asijských a evropských digitalizačních aktivitách souvisejících s masovou digitalizací nebo na ni navazující.

Praktickou částí bakalářské práce je pátá kapitola, která se zabývá podrobným porovnáním vybraných projektů masové digitalizace na základě vybraných kvantitativních a kvalitativních kritérií. Budou popsány celkové výsledky ze srovnání i rozdíly a společné rysy těchto projektů.

V závěru práce jsou shrnuty získané poznatky a nastíněny možnosti dalšího vývoje v této oblasti.

Veškerá použitá literatura je citována v souladu s normou ISO 690:2010. Citace v textu jsou uváděny podle Harvardského systému.

2 Definice masové digitalizace

Pojem masová digitalizace nebyl po několik let přesně definovaný. Dnes stále existuje několik pohledů, které jednotliví odborníci zastávají. Jedni zastávají názor, že jde o přirozený proces vývoje samotné digitalizace, která nemá přesné hranice (SCHMITZ, 2008; MURRELL, 2010). Druzí to vyvracejí a na příkladech jednotlivých projektů masové digitalizace ukazují opak a definují konkrétní znaky (COYLE 2006; MILNE, 2006; VERHEUSEN, 2008).

První skupina odborníků vychází z tvrzení, že vývoj procesu digitalizace je logický. Za poslední desítky let totiž došlo k velkému posunu v technickém a technologickém vývoji, který je do jisté míry určující pro posuny v kvantitě a kvalitě digitalizace. To dokazuje tvrzení D. Schmitze (SCHMITZ, 2008), který ve svém článku uvádí, že masovou digitalizaci, navzdory velkému množství projektů masové digitalizace, nelze čistě oddělit od „large-scale“ digitalizace z prostého důvodu jejich provázanosti a kontinuity. Stejný názor návaznosti těchto termínů zastává i Murrell (MURRELL, 2010).

Praxe i publikační prameny ukazují, že existují různé druhy digitalizace a je nutno striktně tyto druhy digitalizace oddělit. Rozdíly mezi definicemi těchto druhů digitalizace nejsou veliké, přesto pro potřeby práce je nutný výběr jedné definice, která poslouží k jasnému a přesnému výběru projektů masové digitalizace.

Jedna z nejpropracovanějších definic vymezuje masovou digitalizaci jako konverzi materiálu v průmyslovém měřítku. Tím se myslí převod celého fondu knihovny bez jakéhokoli výběru dokumentů do digitální podoby pomocí procesu skenování (COYLE, 2006).

Hlavním cílem masové digitalizace je tedy digitalizovat každou knihu, která kdy byla vytištěna (COYLE, 2006). To je přesným opakem digitálních knihoven sbírek jako American Memory Project, World Digital Library nebo Manuscriptorium, kde jsou dokumenty k digitalizaci pečlivě vybírány na základě určených kritérií, jako je např. původ dokumentu, časové určení vzniku dokumentu atd. K naplnění hlavního cíle masové digitalizace jsou nutné prostředky a postupy, jež splňují dvě nejpodstatnější vlastnosti pro masovou digitalizaci. Jedná se o efektivnost, která je zajištěna maximální možnou automatizací, a především rychlost celého procesu. K dosažení těchto dvou vlastností v postupech masové digitalizace je třeba snaha co nejvíce minimalizovat lidský prvek.

Masová digitalizace je tak postavena na efektivním a rychlém skenování knih, po němž následuje převedení naskenované stránky s textem z „obyčejného“ obrázku na text

v digitální podobě pomocí OCR technologie (optické rozpoznávání znaků) (COYLE, 2006). Tím se text stává použitelným pro další práci, např. vyhledávání v textu, indexaci atd. (COYLE, 2006).

Karen Coyle (COYLE, 2006) rozlišuje další dva druhy digitalizace: nemasová digitalizace (vlastní překlad autorkou použitého termínu non-mass digitization¹) a digitalizace „tzv. large scale“. V literatuře se také objevuje pojem „boutique“ digitalizace, který se používá pro projekty, jejichž parametry se dají srovnat s digitalizací nemasovou a „large scale“ (MILNE, 2006).

Nemasová digitalizace se liší od masové digitalizace tím, že jsou dokumenty k digitalizaci pečlivě vybírány. Projekty spojené s tímto postupem jsou primárně určeny k ochraně skenovaných dokumentů, jejichž digitální kopie slouží ke zpřístupnění širšímu okruhu uživatelů, než je možné u fyzického dokumentu. Jedná se především o dokumenty či sbírky, které jsou vzácné a které mají vysokou historickou hodnotu, např. staré tisky, nebo jsou uživateli žádané. Dalším charakteristickým rysem nemasové digitalizace, odlišujícím se od masové digitalizace, je forma zpracování konečného výstupu. U masové digitalizace je primární výstup pouze fotokopie stran v knize zpracovaná OCR technologií, zatímco u nemasové digitalizace je v mnohých případech vytvářen dokument, který nabízí další nadstavbové služby, jako jsou odkazy k referenčním slovníkům a mnoho dalších (COYLE, 2006).

K digitalizaci označené jako „large scale“ patří velká řada projektů. Pro „large scale“ digitalizaci je na rozdíl od masové digitalizace charakteristická výběrovost, nicméně stejně jako v projektech masové digitalizace se skenuje velké množství stránek. Projekty „large scale“ digitalizace se zaměřují především na kompletní knihovní sbírky a fondy jako jeden celek. Mezi nejzajímavější projekty „large scale“ digitalizace patří projekt JSTOR. Není však zaměřený na knihy, ale na periodika (COYLE, 2006).

Na pomezí „large-scale“ a masové digitalizace se nachází tzv. národní digitální knihovny. Projekty národních digitálních knihoven (dále jen NDK) lze těžko oddělit od obou druhů digitalizace, nicméně se domnívám, že na základě výše popsaných charakteristik jednotlivých druhů digitalizací lze obhájit tvrzení, že některé z nich jsou specifickými projekty masové

¹ Tento termín přebírají někteří autoři např. SONG, Linlin, HUANG, Ruhua. Analysis of Breadth and Glossary in Mass Digitization. *Tushuguan Qingbao Gongzuo*. 2009, Vol. 53, No. 21, s. 25-28. ISSN 0252-3116. nebo VRANA,

digitalizace, např. Gallica, Národní knihovna Norska, Japonska, Jižní Koreje, které jsou jak projektem národní digitální knihovny, tak projektem masové digitalizace. Projekty NDK jsou často zapojeny do globálnějších projektů typu Google Book Search, např. Bavorská státní knihovna nebo Katalánská národní knihovna. V Evropě se podílejí národní digitální knihovny na projektu Europeana, který po naplnění bude poskytovat digitální dokumenty z celé Evropy.

Výběr dokumentů je u projektů NDK přísně ohraničen národní produkcí dokumentů, případně dokumentů týkajících se tematicky dané země za určité období. V tomto bodě se střetávají definice masové a „large-scale“ digitalizace, jež mají jiný pohled na výběr dokumentů. Myslím si, že výběr dokumentů v projektech NDK ve smyslu „large-scale“ digitalizace plně neodpovídá, jelikož nedigitalizují pouze jednotlivé fondy dokumentů z jednotlivých knihoven, ale snaží se digitalizovat všechny dokumenty, které se vztahují k dané národní produkci. Tato snaha v projektech NDK je digitalizovat kompletní národní produkci, přičemž výsledky této aktivity se pak spojují i v projektech, které jsou označovány za projekty masové digitalizace, např. Google Book Search, Europeana. Projekty masové digitalizace jsou tak zčásti tvořeny i menšími projekty digitalizace, než je Google Book Search nebo Open Content Allinace. Masovost projektů, jako jsou NDK, lze spatřovat v agregaci výstupů těchto projektů, jak to činí producenti projektů jako je HaithiTrust, Google Book Search, Internet Archive nebo Europeana. Tato tenká hranice a prolínání digitalizačních projektů zčásti objasňuje, proč není jednotná definice masové digitalizace a záleží na kontextu vytváření projektu NDK a úhlu pohledu na tyto projekty.

Autoři, kteří odvozují definici masové digitalizace od projektu Google Book Search (Library Project) a používají rysy tohoto projektu pro definování masové digitalizace, uvádějí stejné charakteristiky jako K. Coyle (COYLE, 2006). Shodují se na tom, že digitalizace probíhá v průmyslovém měřítku a že knihy jsou digitalizovány tzv. „en masse“ neboli masově. Shoda panuje i v názoru, že selekce materiálů pro digitalizaci je prováděna jen do té míry, aby formát dokumentů byl vhodný pro používané skenovací technologie (MILNE, 2006; VERHEUSEN, 2008).

Často se také uvádí (LACKIE, 2008; BAKSIK, 2011; VERHEUSEN, 2008; ST. CLAIR, 2006), že masová digitalizace u projektu Google Book Search je více digitalizací textu než knih samotných, jelikož jsou vynechávány ilustrace, obrázky a další doprovodné netextové prvky.

S tím koresponduje i fakt, že u těchto projektů bývají vynechány speciální fondy zapojených knihoven, jako jsou fondy map nebo prvotisků (VERHEUSEN, 2008).

Většina autorů odvozuje definici a popis masové digitalizace od nejznámějšího projektu Google Book Search, ale K. Coyle připojuje i pohled na ostatní významné projekty masové digitalizace např. Open Content Alliance. Právě z tohoto důvodu bylo její rozdělení digitalizace vybráno jako základ pro další části této bakalářské práce, jelikož nejvýstižněji popisuje jednotlivé aspekty různých druhů digitalizací a posouvá definici do obecnější roviny než ostatní. Pro potřeby této bakalářské práce je tedy masová digitalizace definována jako konverze dokumentů v průmyslovém měřítku, kdy jsou kompletní knihovní fondy digitalizovány bez individuálního výběru dokumentů. Toho je dosaženo efektivním, rychlým a maximálně automatizovaným způsobem skenování a maximální automatizací dalšího zpracování dokumentů s minimálním zásahem lidského faktoru.

3 Přínosy a problémy masové digitalizace tištěných dokumentů

3.1 Problémové aspekty masové digitalizace

Existuje celá řada aspektů, se kterými je potřeba počítat při spouštění jakéhokoli projektu digitalizace. V obecné rovině se objevují problémy spojené s autorským právem, otázky ochrany a výběru materiálů, vytváření metadat nebo zajištění interoperability (LOPATIN, 2006). Tato kapitola nemá za cíl detailně představit jednotlivá specifika, ale stručně přiblížit jejich podstatné aspekty, které ovlivňují a určují přínosy a zápory masové digitalizace.

Autoři se v literatuře liší v pohledu na to, jaké prvky procesu masové digitalizace se považují za nejdůležitější. V této kapitole jsou zachyceny především prvky shodující se s definicí masové digitalizace K. Coyle (COYLE, 2006), která byla vybrána jako nejvýstižnější a popsána v předchozí kapitole, např. pracovní postupy, výstupy a struktura knih, uživatelské rozhraní nebo ochrana dokumentů. Nebudou ale opomenuty ani pohledy jiných autorů na prvky, které se v definici neobjevují a které jsou také významné pro tyto projekty, např. finanční stránka projektů nebo autorská práva.

3.1.1 Finance

Jedná se o nejdůležitější položku každého projektu masové digitalizace patřící mezi nejnákladnější snahy v oblasti digitalizace vůbec. Finance jsou potřeba na technické vybavení, platy zaměstnanců projektu, výběr² a ochranu materiálů, stejně tak na zajištění kvality digitálních souborů. Nesmí se zapomínat ani na náklady spojené s dalším zajištěním fungování projektu po skončení samotné digitalizace, jelikož bez zajištění peněz na další provoz internetových stránek, nákladů na uložení atp. by neměly projekty žádný smysl.

V roce 2007 byl učiněn odhad u projektu Open Content Alliance na náklady masové digitalizace. Odhad ceny skenování jedné stránky byl 10 centů a 30 \$ za knihu. V roce 2009 byla u projektu OCA tato cena potvrzena jako relevantní údaj (KAHLE, 2009). U dvou dalších projektů se hovoří o ceně mezi 5 a 10 dolary za knihu u Google Book Search, respektive o 6 \$ za knihu u Million Book Project (KAHLE, 2009). Náklady projektu Google Book Search se dají pouze předpokládat, jelikož producenti projektu tato data nezveřejňují. Francouzská národní knihovna odhadovala náklady masové digitalizace na 0,09 € za stránku v roce 2006. Tyto

² Zde je míněn výběr exempláře, který je v nejvhodnějším fyzickém stavu pro digitalizaci.

náklady pokrývají pouze skenování a optické rozpoznávání znaků OCR (SCHMITZ, 2008). Ve srovnání se zahraničím jsou ceny v České republice přibližně stejné, např. z ceníku Digitalizačního centra AV ČR vyplývá, že za barevný sken jedné strany formátu A4 v rozlišení 600 dpi je účtováno 8,80 Kč, což je asi 0.35 €. Průměrné náklady, ve kterých jsou započítány veškeré náklady od skenování přes zpracování po zajištění dlouhodobého uchování dokumentů, na digitalizaci jedné stránky činily v Národní knihovně Nizozemí (Koninklijke Bibliotheek) 1,30 € (VERHEUSEN, 2008).

Další možné náklady, se kterými je potřeba počítat, jsou náklady na vyrovnání spojené s otázkou autorských práv. Tato situace může nastat v souvislosti s rozšířením projektu masové digitalizace o díla chráněná autorským zákonem.

Projekty jsou financovány několika způsoby. Do projektů masové digitalizace se aktivně zapojují soukromé firmy, např. Google, Microsoft nebo Yahoo!, poskytující finanční zajištění projektu. Knihovny mohou vyčlenit i na tyto projekty finance ze svého rozpočtu nebo pomáhat přispívat do společného projektu konsorcia (VERHEUSEN, 2008). Možným přispěvatelem je státní pokladna, případně i Evropská Unie prostřednictvím operačních programů a jejich strukturálních fondů nebo další organizace. Je však nemožné, aby jednotlivé knihovny vyčlenily dostatečné množství potřebných prostředků ve svých rozpočtech na tak velké projekty, jako je masová digitalizace knih a časopisů. Financování projektů masové digitalizace je proto ve většině případů zajištěno kofinancováním jednak soukromými zdroji, tak i státními prostředky, snad jen s výjimkou Francie (NIELSEN, 2008) a Norska (SKARSTEIN, 2010). Francie i Norsko jsou v této oblasti průkopníky díky vlastním projektům (Gallica a NDK Norska). Oba jsou štědře financovány pouze ze státního rozpočtu (KNOLL, 2005; SKARSTEIN, 2010).

3.1.2 Autorská práva

Autorská práva jsou kromě financování nejpálčivějším a nejdiskutovanějším problémem masové digitalizace, a to po celém světě. Toto téma se dostalo nejvíce do popředí všech projektů masové digitalizace a vydalo by na samostatnou práci. Řeší se zde hlavně otázky týkající se tzv. osiřelých děl a děl chráněných autorským právem, ale i samotná délka trvání ochrany autorských práv či nalezení rovnováhy mezi zájmy držitelů autorských práv a uživatelů (NIELSEN, 2008).

Tento problém se týká několika subjektů, jako jsou autoři, nakladatelé, autorské svazy, knihovny atd., kteří se musí mezi sebou dohodnout na tom, jakým způsobem by měla být nastavena pravidla a právní normy upravující vztahy mezi nimi. Každá z výše jmenovaných skupin hájí jak svá práva, tak své finanční zájmy, a proto je dohoda velmi obtížná.

Příkladem dohody o vyrovnání, kdy se některé do sporu zapojené subjekty už dohodly na podmínkách umožňujících využití děl chráněných autorským právem s producenty projektu Google Book Search, který tato díla od počátku skenoval, je tzv. Google Book Search Settlement Agreement. Důvod je prostý. Jsou zde obavy ze strany autorů a nakladatelů, aby získali alespoň nějakou finanční kompenzaci (MURRELL, 2010; HAHN, 2006). Příkladné a prakticky fungující řešení problému digitalizace dokumentů, jež podléhají autorským právům, lze v Evropě nalézt alespoň na národní úrovni jednoho ze členských států unie Francii a Norska. Zbytek Evropy totiž nemá stále otázku autorských práv vyřešenu.

3.1.3 Výběr materiálů k digitalizaci

Další krok v procesu digitalizace úzce souvisí s autorskými právy (díla chráněná a nechráněná autorským zákonem) a fyzickým stavem dokumentů. Při výběru materiálů je brán zřetel na to, jaké formáty dokumentů jsou vhodné pro skenery, které jsou používány (VERHEUSEN, 2008). U tohoto bodu je také třeba zmínit problém duplicit při skenování dokumentů mezi jednotlivými projekty masové digitalizace. (COYLE, 2006). V současné chvíli neexistuje žádné řešení, které by zabránilo tomuto jevu. Nelze proto eliminovat špatné hospodaření s penězi vynakládanými na duplicitní digitalizaci některých děl.

3.1.4 Technologie

Masová digitalizace, jakožto plně automatizovaný proces, musí rychle zpracovat velké objemy dokumentů. Z tohoto důvodu se používají robotické skenery, které zvládnou skenovat dokumenty různých fyzických formátů a materiálů včetně cenných a křehkých. V této oblasti také dochází k neustálému vývoji, který se snaží celý proces skenování maximálně zefektivnit. Největší nevýhoda tohoto prvku masové digitalizace se odráží hlavně na finančních nárocích na pořízení takovýchto moderních nástrojů (NIELSEN, 2008).

3.1.5 OCR technologie

Nástroje a technologie optického rozpoznávání znaků je důležitou součástí převodu tištěné knihy do digitální podoby. OCR technologie převádí naskenovaný obraz textu na

plnohodnotný text. Umožňuje tak další práci s tímto textem, jako je vyhledávání v textu, automatická indexace nebo vytvoření hypertextových odkazů na příslušné další texty (SUTHERLAND, 2008). V literatuře je popsáno mnoho problémů souvisejících s OCR technologií, např. problémy OCR technologie u azbuky a ruských knih (NEUBERT, 2008) nebo problémy technologie OCR při převodu historických tištěných dokumentů v průběhu masové digitalizace (BALK, PLOEGER, 2009). U OCR nástrojů je tedy v první řadě důležité, aby zvládaly rozpoznat jazyky a typy písma, kterým je daný text napsán, a aby uměly s takovým jazykem a typy písma po skenování dokumentů pracovat.

U projektů masové digitalizace však není možné udržet takovou vysokou úroveň kvality optického rozpoznávání znaků jako u projektů nemasové a „large scale“ digitalizace. Není totiž reálné kontrolovat všechny naskenované soubory, jelikož by došlo k enormnímu zvýšení nákladů celého projektu masové digitalizace (VERHEUSEN, 2008). Provádět dodatečnou kontrolu a důsledně dodržovat standardy kvality digitálních kopií v takové rozsahu by vedlo ke zpomalení celého procesu. V tom případě by se již nedalo hovořit o masové digitalizaci (HAHN, 2006).

3.1.6 Metadata

V současnosti existuje mnoho druhů metadat, které jsou používány při projektech masové digitalizace, např. Dublin Core, MARC, MODS, METS nebo vlastní metadata speciálně vytvořená dle požadavků knihovny (LOPATIN, 2007). Jejich důležitost tkví v obsažených informacích. K tomu se váže i nutnost dobré kvality a zpracování metadat, která je důležitá nejen pro přístup k materiálům v digitálním repozitáři, ale i pro reprezentaci informací o materiálu samotném, jako je jeho struktura, tvůrce, formát nebo technické informace (LOPATIN, 2006). Tvorba metadat v projektech masové digitalizace z větší části probíhá automaticky (SCHMITZ, 2008), a proto musí být jejich kvalita kontrolována lidským faktorem. U nástrojů pro automatickou tvorbu metadat se nejčastěji používá především obsahová analýza. Není to však jediná používaná metoda. Existuje také např. metoda automatického generování metadat prostřednictvím sémantické sítě (RODRIGUEZ, 2009). Metadata také umožňují integraci a shromažďování informací o digitálních kopiích z různých portálů a služeb bez omezení na jedno místo (CEYNOWA, 2009).

3.1.7 Standardy souborů

Zde platí to samé, jako u předchozího bodu. V současné době existuje několik typů souborů, které jsou používány pro uložení naskenovaných textů. K. Coyle (2006) uvedla, že v roce 2006 byly nejčastěji používány v projektech masové digitalizace standardy souborů, jako je TIFF, JPEG 2000, DjVu nebo PDF, a tak je tomu i v současnosti. U těchto formátů jsou kladeny požadavky na nejlepší poměr mezi minimální velikostí souboru pro potřebu skladovací kapacity při co nejvyšší kvalitě. Důvody jsou jednoduché: vysoké náklady na uložení a velikost úložného prostoru. Z těchto důvodů probíhá u standardů, jako je TIFF nebo JPEG2000, neustálý vývoj (VERHEUSEN, 2008).

V souvislosti s požadavky na standardy uložení digitálních kopií se také hovoří o schopnosti konverze, ochrany a dlouhodobého uložení těchto souborů. Stabilita souborů je prioritou, kterou se však zatím nepodařilo vyřešit. Pro uchování digitálních kopií je klíčové nastavení určitého standardu kvality zpracování naskenovaného dokumentu. Nicméně zachování kvality zpracování u digitální kopie je poněkud v rozporu s tlakem na rychlou a levnou digitalizaci (COYLE, 2006).

S tímto prvkem masové digitalizace souvisí i problematika trvalého uložení digitálních dokumentů a jejich dlouhodobá ochrana.

3.1.8 Dlouhodobá ochrana digitálních dat

Aby se dalo hovořit o dlouhodobé ochraně digitálních dat, musí daná metoda splňovat tyto čtyři vlastnosti: uchovávání, zpřístupnitelnost, srozumitelnost a dlouhodobý horizont (CUBR, 2009). Původní představy o uložení na optických nosičích, jako jsou CD-ROMy nebo DVD (VERHEUSEN, 2008), byly shledány jako nevhodné na základě provedených výzkumů a praxe. Dnes se pro dlouhodobou ochranu digitálních dat využívá především metody digitální archivace. Požadavky na dlouhodobou ochranu jsou upraveny modelem Open Archival Information System (OAIS), podle kterého jsou vytvářeny digitální repozitáře. V oblasti digitální archivace existují čtyři základní strategie dlouhodobé ochrany. Dělí se podle problémových okruhů, které jsou rizikové pro naplnění cíle dlouhodobé ochrany. Jedná se o strategie technické roviny, informační roviny, systémové roviny a institucionální roviny. Např. u strategie informační roviny se dále dělí na další strategické podskupiny, jako jsou emulace, migrace, zapouzdření, strategie DRS, strategie UVC, modulární emulace, formátové strategie a metadatové strategie (CUBR, 2009).

Digitalizované dokumenty z projektů masové digitalizace jsou tak využity i jako kopie fyzických dokumentů pro dlouhodobé uchování. Naráží se tu ale na problém, který souvisí se samotnou povahou masové digitalizace. Existuje zde rozpor mezi rychlostí masové digitalizace a zachováním kvalitní digitální kopie papírového materiálu (COYLE, 2006). Není možné při rozsahu masové digitalizace zkontrolovat všechny digitalizované dokumenty. Z tohoto důvodu zde existuje možnost uchování nekvalitní digitální kopie fyzického dokumentu.

3.1.9 Přístup k digitalizovaným dokumentům

V oblasti přístupnosti k digitalizovaným dokumentům existují dva hlavní přístupy. Oba jsou založeny na dohodě mezi knihovnami a sponzorem, případně partnerem projektu. Dohoda určuje podmínky, za jakých zpřístupnění probíhá. První varianta dovoluje, aby digitalizované dokumenty byly zveřejněny pouze uživatelům knihovny, popřípadě uživatelům soukromé služby provozované partnerem projektu. Druhou možností je zveřejnění zdarma a volně každému uživateli internetu. V této souvislosti většinou dochází k tlaku na druhou možnost, protože v digitalizaci fondů velkých vědeckých knihoven je viděn i přínos pro rozvojové země třetího světa, které nemají finance na zajištění takovýchto zdrojů (SCHMITZ, 2008). Tyto zdroje jsou pro země třetího světa volně a zdarma dostupné a působí tak jako jakási informační pomoc.

Problematika přístupu k digitalizovaným dokumentům hraje také významnou roli v ochraně autorských práv a dohody, které upravují využití jednotlivých děl. Možným řešením jsou přístupy Open Access a Fair use (HAHN, 2006), které počítají s možnostmi veřejného vlastnictví, svobodného přístupu k informacím a využití díla chráněného autorským právem za specifických podmínek bez souhlasu držitele práv.

V některých případech je obsah z projektů masové digitalizace přístupný z několika různých stránek s odlišnými přístupovými nástroji. Existuje několik možností, jak se k tomuto obsahu dostat. První možností jsou tzv. agregátoři, kteří shromažďují obsah vzešlý z různých projektů masové digitalizace, např. Google Book Search, HathiTrust (dříve MBooks), Internet Archive, Europeana. Přístupnost je zajištěna i z vlastních OPACů knihoven zapojených do projektů, jež poskytují odkazy na plné texty, které jsou uživatelům známé prostřednictvím rozhraní OPACu (POWELL, 2007). Samotné knihovny také mohou zpřístupňovat digitalizovaný obsah přes rozhraní vlastních portálů, např. prostřednictvím národní digitální knihovny.

3.1.10 Uživatelské rozhraní

Uživatelsky přívětivé rozhraní je velmi důležitou součástí přístupnosti a vyhledávání dokumentů, případně dalších služeb a prezentace samotných dokumentů s jejich záznamy. V dnešní době hraje tento aspekt podstatnou roli pro získání obliby takového nástroje či služby.

Nejpodstatnější u uživatelského rozhraní je samotné zobrazení a prohlížení digitálních souborů. U zobrazení a prohlížení těchto souborů jsou podstatné dvě hlavní věci. Jednak standard souboru (např. PDF, DjVu, JPEG 2000 a další), ve kterém je daný dokument uložen, a software, který je použit pro jeho zobrazení, jelikož pro různé typy standardů existují různé možnosti zobrazení. Např. u standardu DjVu se používá stejnojmenný software, který napodobuje zobrazení podobné skutečnému listování v knize. Tato prezentace se může zdát uživatelům přesvědčivá z toho úhlu pohledu, že skutečně listují v digitální knize, avšak některé nevýhody se brzy projeví. Je těžké například přejít na konkrétní kapitoly nebo další funkce, jako je vyhledávání v textu či schopnost zvětšování stránky, protože nefungují, tak jak mají (COYLE, 2006).

Kromě designu a rozvržení stránek samotných je hlavní pozornost věnována tomu, jak jsou prezentovány digitalizované knihy. Možnosti jsou svázány se standardy, které jsou používány v projektech masové digitalizace (COYLE, 2006).

3.2 Výhody a nevýhody masové digitalizace

Projekty masové digitalizace mají obdobné výhody a nevýhody jako každý projekt kteréhokoli druhu digitalizace. Můžeme polemizovat o tom, zda je masová digitalizace v obecné rovině přínosem či ne. Nejde ale pominout fakt, že její projekty vycházejí pouze z požadavků dnešní doby. To s sebou samozřejmě přináší různá pro a proti. Nicméně se jedná o projekty, které mají smysl i přes jisté problémy a nevýhody, naplní-li se snaha vedoucí ke zlepšení dostupnosti informací po celém světě. Vypsání výhod i nevýhod byly vybrány na základě předchozí podkapitoly.

Výhody

- rychlost
- automatizace procesů digitalizace
- lepší dostupnost dokumentů bez ohledu na jejich fyzický počet

- kvantita skenovaného materiálu
- efektivnost procesu skenování

Nevýhody

- důraz kladený na rychlost a efektivnost více než na kvalitu
- velikost uložených dat
- obrovské finanční náklady
- nevyřešené otázky autorského práva
- duplicita digitalizovaných dokumentů mezi jednotlivými projekty

4 Projekty masové digitalizace

4.1 Rozdělení projektů masová digitalizace

V současné době existuje několik modelů, které představují různé způsoby fungování projektů masové digitalizace. Tyto projekty se dají rozdělit do dvou základních skupin. První skupina projektů je založena na vstupu soukromého kapitálu. Firma vstupující do projektu masové digitalizace za poskytnuté finanční prostředky vyžaduje získání digitálního obsahu knih a dalších druhů dokumentů, se kterými může libovolně nakládat.

Druhou možností je vznik konsorcia institucí, které poskytnou svůj fond dokumentů do jednoho velkého projektu. To zajistí přístup k již získaným zkušenostem s realizací projektu a jejich aplikaci v praxi. V tomto případě se však zapojené instituce finančně spolupodílejí na nákladech projektu.

Třetí varianta vychází z modelu tzv. národní digitální knihovny. Zde je hlavním investorem stát, který pomocí grantů a prostřednictvím příslušné instituce masově digitalizuje národní knižní produkci.

Není však vyloučeno ani zapojení knihoven do každé z těchto kategorií a různé kombinace financování.

4.2 Projekty financované soukromými subjekty

4.2.1 Google Book Search

Za začátek projektu firmy Google se považuje říjen roku 2004, kdy byl oficiálně oznámen vznik spolupráce s významnými nakladateli za účelem digitalizace jejich vybrané produkce v projektu Google Print, což byl přímý předchůdce projektu Google Book Search (LACKIE, 2008). V prosinci téhož roku Google oznámil spojení projektu Google Print s pěti vědeckými knihovnami v USA a Velké Británii (University of Michigan, Stanford University, Harvard University, Oxford University a New York Public Library), aby digitalizovali převážnou část jejich fondů v řádech milionů svazků knih a periodik (ST. CLAIR, 2006). Ke změně došlo 17. listopadu 2005, kdy byl projekt Google Print přejmenován na Google Book Search (dále jen GBS) z prostého důvodu. Uživatelé byli zmateni názvem tohoto projektu. Domnívali se totiž, že prostřednictvím této služby si mohou vytisknout dokumenty, které pomocí Googlu našli (LACKIE, 2008).

Významným krokem bylo i první rozšíření spolupráce s knihovnou z neanglicky mluvící země, Complutense University z Madridu v roce 2006 (REDMER, 2006). Při definování pojmu masové digitalizace, při níž byly odvozovány charakteristické rysy masové digitalizace od projektu GBS, bylo také často upozorňováno na fakt, že se tento projekt na začátku své činnosti převážně zaměřoval na digitalizaci literatury v anglickém jazyce (VERHEUSEN, 2008). Tímto krokem však do projektu společnosti Google přibyly i dokumenty v dalších jazycích, jelikož se do projektu GBS postupem času začali zapojovat i knihovny mimo Spojené státy. Dnes se ke spolupráci s projektem hlásí více jak 30 podobných knihoven (WILLETT, 2009).

Projekt GBS je rozdělen do dvou částí: Projekt knihovna (Library Project) a Partnerský program (Partner Program), který je zaměřen na autory a nakladatele (Google Books, c2011). V GBS Projekt knihovna jsou skenovány především knihy, ale i časopisy a noviny ze spolupracujících knihoven (WILLETT, 2009). Dokumenty jsou skenovány speciálně k tomu vyvinutými technologiemi a jejich zpracování od indexace po doplnění základních metadat probíhá automatizovaně. Masovost tohoto projektu dokazuje smlouva s University of Michigan, ve které se uvádí, že má být z fondů univerzitní knihovny během šesti let zpracováno na 7 milionů knih. Pro představu to znamená zpracovat při nepřetržitém provozu více než dvě knihy za minutu (BARTOŠEK, 2009).

Partnerský program funguje jako marketingový program pro autory a vydavatele. Autor nebo vydavatel bezplatně poskytne digitální kopii nebo výtisk své knihy k vystavení online. Za to je mu nabídnuta online propagace prostřednictvím služby Google Books. Knihy jsou zobrazeny s řadou odkazů na internetové nakladatele a obchody, ale pouze v omezeném náhledu, aby si čtenář mohl udělat představu o obsahu dokumentu (Google Books, c2011).

4.2.2 Windows Live Search Books Program

Vznik projektu Windows Live Search Books Program byl chápán jako založení protiváhy a konkurence projektu masové digitalizace společnosti Google. Dohody mezi knihovnami a projektem Live Search Books provázely velké ambice a očekávání. V plánech Microsoftu bylo vybudovat databáze knih, které by bylo možné vyhledávat na webových stránkách vyhledávače nebo v kategorii na hlavní stránce Windows Live Search.

Později Microsoft plánoval integrovat všechny naskenované knihy do svého vyhledávače Bing, s implementovanou funkcí zvanou "hledat v knize", která by uživatelům umožnila vyhledávat v plném textu knihy. Projekt Live Search Books vznikl s cílem vytvořit digitální

archiv milionů knih z předních amerických knihoven (University of California³, University of Toronto, Yale University⁴ a Cornell University⁵, The New York Public Library, and the American Museum of Veterinary Medicine), stejně tak jako více než 50 nakladatelů (ASHLING, 2007). Společně s nimi digitalizace probíhala i na britských ostrovech v Britské národní knihovně (ASHLING, 2007).

Se svými plány na vytvoření projektu Live Search Books Microsoft seznámil veřejnost v prosinci 2005 (DUKE, 2008b). Již předtím byla oznámena v listopadu téhož roku spolupráce s Britskou národní knihovnou (ASHLING, 2007), která měla přinést 100 000 naskenovaných knih (COYLE, 2006; ASHLING, 2007). Microsoft také avizoval dne 8. června 2006 spolupráci s Open Content Alliance a dvěma významnými univerzitními knihovnami University of California a University of Toronto (REDMOND, 2006).

Velkým šokem pak bylo oznámení ukončení projektu v květnu 2008 (FLOOD, 2008). Přitom ještě v březnu téhož roku Microsoft oznámil spolupráci na skenování fondů knihovny Columbia University (DUKE, 2008a) a navazoval v rámci programu Live Search Books spolupráci s řadou vydavatelů (např. Bloomsbury) ve Velké Británii, se kterými podepsal dohody o spolupráci ještě v dubnu téhož roku na knižním veletrhu v Londýně (FLOOD, 2008). Špatnou zprávou to bylo samozřejmě také pro projekt Open Content Alliance, který zčásti společnost Microsoft financoval (DUKE, 2008b). Odborníci Pan Macmillan a Pete Shemilt se shodují na tom, že ukončení projektu má negativní dopad na digitalizaci dokumentů, jelikož byla odstraněna v podstatě jediná konkurence společnosti Google, která ovládla masovou digitalizaci (FLOOD, 2008).

³ Knihovna University of California v té době obsahovala zhruba 34 milionů knih ve 100 pobočkách, což z ní činí největší akademickou knihovnu na světě. Sbírký knihovny University of Toronto, která je považována za jednu z nejlepších výzkumných knihoven v severní Americe, zahrnovaly asi 15 milionů svazků (OSWALD, 2006).

⁴ Projekt Live Search Books v knihovně Yale University se měl nejprve zaměřit na digitalizaci 100 000 autorsky nechráněných vzácných anglicky psaných knih, které pak měly být přístupné čtenářům prostřednictvím rozhraní Live Search Books, a později na výběr dalších dokumentů z celkového množství 13 milionů svazků z 22 univerzitních knihoven Yale (DUKE, 2007).

⁵ Knihovna Cornell University digitalizovala sbírky tematicky zaměřené na zemědělství, americkou historii a techniku, také uzavřela dlouhodobé partnerství se společností Microsoft s vizí, kterou měli i ostatní zúčastněné knihovny. Chtěli digitalizaci svých unikátních sbírek pro vědce a výzkumné pracovníky umožnit přístup k těmto dokumentům přes internet bez ohledu na jejich fyzické umístění. (DUKE, 2006).

Společnost Microsoft od prosince 2006 do května 2008 ve svém projektu zdigitalizovala více než 750 000 knih (CHRISTENSON, 2011) a zindexovala na 80 milionů článků z partnerských knihoven (FLOOD, 2008). Projekt byl rozdělen, obdobně jako projekt Google Book Search, na část určenou pro knihovny a část pro vydavatele. V programu pro vydavatele byly skenovány knihy chráněné autorskými právy až po souhlasu všech zainteresovaných stran, oproti tomu z knihoven byly digitalizovány pouze dokumenty, které už nebyly chráněny autorskými právy. Digitalizace produkce vydavatelů záležela čistě na dohodnutých podmínkách a získání svolení všech zainteresovaných stran (DUKE, 2008b).

4.3 Projekty konsorcií

4.3.1 Open Content Alliance

Projekt masové digitalizace Open Content Alliance (dále jen OCA), který vznikl dne 3. října 2005, představuje úsilí skupiny kulturních, technologických, nekomerčních a vládních organizací po celém světě, které napomáhají vytvoření trvalého archivu digitalizovaných textů a multimediálního obsahu v různých jazycích (Open Content Alliance, 2006). Zakladatelem projektu byl Brewster Kahle, který založil i Internet Archive (ST. CLAIR, 2006). Jedním z hlavních důvodů vzniku projektu byla reakce na digitalizační snahy firmy Google v projektu Google Book Search (dříve Google Print). Veřejné kulturní instituce a autorské svazy se totiž obávali monopolu v oblasti zpřístupnění digitalizovaných dokumentů i dostupnosti informací (WILLETT, 2009).

Mezi zakládající instituce tohoto konsorciálního projektu patří The Internet Archive, Yahoo! Inc., Adobe Systems Inc., The European Archive, HP Labs, The National Archives (UK), O'Reilly Media Inc., Prelinger Archives, The University of California a The University of Toronto (QUINT, 2005). Veřejně o založení projektu OCA informovali zakládající členové konsorcia v San Francisku v říjnu 2005. Připojila se k nim také firma Microsoft, která v roce 2008 projekt OCA opustila. Do projektu se zapojilo již více jak 80 vědeckých knihoven a archivů (WILLETT, 2009). Financování je zajištěno příspěvky jejich členů a dary od mnoha vládních, komerčních a dobročinných organizací. Každý člen projektu OCA musí souhlasit s pravidly otevřenosti tzv. Call for Participation.

Pravidla Call for Participation jsou následující:

1. OCA bude podporovat v co největší možné míře přístup a využitelnost sbírek v archivu za respektování autorských práv vlastníků a přispěvatelů
2. Přispěvatelé si budou určovat název a podmínky, pod kterými budou jejich kolekce distribuovány
3. OCA není povinna přijmout veškerý obsah, který jí je nabídnut a může dát přednost obsahu, který může být přístupný širší skupině uživatelů
4. OCA bude poskytovat metadata kolekcí v rozmanitých formátech
5. OCA vítá snahu vytvořit a nabídnout nástroje (zahrnující poskytování pomoci, katalogů a indexů), které zvýší využitelnost materiálů v archivu
6. Kopie OCA kolekcí budou uloženy mezinárodně v rozmanitých archivech pro zajištění jejich dlouhodobého uchování a dostupnosti všem (Open Content Alliance, 2006).

OCA je rozdělen do následujících šesti pracovních skupin: Metadata and collaborative collection development working group (Pracovní skupina pro metadata a pro rozvoj sbírkové spolupráce), Digital preservation Working Group (Pracovní skupina ochrany digitálních dat), Contribution Working Group (Pracovní skupina příspěvků), Book Format Working Group (Pracovní skupina knižních formátů), Scanning Protocol a Data Transfer Protocol (Skenovací protokol a protokol přenosu dat) (CRAWFORD, 2006).

Internet Archive pro projekt OCA digitalizuje data od přispěvatelů a také dodává do archivu OCA materiál z vlastních sbírek. Společnost Yahoo! zpřístupňuje obsah OCA přes svůj vyhledávač a působí jako správce vyhledávacího stroje webových stránek OCA (Open Content Alliance, 2006). Obsah archivu OCA je zdarma dostupný i dalším vyhledávačům, pro které je volně dostupný k vytváření jejich indexů. Projekt OCA si zakládá na tom, že digitalizuje pouze dokumenty, které už nejsou chráněny autorským zákonem nebo ke kterým jeho realizátoři získali povolení autora. Dokumenty jsou hlavně zpřístupněny prostřednictvím webových stránek Internet Archive a OpenLibrary, která obsahuje přes 23 milionů záznamů s přibližně 1,6 milionem knih s plným textem a je používána jako další přístupové místo pro knihy z projektu OCA společně s dalšími digitalizovanými knižními sbírkami (Open Content Alliance, 2006).

4.3.2 The Million Book Project

Dva předcházející projekty nebyly první tohoto druhu. Jejich předchůdcem byl The Million Book Project. U jména projektu často dochází k záměnám, proto je The Million Book Project uváděn pod jmény Universal Digital Library nebo Universal Library. Nejedná se ale o ekvivalentní názvy, jelikož Universal Digital Library je zastřešující název konsorcia pro digitalizační hnutí, ale The Million Book Project se jmenuje pouze projekt zaměřený na digitalizaci knih (ST. CLAIR, 2008). Tento projekt byl jako první označen za projekt masové digitalizace. Je také nutné zmínit, že projekt masové digitalizace The Million Book Project vznikl původně jako projekt „large scale“ digitalizace, avšak postupem času se přetransformoval na projekt masové digitalizace. Stalo se tak připojením k společnosti the Internet Archive, která se účastní i projektu Open Content Alliance.

Projekt založila National Science Foundation na začátku roku 2000 (ST. CLAIR, 2008), ale projekt přešel pod vedení Carnegie Mellon University v roce 2001 (WILLETT, 2009). O financování se starala americká instituce The National Science Foundation s vládní podporou Indie, Číny, Internet Archive a The University of California Libraries v Merced, která pracovala na zajištění autorských práv. V tomto projektu bylo digitalizováno přes 1,5 milionu svazků včetně velkého množství dokumentů z převážně universitních knihoven v Číně, Indii, Egyptě⁶ a dalších partnerů viz. Příloha B.1 (ST. CLAIR, 2008). Jádrem tvoří dokumenty v čínštině, kterých je skoro jeden milion, a dokumenty v angličtině, kterých je přes 350 000 (WILLETT, 2009). Přístup k digitalizovaným dokumentům je zprostředkován zdarma pomocí stránek Internet Archive nebo vlastních stránek The Universal Digital Library. Také jsou dostupné na serverech zúčastněných institucí v Indii⁷, Číně⁸ a Mellon Carnegie University. Tento projekt nebyl v roce 2007 prodloužen a v současné době nejsou žádné informace o jeho znovuoobnovení (WILLETT, 2009). Na začátku roku 2007 se jednalo o největší projekt digitalizace, jelikož obsahoval na 1,4 milionů svazků. Oproti tomu Google Book Search měl na kontě 1 milion svazků a Open Content Alliance jen okolo 100 000 svazků (ST. CLAIR, 2008).

⁶ Hlavně Alexandrijská knihovna.

⁷ <http://www.ernet.dli.in>

⁸ <http://www.ulib.org.cn>

The Million Book Project si kladl za cíl digitalizovat všechny významné literární, umělecké a vědecké práce na světě. Autoři projektu tím chtěli odstranit nevýhodu dostupnosti dnešních knihoven a přístupnosti vzácných dokumentů (Universal Digital Library, c2001-2008). V projektu byly přijaty dvě hlavní strategie vytváření sbírek podle obsahu dokumentů. První se zaměřila na nejlepší knihy podle „Books for College Libraries“. Druhá část se pak zabývala dokumenty, které tematicky zahrnovali zemědělství.

4.4 Situace mimo Evropu

Mimo Evropu a Spojené státy americké se digitalizační aktivity soustřeďují převážně do ekonomicky vyspělých zemí jako je Japonsko, Čína, Jižní Korea, Austrálie, Nový Zéland nebo Indie. Všechny projekty asijských zemí se dají pokládat za masovou digitalizaci, neboť šíře a objem jejich aktivit v tomto směru přesahuje evropské i americké parametry. U Japonska a Jižní Koreje nejsou tyto snahy náhodné, neboť jde o vysoce informační společnosti, které maximální možnou měrou využívají informační technologie a snaží se zachovat své kulturní dědictví. Čína se snaží dostat na jejich úroveň a investuje nemalé prostředky do tohoto odvětví. Všechny zmíněné země využívají nejmodernějších technologií pro digitalizaci svého kulturního dědictví. Nejvýznamnější projekty masové digitalizace v této části světa se odehrávaly v projektu The Million Book Project v Číně a Indii (Universal Digital Library, c2001-2008) a v projektu Digital Archiving Project, kde se spojují významné národní digitální knihovny tří zemí (Jižní Koreje, Číny a Japonska) pod vedením National Diet Library (NATIONAL DIET LIBRARY, 2008).

Kvalitní digitalizační projekty mají i v Austrálii a na Novém Zélandu. Jedná se o projekty tzv. „large-scale digitalizace. Australské projekty zastupuje Digital collections a PANDORA – Australia's Web Archive Australské národní knihovny (The National Library of Australia). Na Novém Zélandu to je především The New Zealand Digital Library University of Waikato a digitální knihovna Národní knihovny Nového Zélandu (New Zealand Digital Library).

4.4.1 Japonsko

Práce na japonské digitální knihovně (Electronic Library) v National Diet Library⁹ byla započata již v roce 1998 pod názvem National Diet Library Electronic Library Concept (TAYA, 2010). Budování této digitální knihovny se zakládalo na kooperaci významných japonských knihoven, jež tvoří NDL (Tokyo Main Library, Kansai-kan of the National Diet Library a International Library of Children's Literature) a další kulturní instituce, jako jsou archivy a muzea. Zahájení provozu pro veřejnost bylo uskutečněno až v říjnu 2007. Po roce 2009 byla zahájena i spolupráce v mezinárodním měřítku. Do kooperace byly zapojeny země východní Asie, především Čína a Jižní Korea. Ve stejném roce National Diet Library obdržela dodatečné finanční prostředky, cca. 138 711 000 dolarů k vytvoření projektu masové digitalizace. Tyto finanční prostředky odpovídají více než 50 % ročního rozpočtu knihovny ve výši 234 928 000 dolarů (VELDE, 2011). Projekt masové digitalizace byl podpořen, protože se National Diet Library rozhodla změnit způsob dlouhodobého uchovávání. Přejít z mikrofilmu na digitalizaci v březnu 2009 a změny autorského práva s účinností od ledna 2010 umožnily National Diet Library digitalizaci svých fondů za účelem dlouhodobého uchování bez povolení držitelů práv. Poskytnutí vládních finančních prostředků na digitalizaci byla také zčásti reakce na projekt Google Book Search¹⁰. V roce 2010 obsahovala digitální knihovna na 2,3 milionu knih a další čtvrtina fondu National Diet Library měla být digitalizována v roce 2011 (TAYA, 2010). Celý projekt, který umožnil digitalizaci přibližně jednoho milionu položek, byl dokončen v březnu 2011. Japonská digitální knihovna obsahuje nyní téměř všechny knihy vydané v Japonsku před rokem 1969, které jsou ve fondech National Diet Library, 11 000 titulů časopisů, 73 000 vzácných a starých knih (VELDE, 2011). Významnou částí této digitální knihovny je Rare Books Image obsahující staré japonské a čínské knihy, mapy, obrazy a další cenné dokumenty. Mezi další zajímavé části této digitální knihovny se řadí i Digital Library from the Meiji Era, Picture Book Gallery a Digital Library of Children's Literature (TAYA, 2010). Digitalizované dokumenty jsou volně dostupné přes internet, kromě děl chráněných autorským zákonem. Ty jsou dostupné pouze v knihovnách (NATIONAL DIET LIBRARY, 2012).

⁹ Národní knihovna Japonska.

¹⁰ Do projektu masové digitalizace společnosti Google je v Japonsku zapojena knihovna univerzity Keio, která je jediná v tomto směru z celé Asie (Google Books, c2011).

4.4.2 Jižní Korea

Jižní Korea patří mezi průkopníky v digitalizaci a její digitalizační projekty patří mezi nejpokročilejší. National Digital Library je složena z fondů hlavních knihoven v zemi. Jde o The National Libray of Korea, The National Assembly Library, Supreme Court Library, Korea Institute of Science and Technology Information, Korea Education and Research Information Service a Korea Advanced Institute of Science and Technology Science Library. Jádrem této digitální knihovny je tvořeno starými tisky, novinami vydanými před rokem 1945, publikacemi vydávanými Ministerstvem kultury a turismu, disertačními pracemi z humanitních věd a mnoha dalšími dokumenty (CHOI, 2003). V roce 2011 bylo uvedeno, že tato digitální knihovna obsahuje již více než 116 milionů digitálních dokumentů, nicméně k tomuto číslu byla vyslovena pochybnost, že se jedná spíše o počet zdigitalizovaných stran (FRENCH, 2011).

Uchovávání digitálních kopií originálních sbírkových materiálů se ujala The National Libray of Korea od roku 1998. V roce 2000 knihovna zahájila masovou digitalizaci v reakci na vládní informační (znalostní) politiku, která plánovala digitalizovat po dobu dvou let asi 140.000 svazků za cca 1.010.000 dolarů. Do tohoto projektu se současně zapojily veřejné knihovny. The National Libray of Korea masově digitalizovala do roku 2009 na 110 000 000 stránek z 390 000 publikací. Digitalizace je omezena na díla bez autorsko-právní ochrany. Na základě licence omezeného přístupu bylo digitalizováno 150 000 položek s ochranou autorských práv. V roce 2009 The National Libray of Korea zahájila práce na tvorbě digitálního portálu. Projekt digitální knihovny měl zlepšit regionální rozvoj, podporu zaměstnanosti, snížit digitální propast, oživit a posílit všechny knihovnické služby. K dnešnímu dni byla provedena jedna hodnotící studie, která měřila ekonomickou hodnotu digitalizovaných materiálů¹¹. Podle této studie jsou digitalizované texty ceněny pro jejich výběr a pohodlí přístupu. Uživatelé v Jižní Koreji jsou připraveni zaplatit až cca 0,81 \$ pro přístup k digitalizovaným jihokorejským monografiím (VELDE, 2011).

Zajímavá je také digitální knihovna The National Assembly Library, založená Parlamentem Jižní Koreji v roce 1953. Ta je součástí The National Library of Korea. Obsahuje přes 600 000

¹¹ Ryu, Hee-Kyeong, Lee, Doo Young. A study on the economic value measurement of domestic monograph full-text information services. *Journal of the Korean Society for Information Management*. 2006, vol. 23, no. 4, s. 111-128. ISSN 1013-0799.

digitálních položek, které nejsou všechny plně dostupné přes internet. Uživatelé mohou dostat přístup k plno-textovým dokumentům po přihlášení nebo jsou autorizováni jako uživatelé akademických knihoven a institucí. Příklady digitální sbírek jsou: Social Sciences Doctoral and Masters Theses (cca. 300 000 položek) nebo Government Publications (cca. 10 000 položek) (CHOI, 2003).

4.4.3 Čína

V Číně je digitalizace velkou prioritou nejen v knihovnách, ale i v politice. Financování těchto projektů jde přes National Social Science Foundation, což je státní instituce přerozdělující finance. Malými částkami také přispívají vlády provincií, institucí a sesterských knihoven. V roce 1997 přišlo Ministerstvo kultury a jeho Státní plánovací komise s projektem National Pilot Digital Library Project, což byl první projekt tohoto druhu v Číně. Spolupracovaly zde Národní knihovna Číny a pět velkých veřejných knihoven. Od roku 1998 se pak k tomuto projektu připojovaly další instituce např. Ministerstvo školství a další vědecké a technické instituce (LIU, 2004). Výstupem tohoto projektu je Čínská národní digitální knihovna, která obsahovala na konci roku 2009 okolo 7,5 milionu digitálních dokumentů (DAWEI, YIGANG, 2010).

Mezi další významné digitální knihovny patří The China National Knowledge Infrastructure, která obsahuje sedm plno-textových databází rozdělených dle druhů dokumentů na noviny, časopisy, monografie, disertace, konferenční sborníky a dvě specializované databáze na medicínu a vzdělávání. Hlavní programy digitalizace v Číně jsou The China Digital Library Program a Chinese National Science Digital Library, který má pomoci ke zlepšení podpory výzkumu a znalostních služeb. Projekt buduje také vlastní databázi čínské vědecké literatury z různých oblastí vědy např. biologie, fyzika, matematika, geografie, počítačová věda a mnoho dalších (LIU, 2004). Do roku 2009 čínská vláda rozhodla o digitalizaci 1,4 milionu knih za 9 milionů dolarů. V digitalizaci pokračovaly i nadále s dalšími 3 miliony knih, které byly skenovány od roku 2009 (Open Content Alliance, 2006). Pro srovnání, indická vláda financovala v rámci indických digitalizačních projektů naskenování 600 000 - 1 milion knih. Neposkytla však žádné informace o tom, jaké byly celkové náklady (Open Content Alliance, 2006).

4.5 Situace v Evropě

Digitalizační projekty v Evropě dnes obecně řeší dvě zásadní otázky. Jedná se především o nutnou změnu autorského práva a zajištění a změnu způsobu financování digitalizace. V roce 2008 byla zdigitalizována pouhá 2-3 procenta fondů národních knihoven v Evropské unii. Ozývají se hlasy (AYRIS, 2010) volající po rychlém řešení a celkové reformě autorských práv. Pro projekty masové digitalizace je především důležitá úprava podmínek používání osiřelých děl a části práv, které se dotýkají knihoven. Stávající autorská práva v současné době činí evropské digitalizační projekty oproti projektům amerických knihoven do značné míry neatraktivní a komplikované.

V současné době se ve státech Evropské unie pracuje na právním rámci umožňujícím lepší přístup k osiřelým dílům. Podobné předpisy už přijali některé evropské země. První byla Francie. Knihovnám tato právní úprava umožňuje digitalizovat knihy s právní ochranou. Německo o takové právní úpravě uvažuje. Právo v Japonsku a Norsku povoluje národní knihovně digitalizovat v projektech i díla, která jsou chráněná autorským právem (SAMUELSON, 2012).

Hlavním důvodem jsou knihovní fondy evropských knihoven, které jsou tvořeny z 90 % knihami bez komerčního využití. Jedná se hlavně o tzv. osiřelá díla nebo knihy bez autorské ochrany. Z tohoto důvodu evropské digitalizační projekty nenabízí atraktivní obsah, který je nabízen na druhé straně Atlantiku (AYRIS, 2010).

Nevyhnutelná změna také čeká způsob financování evropských digitalizačních projektů, jelikož Evropská unie a její fondy ani veřejné zdroje nepokryjí náklady, které jsou pro požadovanou rychlost digitalizace takového množství dokumentů nutné. Řešením, které by pomohlo urychlit současný nevyhovující stav, je spolupráce veřejného a soukromého sektoru, tzv. public-private partnership (AYRIS, 2010). V Evropě dnes existuje taková spolupráce u několika projektů.

- Britská knihovna – Cengage Gale (technické a finanční partnerství se získáváním příjmů z kulturních institucí)
- El Prado – Google (obsah zpřístupněný přes Google Books)
- Francouzská národní knihovna (BnF) – Francouzští nakladatelé (SNE) (společné vyhledávání a přesměrování odkazů na možnost zakoupení textů)

- Norská národní knihovna – asociace držitelů práv Kopinor (zpřístupňují obsah chráněný autorskými právy; je zde mnoho sporných otázek spojených s teritorialitou)
- Dánská královská knihovna – ProQuest (staré tisky jsou volně dostupné pro výzkum a vzdělání v Dánsku)
- Google Book Library partnership (cca. 20 knihoven po celém světě) - zlomek obsahu chráněného autorskými právy; 10 milionů knih v Google Book Search (AYRIS, 2010).

Tohoto stavu využívá hlavně společnost Google, která poskytuje finance digitalizačním projektům v různých zemích a za to získává obsah pro svůj projekt Google Book Search. V současné době je uzavřeno několik partnerství mezi Googlem a jednotlivými evropskými knihovnami. Mezi knihovny, které uzavřely spojení s Google patří např. Bavorská státní knihovna v Německu¹², Oxfordské univerzitní knihovny, Knihovna Gentské univerzity, Univerzitní knihovna v Lausanne, Katalánská národní knihovna, Městská knihovna v Lyonu, Univerzita Complutense v Madridu. Další evropské knihovny zatím ještě nejsou zveřejněny na stránkách Google Books, ale svou spolupráci již potvrdily obě strany. Jsou to Národní knihovna Nizozemska (Koninklijke Bibliotheek) (COLOMBET, 2010), Národní knihovna ČR (NÁRODNÍ KNIHOVNA ČR, 2011), Rakouská národní knihovna, italské Ministerstvo kultury (WEISL, 2010), Britská knihovna (THE BRITISH LIBRARY, 2011) a Dánská královská knihovna¹³ (SRINIVASAN, 2009).

V následujících podkapitolách budou uvedeny příklady dvou členských států, které vlastními prostředky financují projekty národních digitálních knihoven, které jsou označovány za masovou digitalizaci. Jde o digitalizační projekty Francie a Norska.

4.5.1 Francie

Francouzská vláda finančně podporuje masovou digitalizaci prostřednictvím Conseil National du Livre (Národní knižní centrum), tím se Francie brání projektu společnosti Google. Francie

¹² Např. projekty v Německu spadají pod Mnichovské digitální centrum. Bavorská státní knihovna však dala přednost spojení s firmou Google, se kterou se dohodla na spolupráci z důvodu pomalého a zatím neupřesněného postupu iniciativ na německé centrální i státní úrovni, které na projektech masového charakteru, jako je Německé digitální knihovny popřípadě Evropská digitální knihovna (Europeana), pracují. Spolupráce byla iniciována z jednoduchého důvodu, jelikož Google umí skenovat v průmyslovém měřítku, které zvládne zajistit jak technicky, tak i logisticky. Samotná selekce dokumentů se odehrává podle vhodného stavu dokumentů pro skenování a podle požadavků na formát pro skenovací technologii (CEYNOWA, 2008).

¹³ Zatím nepotvrzeno Googlem.

se snaží za pomoci vlastních prostředků vybudovat digitální knihovnu jako ochranu svého kulturního bohatství. Bibliothèque National de France (BNF) na projekt Gallica¹⁴ získala státní dotaci ve výši zhruba 750 milionů € (DELORME, 2011). Tato finanční podpora umožnila například do konce roku 2009 zdigitalizovat a online zpřístupnit 973 000¹⁵ dokumentů. Během tří let bylo zdigitalizováno na 250 000 dokumentů. Dalších 17 250 komerčně publikovaných dokumentů s ochranou autorského zákona bylo digitalizováno ve spolupráci s Asociací francouzských vydavatelů (Syndicat national de l'édition), Národním knižním centrem (Conseil National du Livre) a Ministerstvem kultury a komunikace (Ministère de la Culture et de la Communication) (VELDE, 2011). Digitalizační politika BNF se zaměřuje na tři hlavní oblasti: národní dědictví, mezinárodní programy a evropská díla. V roce 2011 BNF oznámila plány na příštích pět let, které počítají se zpřístupněním dalších 500 000 knih. (DELORME, 2011). Mezi ně budou patřit i knihy komerčně dostupné¹⁶, které poskytnou nový zdroj příjmů pro vydavatele a autory, a umožní BNF zachovat jejich digitální kopie (VELDE, 2011).

Digitalizace sbírek BNF je prováděna s rostoucím počtem regionálních partnerů, mezinárodních institucí a organizací, včetně Library of Congress, Mellon Foundation nebo společnosti Google. BNF se rovněž podílí na vývoji digitálního systému pro knihovny, který bude obsahovat dokumentaci a podněty k zajištění toho, aby byla digitalizace koordinována a centralizována (VELDE, 2011).

V roce 2011 BNF oznámila partnerství s poskytovateli¹⁷ obsahu a technologií ve výši 11,2 milionu eur na digitalizaci tištěných sbírek BNF a partnerských knihoven. Očekává se, že bude ročně po dobu tří let převedeno do digitální podoby kolem 70 000 děl (DELORME, 2011).

¹⁴ Projekt Gallica je hlavním přispěvatelem digitalizovaného obsahu do projektu Europeana.

¹⁵ Do února 2010 se počet digitálních dokumentů zvýšil na 1 milion.

¹⁶ Je to možné díky nové dohodě týkající se francouzských knih vytištěných ve 20. století. Záznamy těchto knih budou dostupné čtenáři s výňatky textu a budou obsahovat odkazy na internetové prodejce, kde si uživatel bude moci koupit digitální kopii. Dohodu podepsali Frédéric Mitterrand (Ministère de la Culture et de la Communication) Bruno Racine (prezident BnF), Antoine Gallimard (prezident Syndicat national de l'édition) a Jean-Claude Bologne (prezident Société des Gens de Lettres).

¹⁷ Jedná se o firmy Jouve-Safig-Diadeis a BancTec. BNF navazuje na dřívější spolupráci, která vedla k digitalizaci 36,6 milionu stránek v roce 2007.

4.5.2 Norsko

Kulturní instituce v Norsku se považují za nedílnou součást vládní informační a komunikační politiky. Ta klade důraz na zrychlení procesu digitalizace, aby norské kulturní dědictví bylo zachováno, což je považováno za národní odpovědnost pro příští generace. Díky této podpoře ze strany státu patří Norsko k předním představitelům digitalizace kulturního dědictví v Evropě.

V letech 2005 a 2006 vydala Nasjonalbiblioteket (dále Národní knihovna Norska) sérii strategických dokumentů pro podporu rozvoje Národní knihovny digitální (NBdigital). Tento proces zahrnoval systematickou digitalizaci stávajících fondů knihovny, zavedení dohod pro usnadnění přístupu k digitálním materiálům přes digitální úložiště a sklizení digitálního obsahu z webových domén. The Digital National Library - Strategy Manifesto 2005, obsah tohoto materiálu se snažil stanovit přínosy (vzdělávací, ekonomické a politické) digitalizace pro norskou společnost. Druhý strategický dokument, Access to Digital Content in the National Library: Strategy Document for the Digitisation of the National Library's Collection (Přístup k digitálnímu obsahu v Národní knihovně: Strategický dokument pro digitalizaci sbírek Národní knihovny), zavedl komplexní digitalizační program. Program byl upraven ve strategii Digitisation of the National Library's Collection, vydané v roce 2006, která ustanovila požadavky na provedení projektu a na priority pro proces digitalizace.

Cílem projektu je zlepšit přístup ke kulturnímu dědictví, a to jak k historickým materiálům, tak i moderní části kulturního dědictví (NATIONAL LIBRARY OF NORWAY, 2007). K Norské národní knihovně se do projektu připojily ostatní knihovny a kulturní instituce, hlavně univerzitní knihovny a Norwegian national agency for ALM (Norská národní agentura pro archivy, knihovny a muzea) (SKARSTEIN, 2010).

První partnerství vzniklo mezi Národní knihovnou Norska a Norským rozhlasem (NRK) na převedení a uchování archivu rozhlasového vysílání. Následně se knihovna dohodla se třemi významnými deníky (včetně Aftenposten jeden z největších norských deníků) na jejich digitalizaci.

Národní knihovna Norska má v současnosti 16 digitálních linek pro konverzi různých formátů do digitálních kopií. Pro zpřístupnění je nutná dohoda s držiteli autorských práv reprezentovaných institucí Kopinor a na základě Norwegian Copyright Act, podle kterého má knihovna právo digitalizovat své sbírky pro jejich ochranu (SKARSTEIN, 2010). V roce 2011

digitální knihovna obsahovala více než 170 000 knih, z toho asi 50 000 knih chráněných autorským právem, 250 000 novin, 50 000 rukopisů, 610 000 hodin rozhlasového vysílání a 500 000 fotografií, které zabírají 1.900 terabajtů úložného prostoru. Digitální obsah se zvyšuje v průměru o několik terabytů za den (VELDE, 2011). Měsíční přírůstek tvoří 2 000 – 3 000 knih (SKARSTEIN, 2010). Přibližně 30% obsahu tohoto digitálního úložiště představují materiály, které byly vytvořeny v digitální podobě, zbytek je výsledkem digitalizace dokumentů (VELDE, 2011).

Strategie Národní knihovny Norska umožňuje přístup k velkému množství norského kulturního dědictví v digitální podobě. Dokumenty v projektu jsou vybírány podle tří kritérií: první prioritou je digitalizovat všechny veřejně dostupné materiály, druhá priorita je tematická a zahrnující všechny typy médií, třetí priorita odráží požadavky uživatelů a dalších institucí (TAKLE, 2009).

Celkové náklady projektu byly odhadovány na 1 miliardu norských korun (NATIONAL LIBRARY OF NORWAY, 2007). Tyto digitalizační aktivity jsou financovány z dodatečných příjmů a úspor Národní knihovny Norska. Část financí na digitalizaci přinesla vnitřní restrukturalizace, zvýšení rozpočtů a strategické partnerství s vydavateli. Výdaje na digitalizaci se zvýšily z přibližně 3 240 000 dolarů v roce 2006, přes 2 790 000 dolarů v roce 2007 a 2008, na 4 320 000 dolarů v roce 2009. Ministerstvo kultury a církevních záležitostí přispělo v roce 2007 a 2008 přibližně částkou 430 000 dolarů, která byla v roce 2009 zvýšena na přibližně 2 miliony dolarů (TAKLE, 2009).

Národní knihovna Norska zavedla bezpečné digitální repozitáře k uchovávání a šíření děl a k vytvoření licenčního rámce pro usnadnění vkládání a přístupu k dílům chráněných autorským právem. Tyto doplňující strategie umožňují přístup k norskému kulturnímu dědictví přes digitální úložiště (VELDE, 2011).

Norská digitální knihovna umožňuje přístup přes vlastní vyhledávač FAST, který je použit na webových stránkách Bokshylla (The Bookshelf) a na stránkách Norské národní knihovny, do katalogu napojeného na digitální repozitář. Vyhledávat lze podle jména autora, názvu, klíčových slov, slov obsažených v textu nebo citací přes jedno dotazovací pole (SKARSTEIN, 2010). Vyhledávač neumožňuje pokročilé vyhledávání. Záznam dokumentu obsahuje základní metadata a URL na stránku, kde si uživatel může celý dokument prohlédnout. Jednotlivé záznamy jsou barevně označeny. Barvy značí, zda obsah je či není digitalizovaný,

jedná-li se o webové stránky. Nechybí zde funkce, jako jsou úprava vyhledávacího dotazu pomocí faset, možnost personalizace po přihlášení, funkcionalita Webu 2.0. Samozřejmostí je i příjemné uživatelské rozhraní.

4.5.3 Situace v České republice

Projekt „Vytvoření Národní digitální knihovny“ je v České republice teprve na začátku. Zajišťují jej Národní knihovna ČR společně s Moravskou zemskou knihovnou. Žádost na schválení projektu byla podána v únoru 2010. V červnu téhož roku byl tento projekt schválen. O financování projektu se z 85% stará EU z Integrovaného operačního programu. Zbylé prostředky jsou zajištěny ze státního rozpočtu prostřednictvím Ministerstva kultury ČR (Národní digitální knihovna, 2010).

Spolupráce Národní knihovny ČR a Moravské zemské knihovny na projektu Národní digitální knihovny, který má za hlavní cíl digitalizovat převážnou část bohemikální produkce 19. a 21. století a shromažďovat dokumenty primárně publikované v digitální podobě tzv. born digital, vychází z jednoduchého předpokladu. Instrukce zajišťující tento projekt vlastní nejbohatší fondy v České republice, jelikož obě instituce tvoří svůj fond na základě úplného práva povinného výtisku a uchovávají tak většinu monografií, periodik a dalších druhů dokumentů publikovaných na našem území. Jejich fondy obsahují také významný podíl tzv. zahraničních bohemik. Součástí projektu je kromě digitalizace i podchycení digitálních dokumentů českého webu, tedy dokumentů primárně publikovaných v digitální podobě tzv. born digital. Spojením těchto institucí pro realizaci projektu NDK vede k maximální úplnosti sbírky bohemik jako digitálních dokumentů. Odhadované množství dokumentů určených k digitalizaci je cca 1,2 milionu, což pro představu tvoří 310 milionů stran (NÁRODNÍ KNIHOVNA ČR, 2010).

V souvislosti s budoucím zapojením do budování Národní digitální knihovny a spoluprací se společností Google v rámci projektu „Google pro českou kulturu 2011“, jejímž vstupem by se měl celý proces digitalizace urychlit, se hovoří především o digitalizaci vybraných starších knižních fondů z 16. – 18. století. Celkem se jedná o 200 000 historických dokumentů. Hlavní zájem společnosti Google je o Slovanskou knihovnu NK ČR a Historické a hudební fondy NK ČR (NÁRODNÍ KNIHOVNA ČR, 2011).

Fond Slovanské knihovny je převážně tvořen dokumenty v cizích jazycích, které nejsou v zemích původu dostupné. Cílem digitalizace tohoto fondu je volně zpřístupnit dokumenty

pro převážně zahraniční badatele a virtuálně dokumenty vrátit do zemí svého původu. U sbírek Historického a hudebního fondu se jedná především o díla významných evropských autorů, jako je např. Erasmus Rotterdamský, Jan Ámos Komenský, Tycho de Brahe, Johannes Keller, Martin Luther a mnoho dalších (NÁRODNÍ KNIHOVNA ČR, 2011).

4.6 Další projekty

Mezi dalšími zajímavými projekty masové digitalizace můžeme zmínit např. projekt Amazonu Search Inside the Book capability, který je doplňkem služeb pro jejich knihkupectví, kdy si čtenář může prohlédnout knihu předtím, než si jí koupí přes internet (COYLE, 2006).

5 Srovnání vybraných projektů masové digitalizace

5.1 Kritéria výběru projektů masové digitalizace

Následující kapitola se zabývá výběrem projektů masové digitalizace, které budou v další části práce srovnány. Kritéria výběru projektů byla zvolena na základě rozdělení projektů masové digitalizace v předchozí kapitole (rozdělení projektů podle jejich způsobu financování). Z každé z těchto skupin byl vybrán pro účel porovnání jeden projekt, který splňoval další doplňková kritéria. Kritéria byla následující: musí se jednat o aktivní projekty, patří ve své kategorii mezi největší, a to jak počtem zapojených institucí, tak i objemem skenovaných dokumentů, projekty jsou založeny na mezinárodní spolupráci, ke které dochází napříč kontinenty.

Tato kritéria splnily dva projekty masové digitalizace, které jsou hlavními představiteli daného rozdělení projektů Google Book Search a Open Content Alliance. S těmito projekty je rámcově srovnán jeden zástupce NDK, který splňuje podmínky masové digitalizace. Je jím norská NDK. Srovnávání projektů je provedeno na základě kvantitativních a kvalitativních hledisek nejen mezi sebou. Pro porovnání projektů byla vybrána jako kvalitativní kritéria aspekty masové digitalizace na základě 2. kapitoly. Získané informace budou okomentovány a nejdůležitější fakta budou převedena pro přehlednost do Tabulky č. 1.

	Google Book Search	Open Content Alliance	Norská digitální knihovna
Cíle projektů	Rozšíření indexů vyhledávače Google	Vytvořit konkurenci projektu Google a zlepšit přístup k dokumentům	Vytvořit národní digitální knihovnu jako jádro digitální knihovny celého Norska. Zlepšit přístup ke kulturnímu dědictví, a to k historickým materiálům, ale i k moderní části kulturního dědictví
Knihovny a další kulturní instituce	26 knihoven v partnerském programu	47 knihoven a dalších kulturních institucí	Norská národní knihovna
Vznik	Říjen 2004	3. října 2005	2005
Finance	Soukromý investor Google financuje projekt výměnou za digitální kopie vzniklé v projektu	Konsorcium placené z příspěvků od sponzorů, dárců a členů konsorcia	Financování probíhá z Národního knihovnického digitalizačního programu (National Library's digitization programme), který získává prostředky ze státní pokladny
Náklady	35 milionů naskenovaných knih v roce 2010, odhady hovoří o nákladech okolo 1 mld. dolarů, to je od 30 do 100 \$ za knihu	Cena 10 centů za stránku nebo 30 dolarů za knihu. Jsou do toho započteny i náklady na technologie spojené s dlouhodobým uložením atd.	Celkové náklady jsou odhadovány na 1 bilion (miliardu) norských korun. Předpokládána doba trvání 15-20 let

Tabulka č. 1

	Google Book Search	Open Content Alliance	Norská digitální knihovna
Autorská práva	Problémy s autorským právem v Evropě i v USA, vyřešení situace Google Book Settlement Agreement, skenování všech dokumentů bez ohledu na jejich autorská práva	Dokumenty nechráněné AZ, tzv. orphan works v USA, digitalizovány jsou i dokumenty, které jsou chráněny AZ, ale se souhlasem od držitele těchto práv	Knihovna má právo digitalizovat své sbírky pro jejich ochranu. Pro zpřístupnění je nutná dohoda s držiteli autorských práv reprezentovaných Kopinor na základě Norwegian Copyright Act a Legal Deposit Act
Výběr materiálů k digitalizaci	Ovlivněn autorským právem a jejich fyzickým stavem, který umožňuje skenování		Všechny dokumenty knihovnou vlastněné
Technologie	Nejsou dostupné informace	Skenování: Scribe systém OCR technologie: ABBYY FineReader 8.0	OCR software docWorks
Standardy souborů	PDF, ePub, prostý text	PDF, DjVu, ePub, prostý text, Mobi pro čtečku Kindle, Daisy formát, „online čtení“	PDF, i obrazové soubory JPEG
Rozlišení při skenování	300 dpi - 600 dpi nativní obrázky nebo interpolace založená na nižším rozlišení	400–600 dpi, 12 bitů	400 dpi, 24 bitů
Formát souboru pro uložení	JPEG2000	JPEG2000	JPEG2000

Tabulka č. 1

	Google Book Search	Open Content Alliance	Norská digitální knihovna
Obsah	Knihy jak od nakladatelů, tak z knihoven, ale bez speciálních sbírek knihoven, které jsou také Googlem digitalizovány	Široký okruh beletrie, populárních knih, dětských knih, historických textů a textů z akademické sféry	Všechny sbírky Národní knihovny Norska, malá část fondu není chráněna copyrightem (cca. 5 000 titulů) ze 450 000 titulů
Přístup k digitalizovaným dokumentům	Vlastní portál Google Books nebo repozitáře a digitální knihovny spojených institucí	Prostřednictvím Open Library a Internet Archive	Vlastní vyhledávač od firmy FAST
Uživatelské rozhraní	Jednoduché a přehledné rozhraní	IA – rozhraní není moc přehledné a přívětivé OpenLibrary - jednoduché a přehledné rozhraní	Přístup přes OPAC The Bookshelf a přes BokHylla.no,
Vyhledávání	Základní i pokročilé, klasické vyhledávací pole	Základní i pokročilé, klasické vyhledávací pole v Internet Archive i v OpenLibrary, kde je oddělené vyhledávání v plných textech knih	Jedno vyhledávací okno, lze vyhledávat podle autora, názvu, klíčového slova, v plném textu a podle citace, nemá pokročilé vyhledávání
Obsažená média	Knihy, časopisy, noviny	Knihy, časopisy, noviny, audio, video	Knihy, periodika, noviny, fotky, mapy, atd.

Tabulka č. 1

5.2 Srovnání celosvětových projektů masové digitalizace

5.2.1 Cíle projektů

Zapojením a založením vlastního projektu masové digitalizace společnost Google sledovala své vlastní zájmy, jelikož se snažila o rozšíření svých indexů, které používá její vyhledávač. Obsah knih a dalších tištěných dokumentů a v nich zaznamenané znalosti a informace znamenaly velké lákadlo, proč se pustit do takového projektu. Oproti Googlu je OCA „otevřený“ projekt, který se dělí o informace, které získal v průběhu existování projektu. Jedním z hlavních důvodů byla reakce na projekt GBS. Mezi společné cíle obou projektů se může uvést i snaha o zlepšení dostupnosti dokumentů a informací v knihách obsažených pro třetí svět.

5.2.2 Počet partnerů

Partneři projektů ze soukromé sféry

Zapojení soukromých subjektů do projektů masové digitalizace je nutné a v mnoha případech i klíčové. Projekt GBS založila firma Google jako projekt, který je založen na spolupráci veřejného a soukromého sektoru. Širší zapojení soukromých firem funguje u projektu OCA. Mimo zakladatelské instituce Internet Archive se připojili do projektu OCA internetový giganti Yahoo! a Microsoft (ukončení spolupráce v roce 2008) jako součást konkurenčního boje s Googlem, a další firmy HP Labs, Adobe Systems Incorporated, O'Reilly Media, Xerox Corporation nebo William and Flora Hewlett Foundation.

Knihovny a další kulturní instituce

Při srovnání počtu a rozložení zapojených institucí do obou projektů vycházejí následující údaje. Projekt OCA je větší v počtu spolupracujících knihoven a dalších informačních institucí, ale tito účastníci pocházejí především z USA, dvě knihovny z Evropy a po jedné knihovně z Austrálie, Indie a střední Ameriky (viz. Příloha B.2). Oproti tomu projekt GBS s menším počtem knihoven lépe pokrývá a zapojuje instituce z různých částí světa (11 knihoven z USA, 14 z Evropy a 1 z Asie; viz. Příloha B.3).

5.2.3 Finance

Zdroje a způsob financování

Finance vkládá do projektu GBS společnost Google, která za zprostředkování digitalizace a jejího financování získá digitální kopie dokumentů. Tyto kopie pak může využívat pro

poskytování služeb u svých produktů. Jiný model financování je použit u projektu OCA. Ten získává peníze formou příspěvků od členů konsorcia, ale také od sponzorů a dárců.

Náklady

V případě projektu GBS se odhadují celkové náklady okolo 1 mld. dolarů, které se vztahují k číslu 35 milionů knih, jež hodlají naskenovat, viz. Tabulka č. 2 (BAKSIK, 2010). To znamená, že náklady na skenování jedné knihy se pohybují v rozmezí od 30 do 100 dolarů za knihu. V tomto směru je cena skenování u projektu OCA totožná, jelikož Internet Archive uvádí cenu 10 centů za stránku nebo 30 dolarů za knihu. V této ceně jsou započteny i náklady na technologie spojené s dlouhodobým uložením atd. (Open Content Alliance, 2006).

Jméno instituce	Počet svazků
Committee on Institutional Cooperation	10,000,000
University of Michigan	7,000,000
University of Virginia	5,000,000
Stanford University	3,240,000
University of California	2,500,000
Bavarian State Library	1,000,000
Harvard University	1,000,000
University of Oxford	1,000,000
Princeton University	1,000,000
University of Texas, Austin	1,000,000
Cornell University	500,000
University of Wisconsin, Madison	500,000
Ghent University	300,000
National Library of Catalonia	200,000
Columbia University	200,000
Complutense University of Madrid	200,000
Keio University	120,000
University of Lausanne	100,000
New York Public Library	32,000
Celkem	34,892,000

Tabulka č. 2

5.2.4 Autorská práva

Podstatnou otázkou, kterou se zabývají projekty masové digitalizace a která je velmi ovlivňuje, jsou autorská práva. Ta mají značný vliv na výběr dokumentů pro digitalizaci. Jedná se hlavně o dokumenty, které jsou stále ještě pod ochranou autorských práv a tzv. osiřelá díla (orphan works). V USA musel Google vyřešit tento problém smlouvou s autory,

vydavatelé a dalšími zúčastněnými stranami v tzv. Google Book Settlement Agreement. Google byl odsouzen a potrestán v několika případech za porušení autorských práv, např. za zveřejnění plných textů francouzských knih dostal pokutu 300 tisíc eur. V souvislosti s tímto tématem odkazují na část Přístup v 5. kapitole, která má návaznost na zobrazení knih, jelikož autorská práva určují i přístupnost dokumentů ve službě Google Books. Projekt GBS tak dělí knihy a další dokumenty do tří kategorií: dokumenty chráněné autorským zákonem, dokumenty, kterým ochrana autorským zákonem vypršela a dokumenty, které jsou zveřejněny se souhlasem autorů, nakladatelů atd. (BAKSIK, 2010). Jednodušší pozici má projekt OCA, který digitalizuje hlavně dokumenty nechráněné autorským zákonem a tzv. osiřelá díla. Mezi těmito dokumenty se objevují i díla chráněná autorskými zákony. Je to možné díky souhlasu držitelů práv, která od nich získal projekt OCA.

5.2.5 Výběr materiálů k digitalizaci

U obou projektů je výběr dokumentů ovlivněn autorským právem, formáty dokumentů a jejich fyzickým stavem, který umožňuje skenování dokumentů.

5.2.6 Metadata

K tématu metadat projekt GBS příliš informací neposkytuje, a proto nebylo možné zjistit, jaká metadata projekt používá. Nicméně se v literatuře velmi často objevují články, které upozorňují na často chybná a nekvalitně zpracovaná metadata v projektu GBS. Více informací poskytuje projekt OCA, který používá bibliografická metadata MARC binary, MARC XML a Dublin Core, a technická metadata, která jsou udržována ve formátu XML.

5.2.7 Technologie – skenery atd.

Podobné jako u metadat je to i s poskytováním informací k technologiím používaným při skenování. Projekt GBS nezpřístupňuje podstatné informace k používaným skenerům a OCR technologii. Zatímco projekt OCA je v těchto směrech velmi otevřený. Pro skenování využívá projekt Scribe systém, který splňuje požadavky na automatickou obsluhu otáčení stránek nebo na kontrolu obrazu. Scribe skenery se skládají ze dvouúrovňových kamer, které slouží ke skenování špatně dostupných míst ve vazbě knih. K tomu je také nutná nastavitelná „kolébka“ s pohybem 90°, která pohybuje knihou (RIEGER, 2008). Dodavatelem OCR technologie pro projekt OCA je společnost Adobe a její produkt ABBYY FineReader 8.0.

5.2.8 Standardy souborů

Výstupy skenování jsou digitální dokumenty, které se zpřístupňují buď prostřednictvím textových formátů, které jsou získány použitím OCR technologie, nebo obrazových souborů. Projekt GBS zvolil textové formáty PDF, ePub a prostý text, který je dostupný pouze u děl bez ochrany autorských práv. Obrazové soubory JPEG2000 slouží pouze pro prohlížení dokumentů. Naproti tomu projekt OCA nabízí širokou škálu standardů souborů, které si může uživatel stáhnout, a to jak textové, obrazové tak i s podporou audio formátu. Stejně jako u GBS jsou používány textové formáty PDF, ePub, prostý text, navíc poskytuje i formát e-knih Mobi pro čtečku Kindle. Nadstavbou oproti Google je poskytovaný formát Daisy, který slouží lidem, kteří z různých důvodů nemohou využít tištěné dokumenty. Jedná se o formát, který propojuje text knihy a její zvukový záznam, tedy audioknihu. Možný přístup k textu je i odkaz „číst online“. Jde o rozhraní Internet Archive BookReader, které právě propojuje text a audio záznam. Specifický je formát DjVu, který propojuje naskenovaný text a fotografie dokumentů. Dokumenty lze podávat i jako sadu obrazových souborů, jeden obrázek na stránku. Soubory mohou být v JPEG2000, JPG nebo TIFF. Všechny soubory si lze stáhnout v jednom souboru, který je konvertován do souboru typu Zip nebo Tar.

5.2.9 Přístup k digitalizovaným dokumentům

Přístup k digitalizovaným dokumentům je zajištěn „portálem“ Google Books. Dalším možným přístupovým místem jsou digitální knihovny či repozitáře jednotlivých institucí do projektu zapojených. Výjimkou není ani to, že knihy z obou projektů jsou obsaženy v „portálech“ svých konkurentů, protože některé instituce jsou zapojeny do obou projektů, např. Columbia University, University of Texas nebo University of Virginia. Také dokumenty projektu OCA jsou dostupné prostřednictvím stránek OpenLibrary a Internet Archive.

5.2.10 Počet zpřístupněných dokumentů

V počtu naskenovaných a zpřístupněných knih jasně vede projekt GBS s více jak 12 miliony knih v roce 2010 (ODER, 2010). Projekt OCA je v tomto směru chudší. Ve stejném roce bylo uvedeno, že projekt naskenoval přes jeden milion knih (Open Content Alliance, 2006).

5.2.11 Obsah

Široký okruh beletrie, populární knihy, dětské knihy, historické texty a akademické texty jsou obsahem projektu OCA (Open Content Alliance, 2006). Projekt GBS však získává dokumenty

jak od nakladatelů, tak z partnerských knihoven, ale speciální sbírky knihoven jsou z tohoto projektu vyloučeny, i když jsou Googlem digitalizovány. Knihovny si totiž nepřejí jejich zveřejnění v rámci tohoto projektu (BAKSIK, 2010).

5.2.12 Uživatelské rozhraní

Projekt GBS se prezentuje jednoduchým a přehledným rozhraním, které nabízí tři způsoby zobrazení: plné zobrazení, které je možné pro knihy ve veřejném vlastnictví s možností stáhnutí si knihy, zobrazení úryvků a zobrazení pouze záznamu dokumentu (BAKSIK, 2010).

Projekt OCA zpřístupňuje svůj obsah přes stránky Internet Archive a OpenLibrary. Ze srovnání uživatelského rozhraní stránek vychází lépe OpenLibrary, které nabízí přívětivější uživatelské rozhraní. Internet Archive a OpenLibrary nabízí rozhraní Internet Archive BookReader, které je velmi příjemné a snadno ovladatelné pro čtení online.

Vyhledávání

Vyhledávání u obou projektů je rozděleno na základní a pokročilé. Rozhraní projektu OCA OpenLibrary se odlišuje samostatným vyhledávacím polem pouze pro hledání v plném textu tzv. Search Inside. Zatímco projekt GBS využívá vlastní vyhledávač, v projektu OCA se uživatel setká s opensource vyhledávačem Apache Lucene. Při vyhledávání lze použít booleovskou logiku v obou projektech, ale pouze v projektu OCA fuzzy vyhledávání. V pokročilém vyhledávání projekt OCA poskytuje podrobnou nápovědu, jakým způsobem pracovat s vyhledávačem a jak formulovat dotazy. Obsahuje také větší množství vyhledávacích polí a omezovacích polí. Náhled rozhraní pokročilého vyhledávání GBS a OCA (Internet Archive a OpenLibrary) v Příloze A.

Funkcionalita - služby

Projekt GBS nabízí více možností jak pracovat s digitalizovanou knihou a jejími záznamy. Umožňuje mu to napojení na ostatní služby společnosti Google, jako Google Maps (např. lokalizace místa děje atd.), vytvoření vlastního účtu tzv. My Library, který umožňuje personalizaci tohoto účtu a vzhledu podle preferencí uživatele a vytváření si vlastní knihovny. Některé funkce jsou v mnohém shodné s knihovními katalogy OPAC, jako jsou fasety nebo řazení knih dle různých kritérií. Po využití tohoto způsobu omezování selekčního dotazu se po zvolení příslušných omezovacích kritérií objeví lišta, která ukazuje dodatečně zvolená omezení. Je možné si vytvořit citaci exportem z jednoho z těchto citačních manažerů BiBTeX, EndNote nebo RefMan. Velké využití tady mají i prvky Webu 2.0. Uživatel může

napsat recenzi, ohodnotit hvězdičkami knihu, zaznamenat oblíbené pasáže knihy tzv. „Popular Passages“. Pod záznamy se mu objevují knihy se souvisejícími tématy. Součástí záznamů jsou i obsah, citace, seznamy literatury a odkazy. Rozhraní nabízí i možnost, jak se dostat k fyzickému dokumentu. To zajišťuje odkaz „Find this book in a library“ propojující projekt GBS s online katalogem WorldCat, nebo propojení na nakladatelství, kde se mohou knihy objednat jak v elektronické tak klasické podobě (BAKSIK, 2010).

V tomto směru se projekt OCA nemůže s projektem GBS rovnat, jelikož nenabízí tolik možností v rozhraní Internet Archive ani v rozhraní OpenLibrary.

Jako první budou popsány funkce a služby v rozhraní Internet Archive, který působí více jako archiv dokumentů než katalog knih. Po zobrazení výsledků vyhledávání se v rozhraní nachází fasety v pravé části obrazovky, které umožňují setřídit výsledky hledání např. podle relevance, data přidání, počtu stáhnutí nebo průměrného hodnocení. Dalšími možnostmi třídění výsledků je řazení podle sbírky, ke které dokument patří a druhu dokumentu (text, audio, videa atd.). Fasety také nabízejí přeformulování dotazu z nabízených možností zahrnující výběr z volby sbírky, autora, druhu formátu (obrázek, text atd.). Stejná nabídka je i u podobných položek. Ve srovnání s GBS obsahují záznamy dokumentů metadata, která jsou spojena s procesem digitalizace, jako je jméno sbírky, ze které dokument pochází, rozlišení, použitý skener, identifikátory apod. Rozdílem jsou také obsažené statistiky stahování dokumentů. Ty jsou rozděleny do dvou kategorií: týden a celkově staženo. Prohledávat obsah lze i prostřednictvím kategorií sbírek rozdělující dokumenty podle původu. Rozhraní obsahuje i možnosti Webu 2.0, tedy hodnocení dokumentů hvězdičkami a psaní recenzí.

Lepším dojmem než Internet Archive působí rozhraní OpenLibrary. Popis obsahu knih a autorů je propojen s Wikipedií. Záznamy zahrnují i klíčová slova, stejně tak jméno autora i jeho alternativní podoby, místo a období, o němž obsah knihy pojednává. Výsledky vyhledávání seskupují jednotlivá vydání díla pod jeden název. Součástí záznamu jsou odkazy na nakladatelství (Amazon, Android store ad.) pro zakoupení knihy, na knihovny, kde si mohou knihu půjčit (WorldCat nebo služba Over Drive spojující adresář knihoven a nakladatelství ve vybraných zemích s dostupností knihy). Odkaz na plný text v DAISY formátu (audiokniha) je dostupný pouze na základě klíče od Library of Congress, knihy bez autorské ochrany jsou zpřístupněny přes Internet Archive BookReader rozhraní. OpenLibrary nabízí i mnoho prvků Webu 2.0 (náhled obálky, propojení na GoodRead, uživatelský web hodnotící

knihy). Je zde nabízena možnost stáhnout si záznam z katalogu ve formátu RDF, JSON nebo OPDS. Personalizace je možná po přihlášení uživatele, který si může vytvořit a spravovat svůj seznam knih. Může editovat uvedená fakta o autorovi, knize a mají možnost přidat jakýkoli záznam knihy. Výsledky se mohou řadit podle nejnověji přidaných knih, jejich prvního publikování a četnosti vydání. Prohledávání dokumentů umožňuje prohlížení podle autora a předmětu (předmět, osoba, místo, čas).

5.3 Výsledky srovnání

Přestože projekt GBS spolupracuje s menším počtem institucí než OCA, zpřístupňuje větší množství dokumentů. V oblasti poskytovaných služeb a uživatelského rozhraní jsou si projekty podobné. To platí hlavně u GBS a přístupu k OCA přes OpenLibrary, protože rozhraní Internet Archive není na takové úrovni jako předchozí dvě. Oba projekty poskytují velký přínos pro společnost a uživatele, jelikož zlepšují dostupnost dokumentů pro vzdělání i volný čas. Projekty GBS a OCA také slouží jako dlouhodobá ochrana digitálních kopií skenovaných dokumentů, které podléhají různým fyzickým a mechanickým vlivům (např. kyselý papír, křehký papír, přírodní katastrofy atd.).

5.4 Rámcové srovnání projektů masové digitalizace s vybranou národní digitální knihovnou

Porovnávat vyspělé projekty NDK z Asie, hlavně z Jižní Koreje, Číny a Japonska, s projekty masové digitalizace GBS a OCA je nasnadě. Podle mého názoru jsou tyto projekty národních digitálních knihoven rovnocenné s projekty masové digitalizace, a to jak počtem digitalizovaných dokumentů, tak finančními prostředky. Hlavním rozdílem je globálnost, která je velkou výhodou projektů masové digitalizace, které se snaží o vytvoření celosvětové sbírky dokumentů a mohou oslovit více uživatelů. Přestože projekty NDK jsou pouze lokální, vytvářejí velké sbírky digitálních dokumentů. Nicméně jejich dopad je v podstatě omezen na určitou oblast.

U asijských projektů bych jako jejich přednost uvedl propracovanou koncepci, která vhodně reaguje na požadavky uživatelů. Jejich digitální koncepce mají podporu tamních vlád, oporu v zákonech (autorské právo) a vytvořily takové modely fungování, do kterých se zapojují i autoři, nakladatelé a instituce zastupující držitele autorských práv. Další výhodou je i to, že digitalizační aktivity mají základy v 90. letech 20. století právě v Japonsku a Jižní Koreji. To se

dá říci o digitalizačních aktivitách Francie a Norska, které v tomto směru patří k předním zemím v Evropě.

5.4.1 Srovnání projektů masové digitalizace s digitalizačním projektem Norské národní knihovny

V této podkapitole budou popsány společné body těchto projektů i jejich rozdíly, obecnější rovina byla popsána výše. Z informací uvedených v tabulce č. 1 lze odvodit, že projekty GBS a OCA jsou si po všech technických stránkách rovny a žádné velké rozdíly nelze v tomto směru očekávat.

Hlavními odlišnostmi Národní digitální knihovny Norska oproti projektům GBS a OCA jsou jak způsob financování, rychlost digitalizace, objem dokumentů dostupný, tak přístup k dílům chráněných autorským zákonem. Norský digitalizační projekt je výhradně financován státem na základě informační politiky, která digitalizaci uvádí za svou prioritu. To potvrzuje i vyřešení autorsko-právních otázek vydáním zákonů, které umožňují snáze digitalizovat dokumenty chráněné autorským právem (Norwegian Copyright Act a Legal Deposit Act). Norská národní knihovna nespolehlala pouze na zákon, ale dohodla se s institucí Kopinor, zastupující držitele autorských práv, na modelu dostupnosti dokumentů s autorsko-právní ochranou. V tomto směru je projekt v mnohém úspěšnější než GBS a OCA, jelikož GBS začal řešit otázku autorských práv až v okamžiku, kdy byly na společnost podány žaloby. OCA si vybrala snadnější cestu, projekt digitalizuje pouze díla bez autorské ochrany, výjimečně díla chráněná, ale vždy k tomu získá souhlas držitele práv.

6 Závěr

Hlavní cíl bakalářské práce byl popsat pojem masové digitalizace a porovnat její vybrané projekty. Podle výsledků srovnání byly zjištěny rozdíly mezi těmito projekty a jejich vliv na evropské digitalizační aktivity.

Velmi tenká hranice se nachází mezi masovou a „large-scale“ digitalizací. Definici masové digitalizace je proto velmi těžké popsat, jelikož závisí na kontextu a mnoha dalších faktorech, proto se musí digitalizační projekty posuzovat případ od případu a nesnažit se generalizovat. Masová digitalizace je úzce omezenou aktivitou pro pár projektů na světě. Snahu digitalizovat všechny dokumenty na světě určují především finance. Některé projekty masové digitalizace vznikají agregací výsledků menších digitálních projektů.

Projekt GBS dnes dominuje digitalizačním aktivitám v Evropě a USA, a to jak objemem zpřístupňovaných dokumentů, tak i finančním zajištěním. Projekt OCA je zaměřen především na americké knihovny a kulturní instituce. Asie je v tomto směru z velké části samostatná, jelikož asijské digitalizační projekty jsou velmi pokročilé.

V Evropě, kromě vybraných knihoven spojených se společnostmi Google a států podporující své národní digitalizační projekty, jako je Francie a Norsko, nelze o masové digitalizaci mluvit, dokud se nevyřeší problémy autorského zákona a financování. V současné době nejsou v rozpočtech jednotlivých států ani v evropských fondech finanční prostředky na rychlé dokončení digitalizačních projektů. V zámoří a Asii jsou na tom v tomto směru o mnoho lépe, jelikož se do těchto projektů zapojil i soukromý sektor. V Evropě by to znamenalo větší zapojení soukromých subjektů jako je Google.

V Evropě zatím nejsou jednotlivé národní digitalizační projekty schopny spolupracovat v jednom projektu masové digitalizace. Každý stát tak vytváří vlastní národní digitální knihovnu, která patří spíše do „large scale“ digitalizace. Existují však výjimky, projekty NDK Norska a Francie se naopak považují za masovou digitalizaci.

Díky společnosti Google a jejím digitalizačním aktivitám se v Evropě začalo pracovat na změnách právního rámce umožňujícího lepší přístup k osiřelým dílům. Podobné předpisy už přijaly některé evropské země, např. Francie, Norsko, plánováno v Německu atd. Pokud nedojde k vyřešení a inovaci autorského práva a způsobu financování, budou evropské digitalizační projekty stále tak těžkopádné jako doposud a naplnění projektů, jako je Europeana, bude velmi složité.

Co se týče masové digitalizace v Česku, je avizována spolupráce Národní knihovny ČR se společností Google v rámci projektu Národní digitální knihovny. Toto zapojení do projektu masové digitalizace pomůže urychlit digitalizaci sbírek Národní knihovny ČR.

Je tedy otázkou času, kdy a jak budou vytvořeny podmínky pro efektivní fungování projektu masové digitalizace v Evropě, protože evropská digitální knihovna Europeana zatím tento velký potenciál nenaplnila. Evropské státy by si měli vzít příklad z projektů v Asii a Americe nebo z projektů Francie a Norska, kde jsou digitalizační projekty efektivní a konkurenceschopné.

7 Seznam literatury

ASHLING, Jim. Progress Report: The British Library and Microsoft Digitization Partnership. *Information Today* [online]. 2007, vol. 24, Issue 10, s. 18-19 [cit. 2012-06-10]. ISSN 8755-6286. Dostupný komerčně v databázi LISTA: <http://search.ebscohost.com>.

AYRIS, Paul. The Status of Digitisation in Europe. *Liber Quarterly: The Journal Of European Research Libraries* [online]. 2010, vol. 19, no. 3/4, s. 193-226 [cit. 2012-06-10]. ISSN 1435-5205. Dostupné z: <http://liber.library.uu.nl/index.php/lq/article/view/7961/8252>.

BAKSIK, Corinna. Google Book Search Library Project. In: *Encyclopedia of Library and Information Sciences* [online]. 3rd ed. New York: Talor and Francis, 2011, s. 1982-1992 [cit. 2012-07-11]. ISBN 0-8493-9711-1. Dostupné komerčně v databázi Taylor & Francis e-journals: <http://www.tandfonline.com/>.

BALK, Hildelies, PLOEGER, Lieke. IMPACT: working together to address the challenges involving mass digitization of historical printed text. *OCLC Systems & Services* [online]. 2009, Vol. 25 Issue 4, s. 233–248 [cit. 2012-05-03]. ISSN 1065-075X. Dostupný z: Dostupný také z (DOI): <http://dx.doi.org/10.1108/10650750911001824>.

BARTOŠEK, Miroslav. Nástroje Google. 3. Google Book Search. *Zpravodaj ÚVT MU* [online]. 2009, roč. 19, č. 3, s. 16-20 [cit. 2012-05-03]. ISSN 1212-0901. Dostupné z: <http://www.ics.muni.cz/bulletin/articles/609.html>.

CEYNOWA, Klaus. Mass Digitization for Research and Study: the Digitization Strategy of the Bavarian State Library. *IFLA Journal*. 2009, vol. 35, no. 1, s. 17-24 [cit. 2012-05-03]. ISSN 0340-0353.

COLOMBET, Philippe. Google Books goes Dutch. In: *Inside Google Books* [online blog]. 14 July 2010 [cit. 2012-06-10]. Dostupné z: <http://booksearch.blogspot.cz/2010/07/google-books-goes-dutch.html>.

COYLE, Karen. Mass Digitization of Books. *The Journal of Academic Librarianship* [online]. 2006, vol. 32, no. 6, s. 641-645 [cit. 2012-05-03]. ISSN 0099-1333. Dostupné z (DOI): <http://dx.doi.org/10.1016/j.acalib.2006.08.002>. Dostupný také komerčně v databázi LISTA: <http://search.ebscohost.com>.

CRAWFORD, Walt. Discovering Books: The OCA/GBS Saga Continues. *Cites & Insights: Crawford at Large* [online]. 2006, vol. 6, no. 6 [cit. 2012-05-03]. ISSN 1534-0937. Dostupné z: <http://citesandinsights.info/v6i6a.htm>.

CUBR, Ladislav. *Strategie ochrany digitálních dokumentů*. Praha, 2009. 141 s. Vedoucí práce Jan Hutař. Diplomová práce.

DAWEI, Wei, YIGANG, Sun. The National Digital Library Project. *D-Lib Magazine* [online]. May 2010, vol. 16, Issue 5-6 [cit. 2012-07-11]. ISSN 1082-9873. Dostupné z: <http://www.dlib.org/dlib/may10/dawei/05dawei.html>.

DELORME, Bruno. Digitization at the Bibliothèque nationale de France, including an interview with Bruno Delorme. *Serials* [online]. 2011, vol. 21, Issue 3, s. 261-265 [cit. 2012-07-11]. ISSN 0953-0460. Dostupné komerčně v databázi LISTA: <http://search.ebscohost.com/>.

DUKE, Judy. Columbia U, Microsoft partner to digitize books. *Advanced Technology Libraries* [online]. 2008a, vol. 37, Issue 3, s. 2 [cit. 2012-06-10]. ISSN 0044-636X. Dostupné z: <http://site.ebrary.com/pub/atl/docDetail.action?docID=10224458&page=1>.

DUKE, Judy. Cornell Universtiy Library, Microsoft Partner. *Advanced Technology Libraries* [online]. 2006, vol. 35, Issue 12, s. 2-3 [cit. 2012-06-10]. ISSN 0044-636X. Dostupné z: <http://site.ebrary.com/pub/atl/docDetail.action?docID=10160701&page=1>.

DUKE, Judy. Microsoft Shuts Down Live Search Books Project. *Advanced Technology Libraries* [online]. 2008b, vol. 37, Issue 7, s. 1, 11-12 [cit. 2012-06-10]. ISSN 0044-636X. Dostupné z: <http://site.ebrary.com/pub/atl/docDetail.action?docID=10241551&page=1>.

DUKE, Judy. Yale and Microsoft partner on digitisation project. *Advanced Technology Libraries* [online]. 2007, vol. 36, Issue 12, s. 2-3 [cit. 2012-06-10]. ISSN 0044-636X. Dostupné z: <http://site.ebrary.com/pub/atl/docDetail.action?docID=10210522&page=1>.

FLOOD, Alison. Microsoft stops Book Search. *Bookseller* [online]. 2008, no. 5334, s. 8 [cit. 2012-06-10]. ISSN 0006-7539. Dostupný komerčně v databázi LISTA: <http://search.ebscohost.com>.

FRENCH, Amanda L. Imagine a National Digital Library: I Wonder If We Can. In: *AmandaFrench.net* [online blog]. 1 March 2011 [cit. 2012-06-10]. Dostupný z: <http://amandafrench.net/blog/2011/03/01/imagine-a-national-digital-library-i-wonder-if-we-can/>.

Google Books [online]. Mountain View: Google, c2011 [cit. 2012-07-11]. Dostupný z: <http://books.google.com/>.

HAHN, Trudi Bellardo. Impacts of mass digitisation projects on libraies and information policy. *Bulletin of the American Society for Information Science and Technology* [online]. October/November 2006, vol. 33, no. 1, s. 20-24 [cit. 2012-06-10]. ISSN 1931-6550. Dostupné v databázi ProQuest Central: <http://www.proquest.co.uk>.

CHOI, Wontae. The Development of Digital Libraries in South Korea. *Libri* [online]. 2003, vol. 53, no. 2, s. 131-141 [cit. 2012-07-11]. ISSN 0024-2667. Dostupné z: <http://librijournal.org/pdf/2003-2pp130-141.pdf>.

CHRISTENSON, Heather. HathiTrust: A Research Library at Web Scale. *Library Resources & Technical Services* [online]. 2011, vol. 55, no. 2, s. 93-102 [cit. 2012-05-03]. ISSN 0024-2527. Dostupné komerčně v databázi LISTA: <http://search.ebscohost.com/>.

Inside Google Books [online blog]. Google, c2011 [cit. 2012-07-11]. Dostupné z: <http://booksearch.blogspot.com>.

KAHLE, Brewster. Economics of Book Digitization. In: *Open Content Alliance* [online]. San Francisco: Internet Archive, 2009, Posted on 22nd March 2009 [cit. 2012-05-03]. Dostupné z: <http://www.opencontentalliance.org/2009/03/22/economics-of-book-digitization/>.

KNOLL, Adolf. Evropská digitalizace jako dětská hra. *Knihovna plus* [online]. 2005, č. 1 [cit. 2012-05-03]. ISSN 1801-5948. Dostupný z: <http://knihovna.nkp.cz/knihovnaplus51/knoll.htm>.

LACKIE, Rober J. From Google Print to Google Search: The Controversial Initiative and Its Impact on Other Remarkable Digitization Projects. *Reference Librarian* [online]. 2008, vol. 49, Issue 1, s. 35-53 [cit. 2012-06-10]. ISSN 0276-3877. Dostupný komerčně v databázi LISTA: <http://search.ebscohost.com>.

LIU, Wei. The New Development of Digital Libraries in China In: *DLKC'04: March 2 - 5, 2004: International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society* [online]. Tsukuba: University of Tsukuba, 2004 [cit. 2012-07-11]. Dostupné z: <http://www.kc.tsukuba.ac.jp/dlkc/e-proceedings/papers/dlkc04pp120.pdf>.

LOPATIN, Laurie. Library digitization projects, issues and guidelines: a survey of the literature. *Library Hi Tech* [online]. 2006, vol. 24, Issue 2, s. 273-289 [cit. 2012-05-03]. ISSN 0737-8831. Dostupné komerčně v databázi LISA: <http://search.proquest.com/>.

Million Book Project. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- , Last modified on 17 May 2012 [cit. 2012-07-20]. Dostupné z: http://en.wikipedia.org/wiki/Million_Book_Project.

MILNE, Ronald. The Google Mass Digitisation Project at Oxford. *Liber Quarterly: The Journal Of European Research Libraries* [online]. 2006, vol. 16, no. 1-4, s. 265-270 [cit. 2012-05-03]. ISSN 1435-5205. Dostupné z: <http://liber.library.uu.nl/index.php/lq/article/view/7853/8036>.

MURRELL, Mary. Digital + Library: Mass Book Digitization as Collective Inquiry. *New York Law School Law Review* [online]. 2010/2011, vol. 55, Issue 1, s. 221-249 [cit. 2012-05-03]. ISSN 0145-448X. Dostupné komerčně v databázi LISTA: <http://search.ebscohost.com/>.

Národní digitální knihovna [online]. Praha: Národní knihovna České republiky, 2010- [cit. 2012-07-11]. Dostupné z: <http://www.ndk.cz/narodni-dk>.

NÁRODNÍ KNIHOVNA ČR. *Národní knihovna ČR a Moravská zemská knihovna zahájily realizaci projektu Národní digitální knihovna - 28.6. 2010: tisková zpráva* [online]. Praha: Národní knihovna České republiky, 2010 [cit. 2012-07-11]. Dostupné z: http://www.nkp.cz/files/tz_ndk.pdf.

NÁRODNÍ KNIHOVNA ČR. *Národní kulturní bohatství zachovávají nové technologie. Statistika knih a periodik online - 22. 2. 2011: tisková zpráva* [online]. Praha: Národní knihovna České republiky, 2011 [cit. 2012-07-11]. Dostupné z: http://www.nkp.cz/files/tz_ndk_google.pdf.

NATIONAL DIET LIBRARY. *Annual Report of the National Diet Library FY2008: Trends of the past year* [online]. Kyoto: National Diet Library, 2008 [cit. 2012-07-11]. Dostupné z: <http://www.ndl.go.jp/en/publication/annual/2008/2008trends.pdf>.

NATIONAL DIET LIBRARY. *National Diet Library: (Kokuritsu Kokkai Toshokan)* [online]. Kyoto: National Diet Library, 2012 [cit. 2012-07-11]. Dostupné z: http://www.ndl.go.jp/en/aboutus/pdf/pamphlet_en.pdf.

NATIONAL LIBRARY OF NORWAY. *Digitization of books in the National Library: methodology and lessons learned* [online]. Oslo: Nasjonalbiblioteket, 2007 [cit. 2012-07-11]. Dostupný z: http://www.nb.no/content/download/2326/18198/version/1/file/digitizing-books_sep07.pdf

NEUBERT, Michael. "Books.Google.RU"--Google's Mass Digitization of Russian-Language Books. *Slavic & East European Information Resources* [online]. 2008, vol. 9, no. 1, s. 53-62 [cit. 2012-05-03]. ISSN 1522-8886. Dostupný komerčně v databázi LISA: <http://search.proquest.com>. Dostupný také z (DOI): <http://dx.doi.org/10.1080/15228880802104660>.

NIELSEN, Erland Kolding. Digitisation of library material in Europe : problems, obstacles and perspectives anno 2007. *Liber Quarterly: The Journal Of European Research Libraries* [online]. April 2008, vol. 18, no. 1 [cit. 2012-05-03]. ISSN 1435-5205. Dostupné z: <http://liber.library.uu.nl/index.php/lq/article/view/7901/8132>.

ODER, Norman. Google Book Search by the Numbers. In: *Library Journal* [online]. Plain City: Media Source, 2010, Published 12nd February 2010 [cit. 2012-06-10]. Dostupný z: <http://www.libraryjournal.com/article/CA6718929.html>.

Open Content Alliance [online]. San Francisco: Internet Archive, 2006- [cit. 2012-06-10]. Dostupné z: <http://www.opencontentalliance.org/>.

Open Content Alliance. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- , last modified on 11 May 2012 [cit. 2012-07-20]. Dostupné z: http://en.wikipedia.org/wiki/Open_Content_Alliance.

OSWALD, Ed. Two Libraries Join Windows Live Book Search. In: *BetaNews.com* [online]. 2006 [cit. 2012-06-10]. Dostupné z: <http://betanews.com/2006/06/09/two-libraries-join-windows-live-book-search/>.

POWELL, Christina Kelleher. OPAC Integration in the Era of Mass Digitization: The MBooks Experience. *Library Hi Tech* [online]. 2008, vol. 26, no. 1, s. 24-32 [cit. 2012-05-03]. ISSN 0737-8831. Dostupný z (DOI): <http://dx.doi.org/10.1108/07378830810857771>.

QUINT, Barbara. Open Content Alliance Rises to the Challenge of Google Print. *Information Today* [online]. 2005, Posted on 3 October 2005 [cit. 2012-06-10]. Dostupné z: <http://newsbreaks.infotoday.com/nbreader.asp?ArticleID=16110>.

REDMER, Jens. Madrid's Complutense University opens its library to the Word. In: *Inside Google Books* [online blog]. 26 September 2006 [cit. 2012-06-10]. Dostupné z: <http://booksearch.blogspot.cz/2006/09/madrids-complutense-university-opens.html>.

REDMOND, Wash. Microsoft to Collaborate With University of California and University of Toronto Libraries for Windows Live Book Search. In: *Microsoft News Center* [online]. Microsoft, 2006, 8 June 2006 [cit. 2012-06-10]. Dostupné z: <http://www.microsoft.com/en-us/news/press/2006/jun06/06-08BookSearchPR.aspx>.

RIEGER, Oya Y. *Preservation in the Age of Large-Scale Digitization: A White Paper* [online]. Washington D.C.: Council on Library and Information Resources, 2008 [cit. 2012-06-10]. CLIR publication, no. 141. ISBN 978-1-932326-29-1. Dostupné z: <http://www.clir.org/pubs/reports/pub141/pub141.pdf>.

RODRIGUEZ, Marko A., BOLLEN, Johan, VAN DE SOMPEL, Herbert. Automatic Metadata Generation Using Associative Networks. *ACM Transactions On Information Systems* [online]. February 2009, vol. 27, no. 2, s. 7:1-7:20 [cit. 2012-05-03]. ISSN 1046-8188. Dostupné z: <http://public.lanl.gov/herbertv/papers/Papers/2009/TOISrodriguez.pdf>. Dostupný také z (DOI): <http://dx.doi.org/10.1145/1462198.1462199>.

SAMUELSON, Pamela. A universal digital library is within reach. In: *Los Angeles Times* [online]. Published 1 May 2012 [cit. 2012-06-10]. Dostupný z: <http://articles.latimes.com/2012/may/01/opinion/la-oe-samuelson-google-books-and-copyright-20120501>.

SCHMITZ, Dawn. *The Seamless cyberinfrastructure : the challenges of studying users of mass digitization and institutional repositories* [online]. Washington D.C.: Council on Library and Information Resources, 2008 [cit. 2012-06-10]. Dostupný z: <http://www.clir.org/pubs/archives/schmitz.pdf>.

SKARSTEIN, Vigdis Moe. The Bookshelf: digitization and access to copyright items in Norway. *Program: electronic library and information systems* [online]. 2010, vol. 44, no 1., s. 48-59 [cit. 2012-07-11]. ISSN 0033-0337. Dostupný z (DOI): <http://dx.doi.org/10.1108/00330331011019681>.

SRINIVASAN, Anand. Google Finalizes Deal With Danish Royal Library For Books Digitization Project. In: *Go Rumors* [online]. Posted on 28 December 2009 [cit. 2012-06-10]. Dostupný z: <http://gorumors.com/google-finalizes-deal-with-danish-royal-library-for-books-digitization-project/27529>.

ST. CLAIR, Gloriana. Mass digitization Projects: Celebration and Challenges. In: *The Universal Digital Library* [online]. Pittsburg (Pennsylvania): Carnegie Mellon University, 2006 [cit. 2012-06-10]. Dostupné z: <http://www.ulib.org/conference/2006/4.pdf>.

ST. CLAIR, Gloriana. The Million Book Project in Relation to Google. *Journal of Library Administration* [online]. 2008, vol. 47, Issue 1/2, s. 151-163 [cit. 2012-06-10]. ISSN 0193-0826. Dostupné komečně v databázi LISTA: <http://search.ebscohost.com/>.

SUTHERLAND, Juliet. A Mass Digitization Primer. *Library Trends* [online]. 2008, vol. 57, no. 1, s. 17-23 [cit. 2012-05-03]. ISSN 0024-2594. Dostupný z: <http://muse.jhu.edu/journals/>.

TAKLE, Marianne. The Norwegian National Digital Library. *Ariadne: Web Magazine for Information Professionals* [online]. July 2009, Issue 60, Date published 30 July 2009 [cit. 2012-07-11]. ISSN 1361-3200. Dostupný z: <http://www.ariadne.ac.uk/issue60/takle/>.

TAYA, Hiroyuki. Experience and practice of digitization in Japan In: *National Diet Library* [online]. Kyoto: National Diet Library, 2010 [cit. 2012-07-11]. Dostupné z: http://www.ndl.go.jp/jp/publication/proceedings/sympo1201/5-1_Taya.pdf.

THE BRITISH LIBRARY. *The British Library and Google to digitise 250,000 books* [online]. London: British Library, 2011 [cit. 2012-06-10]. Dostupný z: <http://www.bl.uk/reshelp/inrrooms/stp/readerbulletin/jul2011/julyreadbull.html>.

The Universal Digital Library [online]. Pittsburg (Pennsylvania): Carnegie Mellon University, c2001-2008 [cit. 2012-06-10]. Dostupné z: <http://www.ulib.org/>.

VELDE, Janice van de. *Digitisation Research Project* [online]. Melbourne: National and state Libraries Australasia, 2011 [cit. 2012-07-11]. Dostupný z: http://www.nsla.org.au/sites/default/files/publications/NSLA.Digitisation_Research_Project_final_201109_0.pdf.

VERHEUSEN, Astrid. Mass Digitisation by Libraries: Issues concerning Organisation, Quality and Efficiency. *Liber Quarterly: The Journal Of European Research Libraries* [online]. 2008, vol. 18, no. 1, s. 28-38 [cit. 2012-05-03]. ISSN 1435-5205. Dostupný z: <http://liber.library.uu.nl/index.php/lq/article/view/7902/8134>.

WEISL, Annabella. Unlocking our shared cultural heritage: Google partners with the Austrian National Library. In: *Inside Google Books* [online blog]. 16 June 2010 [cit. 2012-06-10]. Dostupné z: <http://booksearch.blogspot.cz/2010/06/unlocking-our-shared-cultural-heritage.html>.

WILLETT, Perry. Mass digitization and its impact on interlending and document supply. *Interlending & Document Supply* [online]. 2009, vol. 37, Issue 3, s. 143-148 [cit. 2012-05-03]. ISSN 0264-1615. Dostupný z (DOI): <http://dx.doi.org/10.1108/02641610910985620>.

Přílohy

PŘÍLOHA A

Rozšířené vyhledávání knih
[Všechno o společnosti Google](#)

Najít výsledky

se **všemi** slovy
 s **přesnou frází**
 s **alespoň jedním** ze slov
neobsahující slova

Hledat Googlem

10 výsledků hledání ▾

Prohledat:	☒ Všechny knihy ☐ Omezený náhled a úplné zobrazení ☐ Pouze úplné zobrazení ☐ Pouze eknihy Google
Obsah:	☒ Veškerý obsah ☐ Knihy ☐ Časopisy
Jazyk	Zobrazit stránky napsané
Titul	Zobrazit knihy s titulem
Autor	Zobrazit knihy s autorem
Vydavatel	Zobrazit knihy s vydavatelem
Datum vydání	☒ Vrátit obsah publikovaný kdykoliv ☐ Vrátit obsah publikovaný mezi
ISBN	Zobrazit knihy s číslem ISBN
ISSN	Najít časopisy s číslem ISSN

kterýkoli jazyk ▾

 například Books and Culture

 např. Hamilton Mabie nebo "Hamilton Wright Mabie"

 například O'Reilly

 ▾ a ▾
 např. 1999 a 2000 nebo leden 1999 a prosinec 2000

 například 0060930314

 například 0161-7370

© 2011 Google

Obrázek 1: Rozhraní Google Book Search pro pokročilé vyhledávání

Internet Archive

Universal access to all knowledge

Web Moving Images Texts Audio Software Patron Info About IA Projects

Home Donate | Forums | FAQs | Contributions | Terms, Privacy, & Copyright | Contact | Archive Jobs | Jobs | Bios

Search:

Anonymous User (login or join us)

Advanced Search

This form allows you to perform an advanced search. You only need to fill in one field below. This can be any field. If you select "not" as your match criteria, you must select one other field.

Any field: contains

AND Title: contains

AND Creator: contains

AND Description: contains

AND Collection: is All collections

AND Mediatype: is All mediatypes

AND Custom field: contains

AND Custom field: contains

AND Custom field: contains

AND Date: YYYY MM DD

AND Date range: YYYY MM TO YYYY MM DD

Obrázek 2: Rozhraní Internet Archive pro pokročilé vyhledávání

PŘÍLOHA A

Advanced Search returning JSON, XML, and more

This will return results in the format of your choice.

Query:

Fields to return (pick one or more):

- avg_rating
- call_number
- collection
- contributor
- coverage
- creator
- date
- description
- downloads
- foldoutcount
- format
- headerImage
- identifier
- imagecount
- language
- licenseurl
- mediatype
- month
- num_reviews
- oai_updatedate
- publicdate
- publisher
- rights
- scanningcentre
- source
- subject
- title
- type
- volume
- week
- year

(optional) Sort results by:

Number of results:

Page:

Indent response: ☐

JSON format: ☒

XML format: ☐

save to file: ☒

HTML table: ☐

CSV format: ☐

[\(show/hide help\)](#)

RSS format **beta!**: ☐

("Fields to return" ignored)

Obrázek 3: Rozhraní Internet Archive pro pokročilé vyhledávání

Log in / Sign Up

SUBJECTS

AUTHORS

LISTS

ADD A BOOK

RECENTLY

ABOUT US



One web page for every book.

Search Inside

☐ Show only eBooks

[Hide options](#)

Title

Author

ISBN

Subject

Place

Person

Publisher

[Full Text Search?](#)

[Top](#) | [Home](#) • [Add a Book](#) • [Subjects](#) | [Authors](#) | [Lists](#) • [About Us](#) | [Help](#) • [Developers](#)

Around The Library

Bryan Tyson updated [The Oxford Handbook of Biblical Studies](#) 1 hour ago

menolly42 updated [Solid ground](#) 1 hour ago

[Chapbooks One to Ten \(Writers Forum Wallet\)](#) was updated anonymously 44 minutes ago

Open Library is an initiative of the [Internet Archive](#), a 501(c)(3) non-profit, building a digital library of Internet [other projects](#) include the [Wayback Machine](#), [archive.org](#), [nasaimages.org](#), [archive-it.org](#) & [opencontentall](#)

Your use of the Open Library is subject to the Internet Archive's [Terms of Use](#). And, here's an [index](#) of all the pages.

Obrázek 4: Rozhraní OpenLibrary pro pokročilé vyhledávání

PŘÍLOHA B
PŘÍLOHA B.1

Seznam knihoven zapojených do projektu The Million Book Project

Čína

Beijing University
Chinese Academy of Science
Fudan University
Ministry of Education of China
Nanjing University
State Planning Commission of China
Tsinghua University
Zhejiang University

Egypt

Bibliotheca Alexandrina

Indie

Arulmigu Kalasalingam College Of Engineering
Goa University
Indian Institute of Information Technology - Allahabad
Indian Institute of Science
International Institute of Information Technology - Hyderabad
Shanmugha Arts, Science, Technology & Research Academy
Tirumala Tirupati Devasthanams
Maharashtra Industrial Development Corporation
University of Pune
(Million Book Project, 2001)

PŘÍLOHA B.2

Seznam knihoven zapojených do projektu Open Content Alliance

Boston Library Consortium	University of Alberta Libraries
The Bancroft Library	University of British Columbia Library
The British Library	University of California Libraries
Columbia University Libraries	University of Chicago
Emory University Library	University of Georgia
European Archive	University of Illinois at Urbana-Champaign
Getty Research Institute	University of North Carolina at Chapel Hill
Indiana University Libraries	University of Ottawa
Johns Hopkins University Libraries	University of Pittsburgh
McMaster University	University of Texas
Memorial University of Newfoundland	University of Toronto
Missouri Botanical Garden	University of Virginia Library
The National Archives	Washington University
National Writers Union	York University Library
Natural History Museum, London	Biodiversity Heritage Library, a
National Library of Australia	cooperative project of:
Perseus Project	American Museum of Natural History
Prelinger Library and Prelinger Archives	Harvard University Botany Libraries
Research Libraries Group	Harvard University, Ernst Mayr Library of the Museum of Comparative Zoology
Rice University Libraries	Missouri Botanical Garden
San Francisco Public Library	Natural History Museum, London
Simon Fraser University Library	The New York Botanical Garden
Smithsonian Institution Libraries	Royal Botanic Gardens, Kew
Universidad Francisco Marroquín, Guatemala	Smithsonian Institution Libraries
	(Open Content Alliance, 2001)

PŘÍLOHA B.3

Seznam knihoven zapojených do projektu Google Book Search

Harvard University Library, Harvard

University of Michigan Library, Michigan

New York Public Library

Bodleian Library, Oxford

Stanford University Libraries (SULAIR)

Bayerische Staatsbibliothek (Bavorská státní knihovna)

Columbia University Library System

CIC (Committee on Institutional Cooperation):

Indiana University

Michigan State University

Northwestern University

The Ohio State University

Penn State University

Purdue University

University of Chicago

University of Illinois

University of Iowa

University of Minnesota

Complutense University of Madrid

Biblioteca de Catalunya (Katalánská národní knihovna)

Cornell University Library, Cornell

Ghent University Library/Boekentoren

Keio Media Centers (Libraries), Keio

La Bibliothèque Municipale de Lyon

Princeton University Library

California Digital Library, California

Cantonal and University Library of Lausanne/Bibliothèque Cantonale et Universitaire (BCU)

University of Texas Libraries, Texas

University of Virginia Library, Virginia

University of Wisconsin Libraries, Wisconsin

Koninklijke Bibliotheek (Národní knihovna Nizozemí)

Österreichische Nationalbibliothek (Rakouská národní knihovna)

Národní knihovna ČR

Ministero per i Beni e le Attività Culturali (Itálie, Ministerstvo kultury)

The British Library

Det Kongelige Bibliotek (Dánská královská knihovna)

(Inside Google Books, c2011)