



Repozitáře šedé literatury

Editor Petra Pejšová

Kolektiv autorů:

Marcus Vaska
Joachim Schöpfel
Iveta Fürstová
Radim Polčák
Jan Mach
Bohdana Frantíková
Petr Karlach
Jindřich Dolanský

Praha 2010

Národní technická knihovna
Vysoká škola ekonomická v Praze

NTK

50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Technical Library



Vydal: Radim Bačuvčík - VeRBuM, Zlín 2010

KATALOGIZACE V KNIZE - NÁRODNÍ KNIHOVNA ČR

Repozitáře šedé literatury / editor Petra Pejšová ;
kolektiv autorů Marcus Vaska ... [et al.]. -- 1. vyd. -- Zlín : VerBuM, 2010
Pod názvem uvedeno: Národní technická knihovna, Vysoká škola ekonomická
v Praze
ISBN 978-80-904273-5-8

08-027.571 * 004.08:930.25 * 005.591-027.542 * 024.5/.6:002.1 * [024.5+659.2]-
021.161 * 021.6 * 316.4.063.34 * 006.032/.033 * (437.3)

- Národní úložiště šedé literatury (projekt)
- šedá literatura
- digitální repozitáře -- Česko
- národní programy a projekty -- Česko -- 2001-2020
- zpřístupňování dokumentů
- zpřístupňování informací
- kooperační systémy knihoven
- mezinárodní spolupráce
- mezinárodní standardizace a normalizace
- kolektivní monografie

004 - Počítačová věda. Výpočetní technika. Informační technologie [23]

Publikace vznikla v rámci řešení projektu *Digitální knihovna pro šedou literaturu - funkční model a pilotní realizace*, který podpořilo Ministerstvo kultury České republiky v rámci programových projektů

Text je zveřejněn za podmínek licence Creative Commons: Uveďte autora - Neužívejte dílo komerčně - Nezasahujte do díla 3.0 Česko

Recenzovali: RNDr. Miroslav Bartošek, CSc., Ing. Petr Žabička
Publikaci doporučila k vydání Vědecká redakce nakladatelství VerBuM

1. vydání

© Petra Pejšová (ed.), Marcus Vaska, Joachim Schöpfel, Iveta Fürstová,
Radim Polčák, Jan Mach, Bohdana Frantíková, Petr Karlach, Jindřich Dolanský

© Radim Bačuvčík - VerBuM, 2010

ISBN 978-80-904273-5-8

Obsah

Předmluva	5
<i>Ing. Martin Svoboda, NTK</i>	
Úvod - popis účelu a obsahu knihy	7
<i>PhDr. Petra Pejšová, NTK</i>	
1. kapitola: Úvod do šedé literatury	11
<i>Marcus Vaska, University of Calgary, Kanada</i>	
2. kapitola: Přístup k evropské šedé literatuře	19
<i>Joachim Schöpfel, Ph.D., University of Lille, Francie</i>	
3. kapitola: Vývoj šedé literatury v České republice	33
<i>PhDr. Petra Pejšová, NTK</i>	
4. kapitola: Partnerská síť NUŠL	43
<i>Mgr. Iveta Fürstová, NTK</i>	
5. kapitola: Zpřístupnění vysokoškolských kvalifikačních prací	55
<i>Ing. Jan Mach, VŠE</i>	
6. kapitola: Právní aspekty šedé literatury	66
<i>JUDr. Radim Polčák, PF MU Brno</i>	
7. kapitola: Popisné údaje pro šedou literaturu	78
<i>Bc. Bohdana Frantíková, NTK</i>	
8. kapitola: Analýza a podklady pro výběr softwarového řešení pro NUŠL	93
<i>Ing. Petr Karlach</i>	
9. kapitola: Pilotní implementace NUŠL	107
<i>Ing. Jindřich Dolanský a PhDr. Petra Pejšová, NTK</i>	
10. kapitola: Audit NUŠL s použitím nástroje DRAMBORA	123
<i>Ing. Petr Karlach</i>	
Závěr	133
Resume	134
Přílohy	135
Seznam zkratk	142
Rejstřík	147

Předmluva

O šedé literatuře, které se věnuje tato publikace, tvrdí řada autorit, že přináší nejčerstvější vědecké poznatky a současně je obecně známo, že je jen obtížně vyhledatelná a ještě hůře dostupná. Dalo by se tedy předpokládat, že všichni, kdo financují výzkum a vývoj, budou usilovně podporovat nástroje, které tuto dostupnost zlepší, už jen proto, aby snížili pravděpodobnost příslovečného „znovuvynalézání kola“ a tak zlepšili efektivitu vynakládaných veřejných prostředků. S touto nadějí jsme po kolapsu *Evropské asociace pro využívání šedé literatury*, jejíž let příliš orlí nebyl a která především nesplňovala to podstatné, tj. spolehlivě dodat originální dokument, iniciovali před pěti lety schůzku reprezentantů *Rady vlády České republiky pro výzkum a vývoj*, *Akademie věd České republiky* a vysokoškolských knihoven a představili tam záměr, jak situaci alespoň v národním měřítku výrazně zlepšit. K tomu ovšem nestačí dohodnout a vytvořit technické řešení a přesvědčit podobně uvažující instituce, že je moudré takového nástroje užívat. Od zmíněné schůzky uplynulo dlouhých pět let a velmi si vážím toho, kam projekt *Národní úložiště šedé literatury*, o němž hovoří tato kniha, dospěl. I když si nedělám iluze o stoprocentní účinnosti právních norem, přece jen předpis, který producentům kteréhokoliv typu šedé literatury uloží jeho povinné zpřístupnění, podstatným způsobem změni ochotu ke spolupráci - jak jsme byli svědky u disertačních prací a ostatních vysokoškolských kvalifikačních prací. Zatímco v mnoha zemích jsou výsledky výzkumu placeného z veřejných prostředků veřejně dostupné, v České republice, kromě toho, co je publikováno v tradičních dokumentech, jsou další informace prakticky nedostupné. Řada významných druhů šedé literatury dosud legislativní podporu nemá a je tedy zatím jen na ochotě a povědomí o užitečnosti spolupráce jednotlivých institucí, jak bude projekt nabývat na významu. Věřím, že úspěch projektu bude možné využít jako argument pro vyjednání legislativní podpory, aby mohl dosáhnout zamýšlených cílů. Ne pro slávu *Národního úložiště šedé literatury*, ale pro lepší využití nemalých veřejných prostředků do výzkumu a vývoje vkládaných. K tomu by měla přispět i tato kniha, jíž přeji bohaté řady laskavých čtenářů.

Ing. Martin Svoboda

ředitel Národní technické knihovny v Praze

Úvod

Potenciál šedé literatury v posledních letech stále více zajímá odborníky z oblasti vzdělávání, výzkumu a vývoje, ale také informační pracovníky. Tato kniha si klade za cíl informovat o šedé literatuře odbornou komunitu a ukázat možnou cestu k vybudování repozitáře šedé literatury. Odborná publikace *Repozitáře šedé literatury* je jednou z prvních tohoto druhu v České republice, ale i v zahraničí, která přináší ucelený pohled na problematiku budování repozitářů šedé literatury. Obdobná vědecká publikace od Dominica J. Farace a Joachima Schöpfela s názvem *Grey Literature in Library and Information Studies* vyšla v zahraničí v září 2010.

Kniha *Repozitáře šedé literatury* představuje pojem šedé literatury, její vývoj, typy a producenty především v evropském kontextu se zaměřením na Českou republiku. Současně přibližuje na příkladu *Národního úložiště šedé literatury* budovaného v *Národní technické knihovně*, které bylo prakticky implementováno, postup budování repozitáře s aspekty šedé literatury. Mezi nejdůležitější aspekty při budování repozitářů šedé literatury patří spolupráce a legislativa. Pro praktické vybudování digitálního repozitáře uvádí kniha nutné komponenty, jako analýza, výběr a implementace software, tvorba metadatového formátu a audit důvěryhodnosti repozitáře.

Úvodní kapitola představuje pojem šedé literatury, její historii, druhy a vývoj. Věnuje se též producentům šedé literatury, kteří jsou velmi různorodí oproti producentům klasické, tzv. bílé literatury. V druhé části rozebírá znaky a vliv šedé literatury na odbornou i laickou veřejnost. V závěru uvádí příklady šedé literatury v různých vědních oborech.

Druhá kapitola se věnuje šedé literatuře v Evropě. Uvádí příklady mandatorních politik a strategií akvizice šedé literatury. Představuje významné kolekce šedé literatury jak v tištěné, tak v elektronické podobě. Velký důraz v této kapitole je kladen na kooperaci v oblasti sběru, uchování a zpřístupnění šedé literatury na všech úrovních od oborových přes národní až k mezinárodním sbírkám.

Vývoj šedé literatury v České republice je popsán ve třetí kapitole společně se současnými systémy a projekty věnovanými kooperaci v této oblasti. Uvedena je zde historická i současná role *Národní technické knihovny* v budování systému sběru šedé literatury v České republice a spolupráce s mezinárodními kooperačními systémy. Na podporu zajištění této role realizuje od roku 2008 projekt, v jehož rámci vzniklo *Národní úložiště šedé literatury*.

Spolupráce s producenty je klíčovou činností při tvorbě repozitářů. Přestože je šedá literatura velmi cenným zdrojem informací, je klasickými způsoby sběru těžko podchytitelná. Velkou roli zde hrají kontakty a budování partnerských sítí. Kooperace má i druhou stranu, a to společný vývoj standardů, metodik a nástrojů, protože tato činnost je velmi náročná a není finančně ani kapacitně možné, aby se jí věnovaly jednotlivé organizace samostatně. Důležité ovšem je,

aby za určitou oblast šedé literatury vždy převzala vůdčí roli určitá organizace, která bude v této činnosti podporována.

Vysokoškolské kvalifikační práce jsou natolik odlišným druhem šedé literatury, a to i z právního hlediska, že jim byla věnována samostatná kapitola. Stav zpřístupnění vysokoškolských kvalifikačních prací na jednotlivých vysokých školách v České republice je velmi různorodý. Zástupci vysokých škol využili zahraniční zkušenosti na jejichž základě vznikl metadatový formát EVSKP-MS. Díky spojení vysokých škol vznikl v roce 2008 *Národní registr vysokoškolských kvalifikačních prací*.

Důležitým aspektem je právní pohled na šedou literaturu. Ten je daný legislativou každé země a zde je popsán pohled na základě české legislativy. Z hlediska práva není důležité přesné vymezení šedé literatury, ale jaké právní režimy mohou mít díla označovaná za šedou literaturu. Významnými pojmy, které tato kapitola rozebírá, jsou autorská práva, zákonné licence, volné licence, majetková autorská práva, školní dílo, zaměstnanecké dílo, exkluzivní licence apod. V závěru kapitoly se nachází doporučení pro pořizovatele databází šedé literatury z právního hlediska a úvaha nad možnostmi využití volných licencí.

V sedmé kapitole je věnován náležitý prostor popisným údajům neboli metadatům. Stěžejní část kapitoly tvoří popis metadatového formátu vytvořeného pro *Národní úložiště šedé literatury*, který je rozdělen na povinné prvky, specifické prvky a technická a administrativní metadata. Před samotným popisem metadatového formátu jsou stručně popsány vzorové výchozí metadatové formáty, kterými byly Dublin Core, OpenSIGLE, ETD-MS, MARC21, EVSKP-MS a ASEP.

Klíčovým krokem při budování každého repozitáře je výběr softwarového systému. Požadavky, které byly kladeny na software vybíraný pro *Národní úložiště šedé literatury*, byly rozděleny do třinácti skupin a lze je použít i pro výběr institucionálních repozitářů. Protože oblast softwaru pro digitální repozitáře velmi dobře pokrývají aplikace z oblasti Open Source, byla provedena analýza nejpoužívanějších systémů, kterými jsou CDS Invenio, DSpace, EPrints, Fedora a Greenstone.

Softwarové řešení pilotní implementace *Národního úložiště šedé literatury* je složeno ze dvou částí. První část představuje CDS Invenio pro digitální repozitář a druhou částí je indexační a vyhledávací systém ESP FAST. Protože přednastavená lokální instalace CDS Invenia pro sbírky šedé literatury je nabízena spolupracujícím organizacím k instalaci na vlastním hardware, je velká část kapitoly věnována právě tomuto systému. Pro vyhledávání nejen nad repozitářem *Národního úložiště šedé literatury*, ale též nad dalšími repozitáři vědeckých zdrojů obsahujících šedou literaturu v České republice, slouží centrální vyhledávací rozhraní provozované pomocí systému ESP FAST.

Závěrečná kapitola se věnuje auditu důvěryhodného digitálního repozitáře opět na příkladu *Národního úložiště šedé literatury*. Tento audit byl proveden jako samoaudit s pomocí nástroje DRAMBORA. V kapitole je stručně popsán průběh auditu a jeho výsledek formou seznamu identifikovaných rizik. Význam auditu

je také v tom, že se v jeho rámci stanovují strategie, jak s riziky nakládat a předcházet jim a případně řešit jejich následky. Nástroj DRAMBORA je k dispozici online a je možno v něm audit průběžně aktualizovat a vyhodnocovat.

Zajímavé čtení za celý autorský kolektiv Vám přeje

Petra Pejšová

1. Úvod do šedé literatury

Marcus Vaska

Ještě před několika lety by pojmu „šedá literatura“ (grey literature) plně porozuměla pouze uzavřená, těsně spjatá komunita vědců. Mnozí tento pojem můžou chápat jako odkaz na nějakou smutnou, těžkou dobu nebo i pochybnou autoritu. Existují také jedinci, kteří vidí obsah této literatury čistě vědecky, ve vztahu k šedé kůře mozkové, a zahrnují tedy do šedé literatury znalosti, které mají vysoce intelektuální charakter (Mason, 2010).

V dnešní době lze však již snadno poznat, co znamenají výrazy **graue Literatur**, **Letteratura grigia**, **literature grise** nebo **grey literature**. Je pozoruhodné, jak se díky úsilí v posledních dvou desetiletích změnilo původně předjímané pojetí šedé literatury. Vznik, výběr a použití těchto **nekonvenčních materiálů** se staly nedílnou součástí výzkumného úsilí. Nástup a použití nových technologií, které představují převládající trend společnosti 21. století, vytvořily vzrůstající povědomí o potřebě zařadit šedou literaturu vedle bílé, čímž se jejich hranice pomalu smazává a v blízké budoucnosti by mohla zmizet úplně.

Pojem šedá literatura: Definice toho, co je skutečně „šedé“

Úloha vytvořit výstižnou **definici šedé literatury** je stejně obtížná jako snaha zvýšit obecné povědomí o existenci této velké skupiny publikací. V diskuzích o roli šedé literatury ve výzkumu někteří vědci odvážně prohlašují, že tato literatura je „neopěvovaný hrdina, pěšák, základ stavby“ (Mason, 2010). Podobná lichotivá označení nepochybně proklamují důležitost těchto materiálů. **Nízkonákladová vydání** a omezený **distribuční systém** šedé literatury mohou však v důsledku vést k přehlížení role tohoto materiálu v procesu vyhledávání.

Nástup a rozvoj šedé literatury jsou doprovázeny řadou pokusů o co nejlepší vysvětlení a popis tohoto pojmu. Tento typ literatury byl již tak často nazýván „šedou“, že není třeba se zabývat ostatními synonymy tohoto výrazu, která se někdy vyskytují. Jsou to zvláště výrazy jako **nekonvenční**, **neformální**, **neformálně publikovaná**, **dočasná** a **nehmotná**. Auger (1994) zachází v tomto ohledu ještě dále a tvrdí, že „dokumenty mohou být nekonvenční v mnohých ohledech a mnoho konvenčně publikovaných dokumentů vykazuje aspekty šedé literatury“ (Mason, 2010, s. 3).

Používá se několik definic šedé literatury, což ukazuje, že tento druh literatury je ze své podstaty skutečně obtížné definovat. Weintraub (2000) definuje šedou literaturu jako „publikace vydané vládou, akademií, podniky a průmyslem, v tištěné i elektronické podobě, které nejsou řízeny komerčními zájmy nakladatelů a kde jsou přitom články, diplomové práce, vládní dokumenty, bulletiny, technické zprávy, sborníky z konferencí a další publikace

distribuvány zdarma, na základě předplatného, nebo jsou na prodej.” Pro srovnání uvádí příslušné heslo v encyklopedii Wikipedia, že šedá literatura je „termín užívaný střídavě inteligencí, knihovníky, pracovníky v oboru medicíny a výzkumu pro označení souboru materiálů, které nejsou běžně dostupné konvenčními prostředky, jako např. nakladatelství” (Grey Literature, 2010). Další definice formulovaná před více než deseti lety na konferenci GL'97 (3rd International Conference on Grey Literature) je asi tou nejpoužívanější v akademické obci: „literatura, která vzniká na všech úrovních v oblasti vlády, akademie, podniků a průmyslu v tištěné i elektronické podobě, a to nezávisle na komerčních vydavatelích” (GreyNet, 1997).

Výše zmíněná **charakteristika** zdůrazňuje, že výzkumní pracovníci by měli v zájmu co nejmenšího **zkreslení** při zpracovávání rešerší (zvláště v případě zpracování **systematických** souhrnných studií) studovat šedou literaturu společně s publikovanými materiály. Vědec by se měl při hledání informací vždy snažit zůstat objektivní. Pokud přitom nezohlední různé podoby a formáty literatury, může tato chyba vést až k neobjektivnímu jednostrannému výzkumu.

Běžným omylem týkajícím se šedé literatury je názor, že tato literatura je pouze **částečně publikovaná**, neboť je distribuovaná jen omezeně, a není tedy tak snadno dostupná široké veřejnosti. Ovšem již původ slova publikovat neboli zveřejňovat explicitně vyjadřuje učinit přístupným veřejnosti. Vše, co je zveřejněno, je tedy publikovaným materiálem, nezávisle na tom, zda zrovna my víme o jeho existenci.

Šedá literatura: historie a vývoj

V závislosti na tom, jak striktně se aplikuje výše uvedená definice šedé literatury, lze přiznat jistou váhu i spekulativním teoriím, že tento druh literatury existoval již od dob antického Říma. Poznámky Archimeda, Sokrata a dalších podobně smýšlejících **antických vědců** lze v zásadě již považovat za určitou hrubou formu šedé literatury (Auger, 1989). Šedá literatura není tedy novým **pojetím**, i když tyto typy dokumentů nebyly vždy označovány jako „šedé“. Šedá literatura byla vlastně označována jako **Reports Literature**.

Jeden z nejstarších dokumentů šedé literatury, *The Behaviour of Aeroplanes in Gusts* od **NASA**, pochází z leteckého průmyslu (Auger, 1989, s. 13). První zaznamenaná technická zpráva šedé literatury vznikla během **vojenských činností** ve Velké Británii v roce 1909 (*Research Memoranda of the Aeronautics Research Council*) a další v roce 1915 v rámci činnosti Národního poradního výboru USA pro letectví. Během druhé světové války došlo k významnému nárůstu množství šedé literatury. Tato dokumentace sloužila jako hlavní prostředek k předávání výsledků výzkumu a zkoušek. Během válečného období vzrostla potřeba větší **bezpečnosti** a **utajení**. Dokumenty musely být produkovány tajně, nikoliv klasickou cestou (Auger, 1989). Jak dále podotýká Gelfand (2000), „**kořeny šedé literatury** leží hluboko v inteligenci a vládních výzkumných komunitách” (Gelfand, 2000, s. 74) a ve většině případů budou v této cestě pokračovat.

Před několika desetiletími byly publikace šedé literatury považovány za zvláštní druh mimo hlavní proud komerčně dostupných knih a časopisů a nebyly tedy permanentní součástí sbírek. Když se však uváží vzrůstající počet záznamů v knihovních katalozích, které se týkají nekonvenčního materiálu, toto pojetí je stále obtížněji obhajitelné. Šedá literatura prošla intenzivním vývojem; od svého prvního uznání v 70. letech až k formálnímu zařazení do vyhledávacího procesu v 90. letech (Auger, 1989).

Typy šedé literatury

Díky technologicky pokročilé společnosti, v níž žijeme, získala šedá literatura mnoho podob. Bez ohledu na formát dokumentů považují znalí uživatelé šedou literaturu „za primární zdroj informací“ (Alberani, 1990, s. 358). Kromě tradičního přístupu k publikacím jako jsou diplomové a disertační práce, články a sborníky z konferencí (publikované i nepublikované verze), překlady (vědecké a technické literatury), bulletiny a vládní zprávy v tištěné podobě vzrostly alarmující rychlostí možnosti elektronického přístupu k těmto materiálům. Mnohé dokumenty šedé literatury jsou díky vyhledávači Google a nástrojům federativního vyhledávání obvykle přístupné na internetu: statistiky, zdroje dat, bulletiny, materiály z kurzů, podniková literatura, knihovní katalogy, digitální knihovny, patentová literatura, konferenční sborníky atd. Nové způsoby distribuce těchto informací vytvořily pak nové možnosti šíření. Jedná se zvláště o multimédia, sociální sítě a další formy neformální, avšak informačně hodnotné komunikace, jako jsou telefonní hovory, zápisy ze schůzí, e-mail, rozhovory, i publikace ve Wikipedii či jiné otevřené encyklopedii.

Dalším typem šedé literatury, který získává na významu, je produkce a distribuce prospektů, plakátů, lístků a rozvrhů. Auger (1994) tento typ definuje jako pomíjivý materiál, „který nese verbální informaci, vzniká tiskem nebo jako ilustrace, ovšem ne ve standardním formátu knihy, periodika nebo brožury“ (Auger, 1994, s. 7).

Expanze prostředků sociální sítě jako spolehlivého zdroje informací vyvolala značnou debatu ve vědecké komunitě. Opatrné posuzování a zkoumání důvěryhodnosti materiálu dostupného na webových stránkách jako Twitter, Second Life a Facebook vedlo část akademické obce k tomu, že tyto zdroje vyloučili z šedé literatury. Funkcí těchto multimediálních produktů je poskytnout náhled na současné povědomí, což je rys šedé literatury, který nelze opomenout. Světová pandemie chřipky H1N1, jež vypukla minulý rok a hrozí opětovným výskytem, byla na Twitteru velmi aktivně diskutována. Zvláště během prvních měsíců, kdy se zdravotnické instituce usilovně snažily zorganizovat program plošného očkování. Příspěvky zde byly poměrně informativní a byly veřejně přístupné dlouho předtím, než byl v denním tisku publikován nějaký článek.

I přes potenciál digitálního věku stále existuje velké množství šedé literatury, která ještě nepřekročila pomyslnou mez a není přístupná v elektronické podobě. Dlouholetý model ručního vyhledávání v časopisech, studia bibliografie

a přímé komunikace s vědci je stále platný a bude patrně ještě nějakou dobu pokračovat (Helmer, 1999).

Producenti šedé literatury

Řada existujících typů šedé literatury dala vznik stále rostoucímu množství organizací, které mají za úkol vytvářet, distribuovat a uchovávat šedou literaturu. Každý producent od **soukromých vydavatelů** a **společností** až po **sdržení**, **rady**, **výzkumné instituce**, **federace**, **podniky**, **odborové organizace**, **trusty** a **univerzity** hraje přitom klíčovou roli v rozšiřování povědomí o šedé literatuře.

Hnutí Open Access pro volný přístup, díky němuž je odborný materiál uživatelům **volně dostupný** na **internetu**, způsobilo nárůst počtu **digitálních knihoven** šedé literatury a **digitálních repozitářů**. Ať už jsou producenti známí pouze na místní úrovni, např. *Grey Literature LibGuide* vyvinutý na *Univerzitě v Calgary*, nebo ať mají širší záběr v národním měřítku, např. *Asociace knihoven vysokých škol České republiky* (dále jen AKVŠ) nebo *Akademie věd České republiky* (dále jen AV ČR), či dosahují světového uznání jako prominentní národní a mezinárodní repozitáře šedé literatury, jako např. *Grey Literature Network Service*¹ (dále jen **GreyNet**) nebo **OpenSIGLE**, evropský informační systém o šedé literatuře. Jejich cíl je vždy stejný: „usnadnit komunikaci mezi jednotlivci a organizacemi produkujícími šedou literaturu” (International Organizations-NUŠL, 2009). Seznam databází šedé literatury v České republice a v zahraničí sumarizují přílohy č. 1 a 2.

Rysy a význam šedé literatury

Rysy šedé literatury lze charakterizovat následovně: **rychlá publikace**, **rozličnost formátů**, **flexibilita**, **detailnost**, **absence recenzního řízení**, **absence komerčního zdroje** a ve většině případů **omezená distribuce**. Rychlost a flexibilita jsou jednoznačnými výhodami šedé literatury, neboť je tak umožněna specifická, cílená diskuze o případně méně známých tématech. Jak ukazuje rozmanitost a množství typů šedé literatury, tato literatura tvoří součást každodenního života, od ranního čtení novin po e-mail a internetové aktivity, jichž se velká většina společnosti denně účastní. Coonin (2003) odvážně tvrdí, že „vše, co čteme kromě časopisů a knih, lze považovat za šedou literaturu.” S ohledem na tuto myšlenku je snadné pochopit, proč byl vznik a vývoj šedé literatury „troj- až čtyřnásobně rychlejší než u konvenční literatury” (Farace, 1997). Již téměř dvě desetiletí jsou ve světě pořádány **konference o šedé literatuře**, kde se diskutuje o významu šedé literatury pro **informační společnost**.

¹ <http://www.greynet.org/>

Okamžitý přístup k informacím je užitečný ve fyzickém i virtuálním světě. Přes expanzi šedé literatury, většinou díky řadě technologických vymožeností (zvláště internetu), zůstávají někteří kritici skeptičtí ohledně skutečné informační hodnoty literatury, která není recenzována a je často aktualizována. Dominic Farace, uznávaná kapacita v oboru šedé literatury, namítá, že takový přístup znesnadňuje šedé literatuře získat větší uznání mezi svými uživateli: „jenom proto, že šedá literatura není řízena komerčními vydavateli, neznamená to automaticky, že neprošla stejně přísným recenzním řízením jako komerčně publikované materiály” (Gelfand, 2000, s. 74). Outtenová (2008) patrně uzavřela tuto debatu argumentem, že i když se nejedná o vědeckou formu publikace, šedá literatura je produktem specialistů v daném oboru a „slouží odborníkům i laikům podobně jako vědecká shrnutí, fakta, statistiky a další údaje, které zprostředkovávají podrobnější pohled na daný předmět.” Se vzrůstajícím zájmem o šedou literaturu, který se projevuje v řadě oblastí včetně např. distančního studia, nelze tento druh literatury přehlížet.

Výběr relevantních studií šedé literatury a jejich začlenění do systematických souhrnných studií mohou být zvláště časově náročné a obtížné. Je též diskutabilní, zda mají být nepublikované studie zahrnuty do souhrnných studií, protože mohou být nekompletní a může být těžké zjistit jejich metodickou kvalitu (Helmer, 1999). Proto vzniká dilema, zda je při snazším přístupu k dokumentům šedé literatury v porovnání s minulostí zachována kvalita. Vzhledem ke svému **mezioborovému charakteru** může daný dokument šedé literatury obsahovat více než jedno vyhledávací heslo. Ze studií vyplývá, že nástroje vyhledávání typicky indexují pouze 16% všeho dostupného obsahu na internetu. Tyto elektronické nástroje jsou však to první, k čemu se uživatel uchýlí ve snaze o zpracování neznámého tématu (Helmer, 1999). Pokud není informace odpovídajícím způsobem indexována, je prakticky nemožné ji nalézt. Tato skutečnost přímo koreluje s potřebou udržovat repozitáře materiálů šedé literatury, jako např. GreyNet. Pak je možné se zaměřit na problém v přístupu na „hluboký web“ (Invisible či Deep Web) a řešit jej. Aby bylo vyhledávání podrobnější, je třeba využívat i další zdroje kromě online databází. Ze studií vyplývá, že „za použití rozšířených vyhledávacích metod kromě těch, jež jsou běžně nabízené vyhledávači, bylo nalezeno dalších 29,3% položek“ (Helmer, 1999).

Je zřejmé, že vzhledem k miliardám stránek na internetu nelze vždy snadno ověřit obsah nebo kvalitu. Vědci využívající tento zdroj informací by proto měli být opatrní a nespoléhat se bez rozmyslu na vše, co si přečtou. Základním rysem materiálů šedé literatury publikovaných na webu je každopádně vysoká rychlost publikace, často bezprostředně po napsání.

Na poli šedé literatury jsou hlavními problémy **absence bibliografické kontroly, neprofesionální formát a omezený náklad** (Auger, 1989). Ohledně nevypovídajícího názvu a **nedostatku popisných informací** Auger uvádí, že mnoho publikací aplikuje v indexování minimalistický přístup: „málo detailů, žádné abstrakty ani rejstříky” (s. 6).

S potenciálem takového množství informací na dosah ruky může být nesmírně obtížné se rozhodnout, kde začít. Prohlížení webových stránek organizací

a pročitání bibliografií a seznamů použité literatury se nabízí jako první krok. Ovšem ne všechny dokumenty jsou dostupné online, některé jsou stále vydávány pouze tiskem. Může být tedy nutné přistoupit k tradičnímu vyhledávání informací v tištěných časopisech, zvláště nejsou-li indexovány v primární databázi distributora. Poslední možností je kontaktovat přímo autora dané publikace nebo příslušnou organizaci. Při dalším vyhledávání v šedé literatuře na dané téma mohou být dobrým výchozím bodem systematické souhrnné studie a abstrakty disertačních prací.

Šedá literatura v různých **oborech**: budování digitálních repozitářů

I přes velké množství materiálů šedé literatury je třeba stále překonávat překážky v podobě nedostatečného povědomí o tomto materiálu a přístupu k němu. Myšlenka otevřených repozitářů, kde jsou příslušné informace přímo volně dostupné na internetu, si získává stále větší podporu (Shearer, 2010). Shearer zdůrazňuje, že budování repozitářů šedé literatury je možné, neboť „většina časopisů (asi 70 %) povoluje autorům archivovat si kopie svých článků v repozitáři“ (Shearer, 2010, s. 25). Následující příklady krátce shrnují důležitost a význam otevřených repozitářů v různých specializacích souvisejících s šedou literaturou.

„Publikované klinické zkoušky jsou většinou rozsáhlejší a vykazují výraznější výsledky léčby než „šedé“ zkoušky (Hopewell, 2007). Badatelé musí proto při psaní souhrnných studií zhodnotit objektivně šedé i publikované zkoušky, aby se snížilo riziko zkreslení.“ To může být někdy obtížné, zvláště v oblasti zdravotnictví, kde rozhodnutí přímo ovlivňují životy pacientů. Nevyhodnotí-li se při péči o pacienta i data šedé literatury může to vést k přehnaně optimistickým odhadům o účinnosti léčby, zvláště pokud jsou rozhodnutí založena pouze na publikovaných zkouškách. Je nutné podrobně prohledávat i vládní zdroje (místní, národní i mezinárodní), zprávy, registry, sdružení, klinické zkoušky a/nebo klinické pokyny a případně databáze s doporučeními založenými na důkazech. I když nejsou tyto typy publikací tak dobře známe jako klasické zdroje, je alternativní a často přehlížený přístup k léčbě potřebný pro kvalitní zhodnocení problému. Problém udělování a/nebo omezování přístupu k elektronickým záznamům o léčbě, které jsou samy součástí šedé literatury, hraje zásadní roli v oblasti posuzování zdravotnické technologie (Helmer, 1999).

Krátkodobý charakter plakátů, programů a podobných publikačních materiálů dělá z šedé literatury klíčovou součástí materiálů v oblasti **hereckého umění**. Tyto dokumenty jsou často jediným prostředkem, k němuž se uživatelé mohou uchýlit, pátrají-li po určitých dramaticích, hercích nebo umělcích (Grey Literature Speaker Series, 2009). Naproti tomu právníký svět se běžně denně zabývá **soudními zápisy**, **zprávami komisí**, **politickými dokumenty**, **soubory formulářů** a samozřejmě **případovými studiemi**, což jsou všechno typy šedé literatury.

Výše uvedené příklady demonstrují význam šedé literatury v konkrétních oblastech, je však třeba zvážit i mezioborový charakter výzkumných studií. Společné úsilí **humanitních** a **sociálních věd** vedlo k produkci několika dokumentů o zločinech holokaustu, a tak dalo „tvář tomu, co bylo ztraceno, aby tak mohli čtenáři a odborníci zachytit obludnost toho, co se stalo šesti milionům Židů“ (Grey Literature Speaker Series, 2009). Záznamy očitých svědků, které samy o sobě tvoří repozitář šedé literatury, jsou zárukou toho, že toto období rasového pronásledování nebude nikdy zapomenuto.

Zachování existence šedé literatury: **budoucí možnosti**

Ačkoliv je příjemné zjištění, že čím dál více studentů, fakult a výzkumných pracovníků se zajímá o šedou literaturu, úkol rozšířit povědomí o tomto druhu literatury není zdaleka dokončen. I přes naši intenzivní snahu existuje stále množství dokumentů šedé literatury pouze v papírové podobě; velká část těchto materiálů není dosud přístupná online. Technologický pokrok nicméně umožňuje výzkumným pracovníkům v oblasti šedé literatury sdílet své znalosti s širší veřejností. Jak vyplývá z tématu konference GL12 a z této monografie, šedá literatura je vskutku **společným dílem**, „spravovaným komunitou učenců, s pomocí univerzitní knihovny nebo technických služeb“ (Shearer, 2010, s. 28).

Od počátků hnutí během posledního desetiletí 20. století je rozmach šedé literatury v informační společnosti nepopíratelný. Tyto úvahy o významu šedé literatury je možná vhodné uzavřít vizí Farace, kterou předestřel před více než 10 lety: „Stejně jako v posledním desetiletí dvacátého století dominovala oblast nabídky informací, tak bude před koncem prvního desetiletí století jednadvacátého dominovat oblast poptávky po informacích“ (Gelfand, 2000, s. 76).

LITERATURA:

ALBERANI, V. - de CASTRO PIETRANGELI, P. - MAZZA, A. *The use of grey literature in health sciences: A preliminary survey*. Bulletin of the Medical Library Association, 1990, sv. 78, č. 4, s. 358-363.

AUGER, C. *Information sources in grey literature*, 2. vydání. Londýn: Bowker-Saur.

AUGER, C. *Information sources in grey literature*, 3. vydání. Londýn: Bowker-Saur.

CARROL, B. - COTTER, G. *A new generation of grey literature: The impact of advanced information technologies*.

FARACE, D. *Rise of the phoenix: A review of new forms and exploitations of grey literature*. Publishing Research Quarterly, sv. 13, č. 2, s. 69.

GELFAND, J. *Interview with Dominic Farace, founder of GreyNet*. The International Journal on Grey Literature, sv. 1, č. 2, s. 73-76.

GreyNet. *Perspectives on the Design and Transfer of Scientific and Technical Information*. GL'97: Third International Conference on Grey Literature, Lucemburk.

HELMER, D. - SAVOIE, I. - GREEN, C. a *British Columbia Office of Health Technology Assessment*. *How do various fugitive literature searching methods impact the comprehensiveness of literature uncovered for systematic review?* Vancouver: British Columbia Office of Health Technology Assessment.

HOPEWELL, S. - McDONALD, S. - CLARKE, M. - EGGER, M. *Grey literature in meta-analyses of randomized trials of health care interventions*. Cochrane Database of Systematic Reviews, (2).

SHEARER, K. *A review of emerging models in Canadian academic publishing*. University of British Columbia: University of British Columbia Library.

ELEKTRONICKÉ ZDROJE:

COONIN, B. *Grey literature: An annotated bibliography*. 2003. [cit. 2010-03-20]. Dostupné z www: <http://personal.ecu.edu/cooninb/Greyliterature.htm>

Grey Literature [online]. 2010. [cit. 2010-03-20]. Dostupné z www: http://en.wikipedia.org/wiki/Grey_literature

Grey Literature Speaker Series [online]. The University of Calgary. 2009. [cit. 2010-03-20]. Dostupné z www: <http://glspeakers.wetpaint.com>

International organizations - NUSL [online]. Národní technická knihovna. 2009. Aktualizováno 19. 10. 2010. [cit. 2010-06-10] Dostupné na www: http://nrgl.techlib.cz/index.php/Mezin%C3%A1rodn%C3%AD_organizace

MASON, M. *Grey literature: History, definition, acquisition, and cataloguing*. 2010. [cit. 2010-03-20]. Dostupné z www: <http://www.moyak.com/papers/grey-technical-literature.html>

OUTTEN, C. *Grey literature*. 2008. [cit. 2010-03-20]. Dostupné z www: http://www.csulb.edu/library/subj/gray_literature

2. Přístup k evropské šedé literatuře

Joachim Schöpfel

Sít'ová spolupráce

Shromažďování šedé literatury je obtížným úkolem odborníků v oblasti knihovnictví a informační vědy (dále jen KIV). Šedou literaturu ve formě zpráv, sborníků nebo formulářů nelze zakoupit jako časopisy nebo knihy. Neexistuje žádná zvláštní agentura ani dodavatel šedé literatury.

Nákup informací je součástí tradiční role knihoven, spolu s jejich výpůjční a archivační funkcí. S ohledem na ekonomickou definici šedé literatury jako „materiálu, který je obvykle dostupný zvláštními cestami a zpravidla není součástí běžného distribučního (...) procesu (...)“ (Schöpfel a Farace, 2010), je zřejmé, že systematické shromažďování šedé literatury vyžaduje zvláštní pozornost, kvalifikaci a postupy.

Materiály šedé literatury jsou často výsledkem trpělivé a dlouhodobé investice do profesionálních kontaktů a sít'ové spolupráce. Sít'ovou spoluprací je zde míněno sdílení informací o obsahu šedé literatury, o nových zdrojích a cestách jejího získávání s dalšími odborníky KIV. Aktuálně jsou dostupné dvě osobní iniciativy Webu 2.0. První je Barrerova osobní stránka se soubory *Really Simple Syndication*² (dále jen RSS), věnovaná sdílení interních znalostí o biomedicínské šedé literatuře a druhá je Giustinioho zpráva *Finding the Hard to Finds*,³ která popisuje strategie shromažďování a vyhledávání šedé literatury.

Odborníci KIV začali s „šedou spoluprací“ mnoho let předtím, než Tim O'Reilly vynalezl Web 2.0. V roce 1980 založila vědecká informační centra EU databázi *System for Information on Grey Literature in Europe* (dále jen SIGLE), aby byla zpřístupněna evropská šedá literatura a zlepšilo se bibliografické pokrytí. Po počátečním financování *Komisi Evropských společenství* utvořila národní centra neziskovou síť pro získávání, identifikaci a šíření šedé literatury s názvem *European Association for Grey Literature Exploitation* (dále jen EAGLE). Každé národní centrum pro vědeckotechnické informace v této síti mělo národní sbírku šedé literatury nebo alespoň garantovalo dodávání dokumentů z partnerských sbírek.

V roce 2005 se obecný výbor rozhodl ukončit činnost EAGLE, protože jeho organizační struktura a ekonomický model se nedokázaly adaptovat na generaci internetových technologií a Googlu; např. SIGLE neumožňoval online

² http://www.netvibes.com/ernestobarrera#Grey_literature

³ <http://www.slideshare.net/giustinid/finding-the-hard-to-finds-searching-for-grey-gray-literature-2010>

katalogizaci, sběr metadat, odkazy na plné texty a další zdroje (Schöpfel et al., 2007).

Tentýž obecný výbor EAGLE však v roce 2005 jednohlasně rozhodl, že evropská spolupráce v oblasti šedé literatury bude zachována a že model z roku 1980 bude transformován do právě vznikajícího prostředí otevřeného přístupu k vědeckým informacím, zvláště v kontextu Berlínské deklarace z roku 2003.

Prvním krokem byla archivace záznamů SIGLE v otevřené a volně přístupné databázi, která je v souladu s protokolem OAI-PMH⁴ pro sběr metadat. Ve francouzském centru *Institute for Scientific and Technical Information of the French National Center for Scientific Research*⁵ (dále jen INIST) byla vyvinuta databáze OpenSIGLE⁶ založená na softwaru DSpace, který vyvinula *Massachusetts Institute of Technology* (dále jen MIT) a většina záznamů SIGLE v ní byla uložena ve zjednodušeném formátu XML (Farace et al., 2009). Následujícím krokem bude federace evropských otevřených projektů pro šedou literaturu, aby byl znovu otevřen přístup k evropské šedé literatuře.

Další sítě šedé literatury existují na místní nebo národní úrovni. Např. na počátku 90. let minulého století spravovala francouzská vláda mezirezortní skupinu *Littérature Grise Administrative* (dále jen LIGRIA) pro zpracovávání šedé literatury.

Jak se lze poučit z této a dalších iniciativ? Mnoho odborníků KIV spravujících sbírky šedé literatury má o spolupráci zájem. V knihovnickém prostředí je spolupráce velmi rozšířeným a oblíbeným fenoménem. Zprostředkování kultury je kolektivní, nikoliv samotářská činnost. V tomto případě nestačí ovšem samotný kladný postoj k síťové spolupráci. Účinná síťová spolupráce vyžaduje zkušenost a kvalifikaci, společné cíle, dobře organizované kontrolní struktury a udržitelný ekonomický model. To vysvětluje, proč zůstává síťová spolupráce někdy osobní záležitostí jednotlivců a nikoliv celých institucí.

Závazná politika

Některé sbírky šedé literatury zaujímají jasnou závaznou politiku, např. vyplývají přímo z národních nebo místních rozhodnutí. Může se jednat o vědecká, kulturní, anebo politická rozhodnutí, např. pro zajištění uchování a přístupu ke konkrétnímu obsahu nebo jako příspěvek k budování kulturního či vědeckého dědictví.

Prvním případem je povinný výtisk šedé literatury. Jedna ze tří speciálních vědeckých knihoven v Německu, *Německá národní knihovna vědy a techniky*⁷ (dále jen TIB) v Hannoveru, která oslavila v červnu 2009 své 50. výročí, se

⁴ Open Archives Initiative Protocol for Metadata Harvesting

⁵ <http://international.inist.fr/>

⁶ <http://opensigle.inist.fr>

⁷ <http://www.tib-hannover.de>

označuje jako místo přenosu vědeckých informací. Úkolem TIB je „získávat a podrobně archivovat literaturu z celého světa v oblasti inženýrství a přírodních věd“ (TIB, 2010). Knihovna klade zvláštní důraz na získávání šedé literatury (např. sborníky z konferencí, výzkumné zprávy, normy a disertační práce v tištěné a elektronické podobě). Sbírky šedé literatury v této knihovně jsou v rámci Německa unikátní. V roce 2010 evidovala TIB více než 210 000 tištěných a 30 000 elektronických německých výzkumných zpráv z oblasti inženýrství a přírodních věd. Každý měsíc přibývá asi 200 nových elektronických a 500 tištěných zpráv. TIB je knihovna pro ukládání elektronických závěrečných projektových zpráv z dotací *Federal Ministry of Education and Research* (dále jen BMBF⁸). Od roku 1996 musí být závěrečná zpráva každého výzkumného projektu doručena do TIB jako tištěná kopie a na elektronickém médiu (viz též Meyer, 2009). Podobně *Moscow Scientific and Technical Information Centre* (dále jen VNTIC) shromažďuje kopie ruských disertačních prací a od roku 1982 jich shromáždilo již 500 000.

Úplně jiný model reprezentuje předepsaná povinnost distribuce sbírek disertačních prací v akademických knihovnách s centrálním přístupem. Francouzská vláda vydala v roce 1985 vyhlášku, která upravovala a zlepšovala distribuci disertačních prací. Místní knihovna skladuje daný dokument, ale záznam o něm je součástí francouzského národního katalogu *Système Universitaire de Documentation*⁹ (dále jen SUDOC), který zprostředkovává objednávání a doručování tištěných kopií (Paillassard et al., 2007).

Třetím modelem jsou bi- nebo multilaterální dohody o získávání a distribuci šedé literatury v kontextu národní politiky vědeckotechnických informací. Tento model dobře ilustruje francouzský příklad. Významná část sbírek francouzských disertačních prací a vědeckých zpráv v centru **INIST** je také uložena na *Ministerstvu vyššího vzdělávání* (disertační práce) nebo v publikující instituci (výzkumné organizaci, ministerstvu). Jiným příkladem je *Dánská královská knihovna*, která byla do července 2002 knihovnou pro ukládání dokumentů Rady Evropy, Spojených národů, NATO, OECD, UNESCO a dalších mezinárodních organizací.

Dlouhodobé shromažďování a uchovávání specifických položek šedé literatury je možné zvláště, pokud jsou tyto jmenovité mandáty doprovázeny veřejným financováním. Někdy též usnadňují projekty digitalizace vědeckého dědictví. Problém těchto jmenovitých mandátů a dohod spočívá v tom, že mohou být nedostatečně nebo krátkodobě financovány, což přináší riziko nekompletní a nekontinuální sbírky.

⁸ Bundesministerium für Bildung und Forschung

⁹ <http://www.sudoc.abes.fr>

Strategie získávání dokumentů

Definice koherentní politiky získávání dokumentů je zásadní součástí fungování knihovny. Tato politika mnohdy odráží potřeby a názory sponzora knihovny, výběr předmětu či strukturu rozpočtu. Některé knihovny mohou též vyvinout specifickou strategii získávání dokumentů šedé literatury nezávisle na externím mandátu. Takové vnitřní instituční strategie mohou dát časem vznik výjimečným sbírkám.

Pravděpodobně nejznámější sbírkou tohoto typu je sbírka *Boston Spa conferences collection* (dále jen Boston Spa) Britské knihovny, čítající asi 450 000 položek. „Britská knihovna má jednu z nejobsáhlejších a nejpřístupnějších sbírek konferencí v anglickém jazyce na světě. (...) Politikou Britské knihovny je, že vlastnictví materiálu je pouze částí celého procesu zpřístupnění, a tak vyvinula řadu produktů pomáhajících uživateli tento materiál nalézt” (Tillett et Newbold, 2006). Tato sbírka si klade za cíl shromažďovat sborníky ze všech vědeckých konferencí v anglickém jazyce. Během zlatého věku nabídky informací věděli odborníci KIV i zákazníci, že Boston Spa má ve sbírkách téměř všechny mezinárodní vědecké konference.

Další samostatnou skupinou dokumentů sbírky Boston Spa jsou vědecké a technické zprávy z několika tisíc veřejných a soukromých britských, amerických a mezinárodních zdrojů.¹⁰ Roční zpráva Britské knihovny z let 2008 - 2009 zmiňuje 10,5 milionu zpráv na mikrofilmu a na webových stránkách se uvádí 4,9 milionů veřejně dostupných zpráv.

Jako komplement ke svému mandátu německých zpráv investuje TIB do systematického sběru zahraničních vědeckých zpráv, zvláště z *National Technical Information Service* (dále jen NTIS) z Ministerstva obchodu Spojených států, z NASA, ale i z významného množství evropských institucí, v celkovém počtu téměř 2 miliony.

Tyto dva příklady též ukazují, že zatím nelze mluvit o formální evropské koordinaci národních nebo místních šedých sbírek. Stále převažuje národní nebo instituční politika, např. strategie uchovávání, národní nezávislost, bilaterální dohody apod.

Někde na půli cesty mezi mandátem a institučním uspořádáním se nalézají sbírky diplomových a disertačních prací. Akademické knihovny jsou obvykle pověřeny shromažďovat práce ze své domovské vysoké školy, ale současně sbírají více či méně systematicky práce z jiných vysokých škol s ohledem na jejich pravidla a kritéria (obory, předměty, instituce atd.). Na první pohled v tom není rozdíl. Když však knihovny posuzují a pročišťují své sbírky, ponechají si „místní sbírku” a práce z jiných vysokých škol jsou často vyhodnoceny k odpisu.

¹⁰ NASA, NTIS, ERIC, DOE, FAO, INIS, ESA atd.

Od tištěných sbírek k digitálním

Od vynálezu a úspěchu internetu opouští knihovny tištěný svět Gutenberga. Činí tak dvěma způsoby. Převádějí své tištěné sbírky na digitální a sbírají a archivují přímý digitální materiál.

Šedá literatura se těmito dvěma cestami vydává. Pro disertační práce v tištěné podobě nebo na mikrofilmu vznikly významné retro-digitalizační projekty. Britská knihovna digitalizuje práce z britských vysokých škol pro nový portál EthOS-portal,¹¹ který spojuje možnost volného přístupu k elektronickým vysokoškolským kvalifikačním pracím (dále jen **eVŠKP**) z otevřených repozitářů a dodávku prací digitalizovaných na vyžádání.

Ve Francii vyvíjí *L'Association Nationale de la Recherche et de la Technologie*¹² (dále jen ANRT) aktivitu s cílem digitalizace v rámci své služby *Thèses à la carte*. Německé digitalizační projekty financované *Deutsche Forschungsgemeinschaft*¹³ (dále jen DFG) zahrnují hlavně rané primární zdroje kulturního dědictví, ale též odborné publikace (rukopisy, časopisy atd.). Britský Joint Information Systems Committee¹⁴ (dále jen JISC) investuje od roku 2003 do digitalizace obsahu zvláštních sbírek, např. přibližně 600 svazků historických zpráv o sčítání lidu uložených na *Univerzitě v Essexu* a 10 000 diplomových prací pro EthOS.

Seznam evropských digitalizačních programů pro šedou literaturu je dlouhý a lze do nich zahrnout i retro-digitalizační program *Delf University of Technology* (dále jen TU Delft) pro jejich vyhledávač elektronických diplomových prací, program *Univerzity v Uppsale* pro více než 11 000 diplomových prací z 18. a 19. století, katalánskou síť elektronických diplomových a disertačních prací *Tesis Doctorals en Xarxa*¹⁵ (dále jen TDX) nebo koordinační aktivitu digitalizačního projektu *Digital Libraries Team* skupinou PSNC¹⁶ v polské Poznani.

O poznání méně digitalizačních projektů je věnováno zprávám, konferenčním sborníkům a dalším formám šedé literatury. Pokud je autorovi známo, žádná evropská iniciativa není srovnatelná s americkým digitalizačním projektem sbírek zpráv *Department of Energy*¹⁷ (dále jen DOE) nebo se spoluprací *Office of Scientific a Technical Information* (dále jen OSTI) s dalšími centry (FERMI,

¹¹ <http://ethos.bl.uk/>

¹² <http://www.anrtheses.com.fr/>

¹³ <http://gepris.dfg.de>

¹⁴ Digitalizace JISC a program elektronického obsahu: <http://www.jisc.ac.uk/digitisation>.

¹⁵ <http://www.tesisenxarxa.net/>

¹⁶ Poznan Supercomputing and Networking Center,

<http://www.geysers.eu/index.php/theproject/partners/psnc>

¹⁷ <http://www.energy.gov/>

LANL atd.).¹⁸ Tři současné iniciativy ve Francii se týkají zpráv *Bureau de recherches géologiques et minières* (dále jen BRGM) pro portál věd o zemi, platformy *Libre accès aux rapports scientifiques et techniques*¹⁹ (dále jen LARA) pro vědecké zprávy z různých institucí a matematického archivu *Numérisation de documents anciens mathématiques*²⁰ (dále jen NUMDAM) s 29 semináři za období 1948 až 2007.

Tyto programy sdílejí tři společné rysy, a to jasně definované sbírky šedé literatury, charakter vědeckého dědictví a nízký stupeň koordinace s dalšími iniciativami.

Digitalizované položky šedé literatury jsou někdy sloučeny s digitálním materiálem nebo s existujícími elektronickými sbírkami. Jedním z příkladů je francouzský národní repozitář eVŠKP *Theses en Ligne* (dále jen TEL) s více než 10 000 pracemi z posledních pěti let (2005 - 2010) a téměř 2 000 digitalizovanými pracemi publikovanými do roku 1990.

Další program, britského *The Centre for Environmental Data Archival*²¹ (dále jen CEDA) se sídlem v radě STFC²² v *Rutherford Appleton Laboratory*, je věnován šedé literatuře o pozorování země a atmosférickým vědám. Patrně všechny položky CEDA (více než 600) jsou původem digitální. Další evropské repozitáře s digitalizovaným nebo digitálním materiálem šedé literatury lze nalézt v adresáři *The Directory of Open Access Repository* (dále jen **OpenDOAR**), kde z více než 776 registrovaných míst jich 54% obsahuje eVŠKP, 42% nepublikované zprávy a formuláře a 40% články z konferencí a workshopů.²³ Ve Francii obsahují $\frac{3}{4}$ otevřených repozitářů šedou literaturu.

Princip volného přístupu

Adresář OpenDOAR demonstruje zásadní posun od tištěných k digitálním sbírkám, které jsou produkovány v rámci souvislého ekonomického modelu a distribučními cestami šedé literatury.

Šedá literatura je definována svými nekomerčními cestami šíření. S rozvojem iniciativy *Open Access* (dále jen OA) jsou dokumenty šedé literatury logicky ukládány v těchto nových repozitářích, zvláště institucionálních (Schöpfel et al., 2009), ale též předmětových nebo jiných typech otevřených archivů.

Průběžný průzkum v letech 2005 - 2009 popisuje, jak pět mezinárodních center vědeckotechnických informací přijalo strategii otevřeného publikování

¹⁸ Viz např. zpráva FERMI z roku 1947 „The Future of Atomic Energy”, dostupná na <http://www.osti.gov/accomplishments/documents/fullText/ACC0043.pdf>

¹⁹ <http://lara.inist.fr/>, viz též Stock et al. (2006)

²⁰ <http://www.numdam.org/>

²¹ <http://cedadocs.badc.rl.ac.uk/>

²² Scientific a Technology Facilities Council, <http://www.scitech.ac.uk/Home.aspx>

²³ Informace platná ke květnu 2010.

v různých prostředích, s různými cíli a s větším či menším úspěchem (Boukacem-Zeghmouri et al., 2006; Schöpfel et al., 2009). Celkový počet položek volně dostupných v OA je těžké odhadnout; může se jednat o asi 3,5 až 4 miliony položek s významným podílem šedých materiálů. Tento zvláštní materiál se typicky vyznačuje četností položek a malou poptávkou. Luzi et al. (2008) popisuje přípravu institucionálního repozitáře italskou *Consiglio Nazionale delle Ricerche* (dále jen CNR), kde alespoň jedna třetina položek šedé literatury (zprávy, ústní prezentace, diplomové práce atd.) v existujícím úložišti CNR je otevřeně přístupná.

Otevřený archiv může být pro tento typ materiálů nejlepším řešením z důvodu nízkých nákladů na získávání, správu, uchovávání a distribuci. To je však pouze předpoklad nepodpořený empiricky, protože není dosud znám žádný ekonomický nebo finanční důkaz.

Budou v blízké budoucnosti všechny dokumenty šedé literatury dostupné na webových stránkách otevřených archivů? Navzdory tvrzení Willinskeho (2006), že „otevřený přístup je veřejně prospěšný“ a že „oddání se vědecké práci s sebou nese závazek distribuovat tuto práci co nejširšímu okruhu čtenářů“, se významná část šedé literatury pravděpodobně do otevřených archivů nedostane. A to z důvodu nedostatečného zájmu nebo rozpočtu pro digitalizaci starších tištěných materiálů, omezeného přístupu, nebo protože tyto položky jsou již dostupné na jiných stránkách (např. osobních stránkách, stránkách institucí s odkazy na soubory PDF apod.), i když ne vždy jsou dobře indexovány.

Podíl dokumentů šedé literatury publikovaných na internetu stále vzrůstá. Tento vývoj je úzce spjat s produkcí šedé literatury v elektronickém prostředí a s retrospektivními aktivitami, vedoucími k opětovnému vydání. Díky internetu je dostupné mnohem větší množství zdrojů šedé literatury (např. základní výsledky výzkumu, poznámky a osobní komentáře, přednášky, bulletiny, produktové katalogy apod.).

Nové informační a komunikační technologie obecně usnadňují archivování zdrojů a hnutí pro volný přístup k tomu dodává silné podněty. Otázky, kdo by měl co archivovat, kde, kdy a na jak dlouho, zůstávají však většinou nezodpovězeny. Z hlediska informační politiky a souvisejících finančních aspektů jsou odpovědi na tyto otázky velmi potřebné, i když se třeba týkají pouze části zdrojů šedé literatury (Schöpfel et al., 2010).

Od sbírky k otevřenému portálu

„Knihovna je sbírkou zdrojů, pramenů a služeb (...); je organizována pro použití a spravována (...).“²⁴ Může být otevřený repozitář nazván sbírkou? Je součástí knihovních sbírek?

²⁴ „Knihovna“ Wikipedie, otevřená encyklopedie (záznam z 13. března 2010).

Toto jsou pravděpodobně všechno otázky z minulosti. Potřebujeme vůbec v přicházející éře Googlu dobře strukturované a spravované sbírky šedé literatury? Nebo potřebujeme nástroje pro vyhledávání, stahování a přístup k šedým dokumentům? Lze si sbírky šedé literatury představit jako určitou globální síť?

Místo odpovědi bychom čtenáři rádi představili některé současné výsledky, produkty a služby.

Před několika lety ukončila *Royal Netherlands Academy of Arts and Sciences* (dále jen KNAW) všechnu pořizovací činnost v *Institute for Scientific Information Services* (dále jen NIWI), dříve jednoho z největších distributorů dokumentů. Místo toho investovala do tvorby nového portálu *National Academic Research and Collaborations Information System* (dále jen NARCIS), který zprostředkovává otevřený přístup z holandských vysokých škol a z výzkumných institucí k publikacím a souborům dat (např. popisům výzkumných projektů, institucí a pracovníků a k výzkumným zprávám). V tomto prostředí se hranice mezi šedou a bílou, tj. komerční literaturou, postupně smazává.

*National Documentation Centre*²⁵ (dále jen EKT) v Athénách spravuje databázi *Hellenic Dissertations*²⁶ odkazující na 13 000 prací z řeckých vysokých škol.

Knihovnický vysokoškolský portál *Die Eidgenössische Technische Hochschule Zürich* (dále jen ETH) zprostředkovává přístup ke 2,1 milionům zpráv z dalších knihoven, k databázím a vyhledávacím nástrojům.

Irish Virtual Research Library and Archive je určena k realizaci latentního potenciálu archivních sbírek na *University College of Dublin*.

Portál *Academic Archive On-line*²⁷ (dále jen DiVA) zpřístupňuje 270 000 výzkumných publikací a studentských prací z 27 švédských a norských vysokých škol a institucí vyššího vzdělání u něhož 44% obsahu tvoří šedá literatura.

Scirus, volně dostupný akademický vyhledávač od společnosti *Elsevier*, indexuje přes 30 otevřených repozitářů označených jako upřednostňované elektronické zdroje, které zahrnují evropské instituce a šedou literaturu.

Portál eVŠKP DART-Europe²⁸ zpřístupňuje více než 130 000 plnotextových výzkumných prací z 233 vysokých škol z 16 evropských zemí.²⁹ DART-Europe je partnerem výzkumných knihoven a knihovnických konzorcií pro evropskou pracovní skupinou *Networked Digital Library of Theses and Dissertations* (dále jen ND LTD). ND LTD zajišťuje přístup k téměř 750 000 pracím.

²⁵ <http://www.ekt.gr/en/index.html>

²⁶ http://argo.ekt.gr/opac2/Help/Databases/ENU/01_PhD_EN_.html

²⁷ <http://www.diva-portal.org/>

²⁸ <http://www.dart-europe.eu/>

²⁹ Informace k březnu 2010.

Dalším současným nástrojem pro vyhledávání zdrojů je německý projekt *Akademisches Publikationsmanagement*³⁰ (dále jen PUMA) pro spravování akademických publikací.

To bylo pouze několik ilustrací. Je nemožné uvést vyčerpávající seznam všech evropských iniciativ. Jejich pojetím je snaha o shromažďování položek, která byla nahrazena konceptem přístupu. Tato funkce portálů většinou zůstává v rámci knihovny, ovšem na scénu vstupují další hráči, jako např. nakladatelé, vyhledávače, výpočetní centra apod. Tyto subjekty se zaměřují na výběr, šíření a přístup, ale nikdy nespravovaly knihovnické sbírky, kde je naopak kladen důraz na uchovávání a organizaci.

Roosendaal et al. (2010) popisuje velmi jasně dynamiku tohoto nového publikačního vzoru a souvisejícího ekonomického modelu. Zřejmými výhodami je zpracování velkého množství informací, jediné přístupové místo, nelokalizovaný přístup, účinné nástroje pro vyhledávání a výběr.

Ve studii Stockové (2007) o evropských **eVŠKP** byly uvedeny některé problémy otevřených repozitářů, kterými byly částečný nebo omezený přístup k plnému textu, záznamy bez plného textu, chybějící nebo nekompletní metadata a jazykové bariéry. Dalšími problémy jsou absence norem a součinnosti.

Podle současných zkušeností se prohledávání a sběr šedé literatury nikdy nestane tak přímočarým jako u časopisů a knih tradičního nakladatelského sektoru. Nové nástroje pro sbírání a ukládání neučiní šedou literaturu méně pomíjivou a přechodnou než dosud. Výzkum ukazuje, že dokud daná organizace nezformuluje politiku týkající se šedé literatury, která bude podpořena dotací z rozpočtu, nebude možné zaručit implementaci technologie. V důsledku toho bude prostředí, v němž šedá literatura koexistovala dodnes, v nejbližší budoucnosti patrně stále nestabilní (Schöpfel et al., 2010).

Od knihovny k eScience

Výzkumné prostředí se vyvíjí a stává se stále více závislé na datech, s rostoucími potřebami jejich získávání, uchovávání, zpracovávání, spravování atd. Objevují se již nové služby pro integraci dat, které přeměňují vyhledávání dat na internetu z pouhých seznamů výsledků na sofistikované nástroje, formulující odpovědi na strukturované otázky (Fry, 2009). Jak ale podotýká Osswald (2008), dosud hrály odborné knihovny pouze malou, pokud vůbec nějakou, roli v projektech **eScience**³¹ implementovaných v EU.

Přístup k výsledkům výzkumu, k publikacím i k datům, je „poslední klíčovou složkou výzkumné infrastruktury (...). Tak posune revoluce eScience knihovny a repozitáře do centra zájmu ve vývoji výzkumné infrastruktury příští generace” (Hey et al., 2006).

³⁰ <http://puma.uni-kassel.de/>

³¹ <http://en.wikipedia.org/wiki/E-Science>

Portály jako **NARCIS** již obsahují ve zdrojové nabídce soubory dat. To je zcela přirozené s ohledem na nejnovější vývoj v oblasti akademických publikací, které představují dynamické publikace, 3D ilustrace, primární soubory dat obsažené v časopisech apod. Jaká je ale „barva“ těchto souborů dat? Jsou součástí odborné literatury? Nebo časem nahradí, alespoň částečně, vědecké publikování?

V dnešní době je koncept „článku budoucnosti“³² v centru vědecké a odborné diskuze. Komerční nakladatelé intenzivně investují do pokročilého editačního softwaru, aby mohli integrovat data do publikací.

Co se děje v tomto prostředí s šedou literaturou? Zatímco uvádění základních dat je v časopisech relativně novým jevem, u diplomových prací a zpráv není doplňkový materiál ničím novým. Jejich součástí jsou běžně CD-ROMy, mapy, tabulky či rozsáhlé datové přílohy.

Šedá literatura dodává základní materiál pro vytěžování dat a služby alert, které upozorňují na změny. Sledování abstraktů z farmakologických konferencí má např. význam pro ekonomické zpravodajství (např. pro analýzu průmyslových trendů) a průzkum oblastí popisujících stav techniky a bibliografií disertačních prací zase napomáhá scientometrii.

Otázkou není, zda má šedá literatura souvislost s **eScience**, ale jak lze co nejlépe využít odborné informace obsažené v diplomových pracích, zprávách, krátkých člancích, formulářích apod. Jedním z řešení je vytvoření velkých repozitářů dat vědeckými komunitami ve spolupráci s jejich knihovnami a vývoj nových modelů publikace dat. Osswald (2008) varuje, že knihovny mohou ztratit významnou část svých úkolů v odborné společnosti, pokud se nepokusí zapojit se do projektů eScience. Toto riziko je reálné. Některé současné iniciativy však dávají naději, že se knihovny stanou součástí vznikající vědecké kybernetické infrastruktury. Nejslibnějším evropským projektem se v současnosti jeví DataCite,³³ který „podporuje sdílení dat, usnadněný přístup a lepší ochranu investic do výzkumu“ (DataCite, c2010).

Dalším krokem by mělo být propojení mezi volně přístupnými daty a publikačními archivy, vědeckou komunitou a institucemi,³⁴ pokud chtějí omezit kontrolu vědeckých výsledků komerčními nakladateli a globálními informačními společnostmi.

Soubory dat jsou každopádně tvrdým oříškem pro nakladatele a knihovny v oblasti ověřování a uchovávání. Možná je jejich opravdové místo mimo komerční distribuční cesty a ne ve „článku budoucnosti“. Zítra budeme mít možná ne jeden, ale mnoho informačních systémů typu NARCIS a snad i unikátní portály umožňující přístup a propojení holandských, britských, německých, švédských a českých souborů dat a publikací. Je třeba snít.

³² Viz Elsevier „Article of the Future“ initiative, <http://beta.cell.com/>

³³ Project leader: TIB Hannover, <http://www.datacite.org/>

³⁴ Například v aktuální verzi výzkumného informačního systému (CRIS).

Budoucnost sbírek šedé literatury

Tato kapitola má za cíl dát čtenáři představu o rozmanitosti a dynamičnosti evropské šedé literatury. Na několika stránkách je samozřejmě nemožné uvést všechny významné sbírky, jako např. zvláštní sbírku více než 60 000 vzácných publikací a samizdatové literatury *Jagellonské knihovny v Krakově* nebo soubor 15 000 digitálních map v *Institut Cartogràfic de Catalunya* (dále jen ICC) v Barceloně. Čtenář najde odkazy na další zdroje na stránkách různých sítí KIV, např. **GreyNet** nebo *Ligue des Bibliothèques Européennes de Recherche*³⁵ (dále jen LIBER) včetně diskusních skupin o digitalizaci, získávání zdrojů, sbírkách kulturního dědictví a jeho ochraně.

V probíhající diskuzi o nových ekonomických modelech akademického publikování, eScience a volném přístupu k veřejným výsledkům výzkumu budou nekomerční distribuční cesty stále hrát hlavní roli jako vektory vědecké komunikace, spolu s komerčním publikováním.

Otevřené archivy budou nabízet stále více odpovídajících služeb a funkcí pro alespoň některé části šedé literatury, ne-li pro všechny. Bibliografická kontrola šedé literatury bude však i nadále problematická i navzdory trendu standardizace digitálních dokumentů. A knihovny spolu s vědeckými komunitami budou muset nalézt nové podoby základních funkcí odborného publikování, které budou brát v potaz i otevřené repozitáře, nekomerční materiály a soubory dat.

Tato kapitola pojednává o evropské šedé literatuře a autor sám je velkým příznivcem evropské myšlenky. Ale filozofie internetu nehledí na hranice, národy a nadnárodní struktury. Problém spočívá v jazykových bariérách, metadatech a formátech a proto je tato kapitola zakončena apelem na potřebu součinnosti a norem.

LITERATURA:

SCHÖPFEL, J. - FARACE, D., J. „*Grey literature*,“ *Encyclopedia of Library and Information Sciences*, Third Edition [online]. M. J. Bates and M. N. Maack, Eds. CRC Press, 2010, s. 2029 - 2039. Dostupné z [www](http://dx.doi.org/10.1081/E-ELIS3-120043732):
<http://dx.doi.org/10.1081/E-ELIS3-120043732>

BOUKACEM-ZEGHMOURI, C. - SCHÖPFEL, J., *Document supply and open access: an international survey on grey literature* [online]. Interlending & Document Supply, 2006, sv. 34, č. 3, s. 96-104. Dostupné z [www](http://archivesic.ccsd.cnrs.fr/sic_00181485/en/):
http://archivesic.ccsd.cnrs.fr/sic_00181485/en/

³⁵ <http://www.libereurope.eu/>

- FARACE, D. - FRANTZEN, J. - STOCK, C. - HENROT, N. - SCHÖPFEL, J., *OpenSIGLE, Home to GreyNet's Research Community and its Grey Literature Collections: Initial Results and a Project Proposal* [online]. In GL10 Conference Proceedings. Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society. Amsterdam, 8.-9. prosince 2008, 2008. Dostupné z www.: http://archivesic.ccsd.cnrs.fr/sic_00379643/en/
- HEY, T. - TANSLEY, S. - TOLLE, K. (Eds.) *The fourth paradigm. Data-intensive scientific discovery*. Microsoft Corporation, 2009.
- HEY, T. - HEY, J., *E-science and its implications for the library community* [online]. Library Hi Tech, 2006, sv. 24, č. 4, s. 515 - 528. Dostupné z www: <http://dx.doi.org/DOL%20%2010.1108/07378830610715383>
- LUZI, D. - Di CESARE, R. - RUGGIERI, R. - CERBARA, L., *Towards an institutional repository of the italian national research council: A survey on open access experiences* [online]. In Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society, 8.-9. prosince 2008, 2008. Dostupné z www: <http://opensigle.inist.fr/bitstream/10068/698002/2/GL10%2^LuzLe^al%2^Confere^Preprin1>
- MEYER, T., *Die Zentralen Fachbibliotheken und ihre Rolle für die Fachinformation in Deutschland*. Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, 2009, sv. 248.
- OSSWALD, A., *E-science and information services: a missing link in the context of digital libraries* [online]. Online Information Review, 2008, sv. 32, č. 4, s. 516 - 523. Dostupné z www: <http://dx.doi.org/10.1108/14684520810897395>
- PAILLASSARD, P. - SCHÖPFEL, J. - STOCK, C., *Dissemination and preservation of French print and electronic theses* [online]. The Grey Journal, 2007, sv. 3, č. 2, s. 77 - 93. Dostupné z www: http://archivesic.ccsd.cnrs.fr/sic_00380488/en/
- ROOSENDAAL, H., E. - ZALEWSKA-KUREK, K. - GEURTS, P. A. - HILF, E. R. *Scientific Publishing: From vanity to strategy* [online]. Chandos Publishing, 2010. Dostupné z www: <http://dx.doi.org/10.3152/030234207X244810>
- SCHÖPFEL, J. - PROST, H., *Document supply of grey literature and open access: an Update* [online]. Interlending & Document Supply, 2009 sv. 37, č. 4, s. 181-191. Dostupné z www: <http://dx.doi.org/10.1108/02641610911006274>
- SCHÖPFEL, J. - BOUKACEM-ZEGHMOURI, C. - PROST, H., *Usage of grey literature in open archives*. V GL11 Conference Proceedings. Eleventh International Conference on Grey Literature: The Grey Mosaic: Piecing It All Together. Washington D.C., 14-15 December 2009, prosinec 2009.
- SCHÖPFEL, J. - STOCK, C. - HENROT, N., *From SIGLE to OpenSIGLE and Beyond: An In-depth Look at Resource Migration in the European Context* [online]. The Grey Journal, 2007, sv. 3, č. 1, s. 45-51. Dostupné z www: http://arcmvesic.ccsd.cnrs.fr/sic_00181592/en/

- SMITH, V., *Data publication: towards a database of everything* [online]. BMC Research Notes, 2009, sv. 2, č. 1, s. 113+. Dostupné z www: <http://dx.doi.org/10.1186/1756-0500-2-113>
- STOCK, C. - ROCKLIN, E. - CORDIER, A. *Lara—open access to scientific and technical reports* [online]. Publishing Research Quarterly. Květen 2006, sv. 22, č. 1, s. 42-51. Dostupné z www: <http://dx.doi.org/10.1007/s12109-006-0007-3>
- STOCK, C., *Open access to full text and etds in europe: improving accessibility through the choice of language?* [online]. In. Ninth International Conference on Grey Literature: Grey Foundations in Information Landscape, 10-11 December 2007, 2007. Dostupné z www: <http://opensigle.inist.fr/handle/10068/697889>
- TILLET, S. - NEWBOLD, E., *Grey literature at the British library: revealing a hidden resource* [online]. Interlending & Document Supply, 2006, sv. 34, č. 2, s. 70-73. Dostupné z www : <http://dx.doi.org/10.1108/02641610610669769>
- WILLIMSKY, J., *The Access Principle: The Case for Open Access to Research and Scholarship (Digital Libraries and Electronic Publishing)* [online]. The MIT Press, prosinec 2005. Dostupné z www : <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0262232421>

ELEKTRONICKÉ ZDROJE:

- ANRTheses [online]. [cit. 2010-06-10]. Dostupné z www: <http://www.anrtheses.com.fr/>
- CEDA Repository [online]. [cit. 2010-06-10]. Dostupné z www: <http://cedadocs.badc.rl.ac.uk/>
- DART-Europe E-theses Portal [online]. The DART-Europe E-theses Portal is endorsed by LIBER. c1999-2010. [cit. 2010-06-10]. Dostupné z www: <http://www.dart-europe.eu/basic-search.php>
- DataCite [online]. DataCite. c2010. [cit. 2010-08-27]. Dostupné z www: <http://www.datacite.org/>
- DIVA. Academic Archive [online]. Uppsala University. c2000-2010. [cit. 2010-06-10]. Dostupné z www : <http://www.diva-portal.org/>
- DOE [online]. U.S. Department of Energy. [cit. 2010-08-27]. Dostupné z www: <http://www.energy.gov/>
- EThOS - Beta. Eletronic Theses Online Service [online]. The British Library Board. [cit. 2010-06-10]. Dostupné z www: <http://ethos.bl.uk>
- GEPRIS [online]. Deutsche Forschungsgemeinschaft. [cit. 2010-06-10]. Dostupné z www: <http://gepris.dfg.de>
- INIST-CNRS [online]. [cit. 2010-06-10]. Dostupné z www: [<http://international.inist.fr/>](http://international.inist.fr/)

JISC [online]. JISC Comms. Aktualizováno 25. 8. 2010. [cit. 2010-08-27]. Dostupné z www: <http://www.jisc.ac.uk/digitisation>

LARA [online]. INIST - CNRS. C2002-2005. [cit. 2010-08-27]. Dostupné z www: <http://lara.inist.fr/>

NUMDAM [online]. Recherche et téléchargement d'archives de revues mathématiques numérisées. 2006 [cit. 2010-06-10]. Dostupné z www: <http://www.numdam.org/>

OpenSIGLE [online]. INIST-CNRS. [cit. 2010-06-10]. Dostupné z www: <http://opensigle.inist.fr/>

PUMA [online]. Deutsche Forschungsgemeinschaft. [cit. 2010-06-10]. Dostupné z www: <http://puma.uni-kassel.de/>

SUDOC [online]. ABES/OCLC. c2010. [cit. 2010-08-27]. Dostupné z www: <http://www.sudoc.abes.fr>

TDX [online]. [cit. 2010-08-27]. Dostupné z www: <http://www.tesisenxarxa.net/>

TIB [online]. German National Library of Science and Technology. Aktualizováno 26. 3. 2010. [cit. 2010-08-27]. Dostupné z www: <http://www.tib-hannover.de>

3. Vývoj šedé literatury v České republice

Petra Pejšová

Vývoj šedé literatury

Vývoj šedé literatury v České republice lze v současné době rozdělit do tří období s dvěma významnými milníky, kterými byly roky 2005 a 2008. Do roku 2005 byl sběr šedé literatury nastaven na podporu evropské spolupráce řízenou asociací EAGLE. V letech 2005 až 2008 nefungoval evropský systém a v České republice představuje toto období čas příprav, ale zároveň propast v centralizovaném sběru šedé literatury. V roce 2008 se rozběhly dva dlouho připravované významné projekty Národní úložiště šedé literatury³⁶ (dále jen **NUŠL**) a *Národní registr VŠKP a systém na odhalování plagiátů* v systému **theses.cz**³⁷ (dále jen **theses.cz**).

První organizovaný sběr šedé literatury započal v České republice počátkem 90. let 20. století. Impulesem k této činnosti byl zájem asociace **EAGLE** sdružit instituce podchycující šedou literaturu z evropských zemí. Zástupci za Českou republiku v EAGLE se staly *Národní technická knihovna* (dále jen **NTK**)³⁸ a *Knihovna Akademie věd ČR* (dále jen **Knihovna AV ČR**).

Jak již bylo uvedeno v předchozí kapitole, EAGLE provozovala systém **SIGLE**, díky němuž bibliograficky podchycovala produkci šedé literatury v Evropě a zpřístupňovala ji uživatelům. NTK se stala členem EAGLE v roce 1995, ale již s předstihem v průběhu roku 1994 vytvořila *Kooperační systém šedé literatury* (dále jen **KSSL**). Tento systém byl založen na smluvním základě a přebíral bibliografické záznamy o šedé literatuře (zejména o disertacích) od kooperujících českých vysokých škol. Převzatá metadata byla konvertována do specifického datového formátu **SIGLE** a v dávkách zasílána do *Fachinformationszentrum Karlsruhe* (dále jen **FIZ Karlsruhe**), kde byla data zpracovávána. Případně spolupráce probíhala tak, že vysoké školy zasílaly tištěné disertace do NTK, kde byly vytvořeny bibliografické záznamy těchto disertací dle pravidel **SIGLE** a opět v dávkách zasílány do **FIZ Karlsruhe**.

Vysoké školy měly možnost předávaná data vybírat dle relevance. Zdaleka tedy nešlo o úplnou evidenci produkce disertací v České republice. Role Knihovny AV ČR, která měla podchycovat šedou literaturu z produkce ústavů AV ČR, nebyla za celou dobu existence systému **SIGLE** naplněna. NTK byla jediným aktivním přispěvatelem a národním integrátorem za Českou republiku v EAGLE.

³⁶ <http://nusl.techlib.cz/>

³⁷ <http://www.theses.cz/>

³⁸ Do 1.7.2009 Státní technická knihovna (STK).

Česká republika byla s 5 778 záznamy, což byl pouze zlomek produkce, druhým nejaktivnějším státem v systému SIGLE ze zemí bývalého východního bloku. Příčinou nízkého počtu byly jednak poměrně striktní požadavky SIGLE na jazykové vybavení dokumentu i záznamu a zcela dobrovolná účast českých producentů v systému. Nejvíce záznamů celkem 19 524 dodala do systému SIGLE Ruská federace, na třetím místě bylo s 3 961 záznamy Lotyšsko a na čtvrtém místě s 850 záznamy Slovenská republika (OpenSIGLE, 2010).

Silnou stránkou systému SIGLE byl relativně propracovaný systém věcného třídění SIGLE Subject Category List (Novák, 2008). Slabými stránkami systému SIGLE byl jednak specifický metadatový formát, který byl nekompatibilní s **formáty MARC**, ale především absence přímé vazby na získání plného textu dokumentu. Většina členských organizací tento nedostatek řešila pomocí služeb dodávání dokumentů (document delivery services). Tato infrastruktura však v České republice neexistovala, a proto se požadavky na výpůjčky řešily cestou standardní *Mezinárodní meziknihovní výpůjční službou* (dále jen MMVS) bez jakékoli zjištěné vazby na systém KSŠL/SIGLE.

V posledních letech (2003 - 2005) nedokázala EAGLE zajistit systému SIGLE odpovídající vývoj směrem k modernímu metadatovému formátu, síťové spolupráci a především k přímému elektronickému zpřístupnění dokumentů. Systém SIGLE se tak stal zastaralým a tato situace vedla klíčové členy EAGLE k vystoupení z asociace a následně k zániku celé asociace.

Bezprostředně po ukončení činnosti EAGLE ukončila NTK také český KSŠL. Zároveň vznikla idea vybudování národního centra pro šedou literaturu se zaměřením i na další typy šedé literatury, jejich zpracování a zpřístupnění v elektronické podobě. V roce 2005 iniciovala NTK několik jednání se zástupci vysokých škol, ústředních knihoven a dalších institucí, která se zabývala možnostmi spolupráce v oblasti šedé literatury v České republice. Pro NTK z těchto jednání vyplynul závěr, že zde byl zájem o tuto činnost a začala s přípravou národního centra pro šedou literaturu. Zároveň se NTK aktivně připojila k *Odborné komisi pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací AKVŠ* (dále jen Komise eVŠKP).

Tyto aktivity se ovšem nerozběhly tak rychle, jak se předpokládalo. Přípravná fáze nakonec trvala tři roky, až do roku 2008. V tomto období stále díky setrvačnosti byly zpracovávány do knihovního katalogu NTK bibliografické záznamy o disertačních pracích zasílaných vysokými školami a další šedá literatura běžně získávaná do fondu NTK, a to především konferenční materiály, závěrečné a výroční zprávy a firemní literatura. Současně se systematicky zpracovávala šedá literatura v rámci kooperačního systému ústavů AV ČR v *Automatizovaném systému evidence publikací* (dále jen **ASEP**) a to konferenční materiály, výzkumné zprávy, disertace a interní tisky. Další druhy šedé literatury systém ASEP nepodchycoval. Veškeré výstupy z dotačních programů, především závěrečné zprávy, archivovaly jednotlivé grantové agentury a ministerstva. Bohužel již většinou nebylo zajištěno zpřístupnění zpráv veřejnosti. Šedá literatura byla také částečně podchycována v institucionálních repozitářích či osobních archivech významných osobností,

částečně i v katalozích především odborných knihoven, jednalo se zejména o disertace, výroční zprávy a sborníky.

Velmi palčivým problémem byla v tomto období archivace elektronických vysokoškolských kvalifikačních prací (dále jen **eVŠKP**) na jednotlivých vysokých školách. Na podporu řešení tohoto problému vznikla v rámci AKVŠ již v roce 2004 Komise eVŠKP, která se zaměřila na přípravu standardů a kooperaci činností směřujících k vybudování repozitáře, který by sloužil jako národní registr eVŠKP. V oblasti archivace a zpřístupnění eVŠKP byla nejaktivnější Masarykova univerzita v Brně, která zároveň začala řešit i problematiku plagiátorství. Opomíjené zůstaly na vysokých školách další typy šedé literatury, které vznikaly jak díky rozsáhlé publikační činnosti samotných pracovníků (studijní materiály, příspěvky do sborníků), tak v rámci vědecké a výzkumné práce jednotlivých vysokoškolských pracovišť v podobě technických a výzkumných zpráv.

Téměř nepodchycená zůstávala šedá literatura publikovaná v dalších výzkumných organizacích, jako byly podnikové či soukromé výzkumné ústavy, dále v institucích státní správy a samosprávy (ministerstva, kraje, státem zřizované organizace). Do této sféry šedé literatury patřila též z velké části firemní literatura, pravidelně či nepravidelně vydávané publikace typu konferenční sborníky, katalogy, programy apod., které vznikaly k pořádaným konferencím, výstavám či workshopům. Tato šedá literatura se jen částečně dostávala do sbírek knihoven. Významnými výjimkami bylo budování fondu firemní literatury např. v NTK a Zemědělské a potravinářské knihovně. Zajímavé je porovnat tuto situaci se zahraničím, která je popsána v druhé kapitole v části Strategie získávání dokumentů.

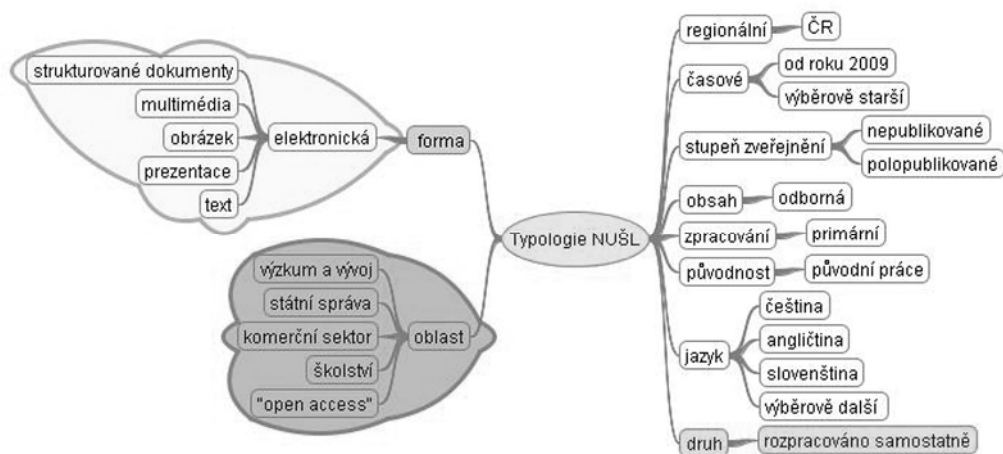
V roce 2008 po období příprav započala systematická práce na centrálním sběru šedé literatury v České republice, a to hned dvěma projekty. Prvním je projekt **NUŠL** pod vedením NTK ve spolupráci s Vysokou školou ekonomickou v Praze (dále jen **VŠE**). Tomuto projektu, jakožto komplexnímu národnímu řešení s návazností na zahraničí, je v této kapitole věnován náležitý prostor. Druhým projektem je systém theses.cz, který začala řešit Masarykova univerzita v Brně ve spolupráci s dalšími vysokými školami. Systém theses.cz centralizovaně shromažďuje eVŠKP, nad nimiž provozuje systém na odhalování plagiátů. Více o tomto projektu stejně tak jako problematice šedé literatury na vysokých školách se věnuje šestá kapitola.

Pilotní projekt NUŠL

Projekt NUŠL se začal realizovat díky podpoře *Ministerstva kultury České republiky* (MK ČR) v rámci programových projektů výzkumu a vývoje jako výzkumný záměr pod názvem *Digitální knihovna šedé literatury - funkční model a pilotní realizace*. Tento projekt je rozdělen do tří etap, které probíhají od roku 2008 do roku 2011. Jeho hlavním cílem je systematicky shromažďovat, dlouhodobě archivovat a zpřístupňovat odbornou šedou literaturu především z oblasti výzkumu a vývoje, státní správy a školství, ale i z komerčního sektoru

a „open access“ na národní úrovni. Na podporu tohoto cíle NTK vytváří fungující síť partnerských organizací, funkční model a pilotní aplikaci. Dále na základě ověřené technologie a metod definovaných v rámci projektu vznikají doporučení a standardy pro další instituce, které se rozhodnou vytvářet digitální repozitáře šedé literatury. Doporučení a standardy představují zejména doporučený metadatový formát, výměnné formáty a šablony, vzory licenčních modelů a řešené právní problematiky, metodiku ochrany, archivace a zpřístupňování digitálních dat.

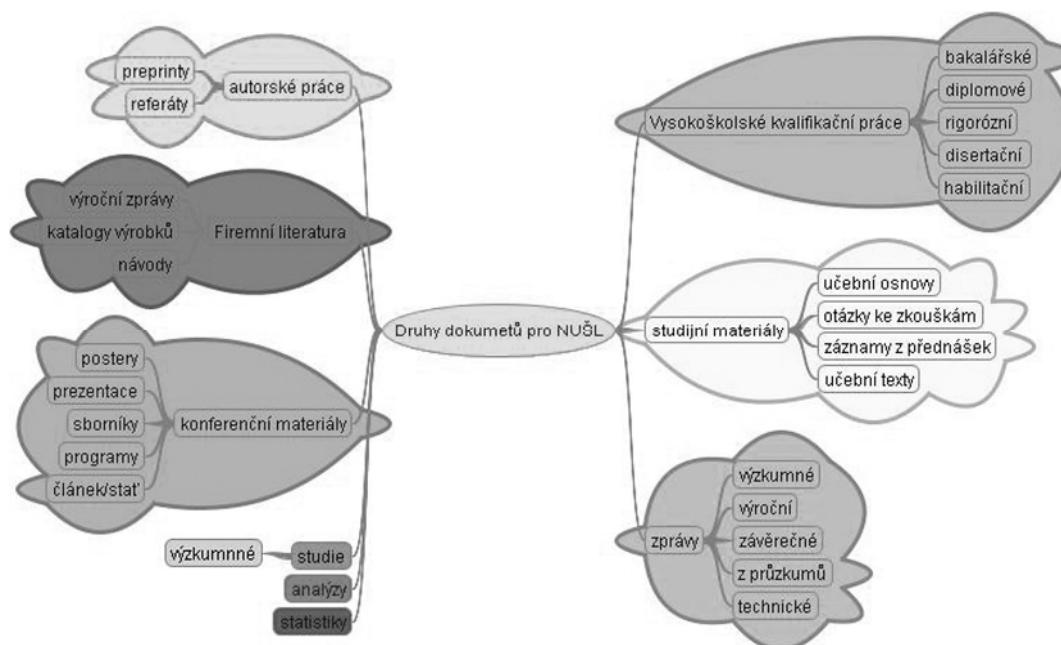
Před samotným sběrem šedé literatury do NUŠL bylo nutné definovat typologii, která by jasně vymezila předmět sběru. Většina typologií včetně typologie GreyNet se zaměřuje pouze na druh dokumentu. V případě NUŠL byla vytvořena vlastní typologie, která zohlednila i další kritéria. Typologie NUŠL je strukturována do dvou úrovní, aby mohla být lépe vyjádřena jednotlivá hlediska. Základní typologie (první úroveň) znázorněná níže myšlenkovou mapou popisuje veškerá kritéria pro dokumenty, které jsou přijímány do NUŠL. Mezi základní kritéria patří sběr dokumentů v elektronické formě od roku 2009 a výběrově starších. Elektronickou formu představují především „born digital“ nebo digitalizované dokumenty. Co se týče metadat, jsou sbírány všechny dostupné záznamy o šedé literatuře vzniklé v České republice bez omezení. Dodávky šedé literatury jsou zajišťovány od partnerských organizací z oblasti školství, výzkumu a vývoje, státní správy, komerčního sektoru a ověřených „open access“ zdrojů. Původem vzniku šedé literatury v NUŠL musí být Česká republika a její obsah musí být původní a odborný. Preferovaný jazyk práce je český, anglický, slovenský a výběrově další evropské jazyky.



Obrázek 1: Základní typologie NUŠL (první úroveň)

Ze základní typologie je dále rozpracována charakteristika druhu dokumentu (druhá úroveň), kde jsou hlavními skupinami vysokoškolské kvalifikační práce, studijní materiály, zprávy, autorské práce, firemní literatura, konferenční materiály a další. Jednotlivé skupiny již obsahují konkrétní druhy dokumentů. Předmětem projektu je též ověřit, které z druhů dokumentů budou početnější,

a rozšířit tuto typologii o nové druhy dokumentů potřebné pro vědu a vzdělávání.

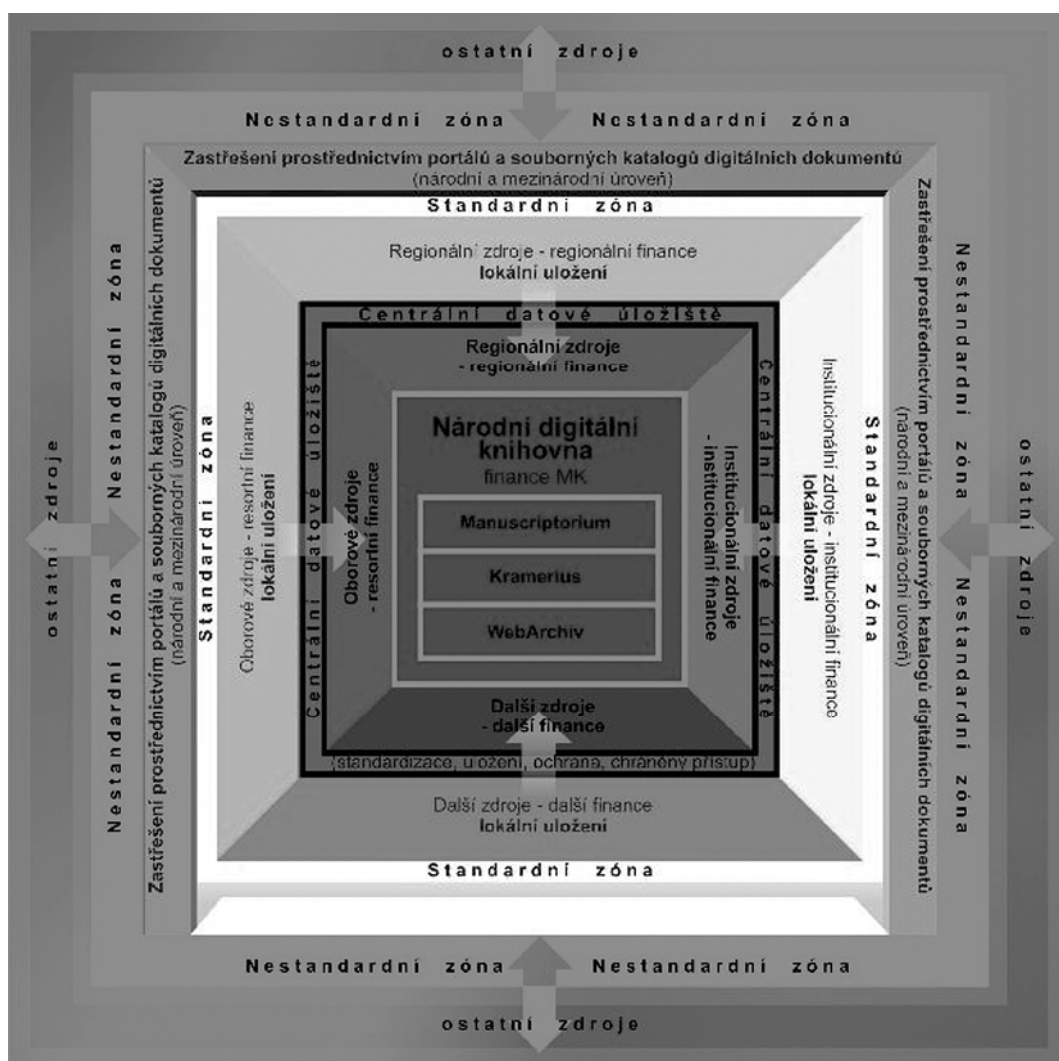


Obrázek 2: Typologie dle druhu dokumentu (druhá úroveň)

Návaznosti a propojení systémů šedé literatury

Záměrem NUŠL není budování izolovaného řešení, ale právě naopak je jím snaha doplňovat a navazovat na další systémy či projekty, které řeší dlouhodobou archivaci a zpřístupňování digitálních dokumentů v České republice i v zahraničí. Význam spolupráce v síťovém prostředí je zdůrazněn v úvodu druhé kapitoly, ve které následuje množství příkladů o budování repozitářů či portálů šedé literatury.

Jedním z projektů, na který NUŠL navazuje, je *Česká digitální knihovna* (dále jen ČDK). *Národní knihovna České republiky* (dále jen NK ČR) vytvořila koncepci ČDK - viz Obrázek č. 3. Jádrem České digitální knihovny je *Národní digitální knihovna* (dále jako NDK), která je uprostřed obrázku zobrazena červenou barvou. NDK se orientuje pouze na publikované dokumenty v rámci ochrany kulturního dědictví České republiky. NUŠL naopak podchycuje dokumenty oficiálně nepublikované či polopublikované. V základním funkčním schématu koncepce ČDK se NUŠL nachází v tzv. standardní zóně, která předpokládá zastřešení prostřednictvím portálů a souborných katalogů digitálních dokumentů. NUŠL je jedním z nich. Samotný NUŠL přitom sklízí metadata a plné texty z lokálních úložišť oborových, institucionálních a dalších zdrojů v šedé zóně. NTK tím přebírá zodpovědnost za dlouhodobou archivaci a zpřístupnění sklizených dat.



Obrázek 3: Česká digitální knihovna (Portál Národní digitální knihovny, 2006) - šedá literatura se nachází ve standardní zóně

Významnou oblastí produkce šedé literatury jsou také pracoviště **AV ČR**. V AV ČR není šedá literatura sledována samostatně, je vymezena spolu s ostatními výsledky vědecké činnosti, které vykazují jednotlivá pracoviště v rámci hodnocení vědy (související s dalším financováním těchto vědecko-výzkumných institucí). Tyto výsledky vědecké činnosti jsou sledovány prostřednictvím systému **ASEP**, který spravuje Knihovna AV ČR. V rámci budování partnerské sítě spolupracujících organizací byla jak s ústavu AV ČR tak s Knihovnou AV ČR navázána úzká spolupráce.

Dalším projektem, na který by řešitelé NUŠL rádi navázali, je systém theses.cz. Přínosné by bylo přebírat sjednocená data přímo ze systému theses.cz namísto přebírání dat z jednotlivých lokálních repozitářů do digitálního repozitáře NUŠL, který zaručuje dlouhodobou archivaci dat. Dalším možným způsobem spolupráce je připojit systém theses.cz jako zdroj do vyhledávacího

uživatelského rozhraní NUŠL bez nutnosti sklízet data přímo do digitálního repozitáře NUŠL. Vyhledané záznamy v NUŠL by v tomto případě odkazovaly na dokumenty přímo do systému theses.cz. Tento způsob spolupráce nezajistí dlouhodobou archivaci dat v digitálním repozitáři NUŠL. Ani jedno z řešení bohužel v současné chvíli není možné, protože systém theses.cz nemá implementován protokol **OAI-PMH** pro výměnu dat a umožňuje pouze prohlížení obsahu přes webové stránky.

Obecně možnostem spolupráce na národní úrovni a budování partnerské sítě v rámci České republiky se věnuje čtvrtá kapitola. Aby se nákladně nevyvíjela jednotlivá izolovaná řešení na národních úrovních, je v oblasti šedé literatury nezbytná mezinárodní spolupráce. A to jak v oblasti sdílení standardů a metodik při zpracování a archivaci, tak i v oblasti sjednoceného přístupu k samotné šedé literatuře.

V současnosti je navázána spolupráce s mezinárodní organizací pro šedou literaturu **GreyNet** se sídlem v Amsterdamu. GreyNet byl založen v roce 1992 za účelem pořádání série *International Conference on Grey Literature*³⁹ (dále jen GL). Cílem GreyNetu je usnadnit dialog, výzkum a komunikaci mezi lidmi a organizacemi v oblasti šedé literatury. Po zániku EAGLE se GreyNet snaží o obnovení jejích funkcí, především o znovuoobnovení kooperující sítě organizací zabývajících se šedou literaturou.

Jak je již zmíněno v druhé kapitole, **INIST** obnovil systém SIGLE, který je nyní znám pod názvem **OpenSIGLE**.⁴⁰ Nejprve v roce 2006 vznikl prototyp nového systému, založený na softwaru **DSpace**⁴¹ z oblasti Open Source. V této fázi byly převedeny záznamy z německého FIZ Karlsruhe do DSpace ve francouzském INISTu. Cílem bylo zajistit pokračování přístupu k již vloženým datům a umožnit vkládat nové záznamy včetně digitálních dokumentů. V závěru roku 2008 bylo prvním krokem v tomto směru vložení kolekce sborníků z GL za posledních 5 let (GL5 až GL9) do OpenSIGLE včetně plných textů a prezentací. V současnosti je řada sborníků doplněná tak, že obsahuje souvislou řadu GL1 až GL10. Další vývoj OpenSIGLE zajišťuje INIST, který vytváří technologické podmínky a vypracovává metodiky a standardy pro přijímání dat a zajištění linků na fulltexty a externí zdroje.

Po fyzickém vytvoření repozitáře NUŠL plánuje řešitelský tým v druhé polovině roku 2010 též jeho zapojení do projektu *Digital Repository Infrastructure Vision for European Research*⁴² (dále jen **DRIVER**) a **WorldWideScience.org**.⁴³ DRIVER je nadstavbou nad zapojenými evropskými institucionálními repozitáři typu „open access“. Jedná se o vyhledávací portál, který vyhledává v článcích, disertacích, knihách, zprávách apod. z více než 200 institucionálních

³⁹ <http://www.textrelease.com/textreleasehome.html>

⁴⁰ <http://opensigle.inist.fr/>

⁴¹ <http://www.dspace.org/>

⁴² <http://www.driver-community.eu/>

⁴³ <http://worldwidescience.org/>

a obsahově zaměřených repozitářů z 23 zemí Evropy a v 25 jazycích. Projekt WorldWideScience.org je globální bránou vědy, jejímž cílem je prostřednictvím mnohostranných partnerství umožnit federativní vyhledávání v národních a mezinárodních vědeckých databázích a portálech a zrychlit tak vědecký objev a pokrok.

Podpora spolupráce a odborné diskuse o šedé literatuře

V roce 2008 vznikly první české webové stránky plně se věnující šedé literatuře z tuzemska i ze světa. Na webových stránkách jsou průběžně aktualizovány informace o projektu NUŠL a zveřejňovány jeho výstupy. Tyto výstupy představují standardy (jako jsou metadatový formát, typologie, datové formáty, formát pro výměnu dat apod.) a metodiky (jako jsou návod k vyplňování polí metadatového formátu, lokální instalace digitálního repozitáře, správa sbírek a úpravy šablon v lokální instalaci apod.). Dále slouží k podpoře spolupráce NUŠL s producenty dat šedé literatury k níž patří legislativní problematika šedé literatury a vysvětluje zvolené softwarové řešení, jehož součástí je lokální instalace digitálního repozitáře zde dostupná ke stažení. Z důvodu širších souvislostí jsou zde uvedeny vybrané informační zdroje šedé literatury, dále odkazy na portály podporující výzkum a vývoj ve vědě v České republice.

NUŠL Národní úložiště šedé literatury

O PROJEKTU

Uživatelské rozhraní NUŠL je dostupné na www.nusl.cz
 Digitální repozitář NUŠL je dostupný na <http://invenio.ntkcz.cz>

Národní technická knihovna v Praze (NTK, příspěvková organizace přímo řízená Ministerstvem školství, mládeže a tělovýchovy ČR) v souladu se svým statutem řeší otázky **budování Národního úložiště šedé literatury**. V současnosti je tento projekt podporován Ministerstvem kultury v rámci projektu **"Digitální knihovna pro šedou literaturu - funkční model a pilotní realizace"**. Na tomto projektu Národní technická knihovna spolupracuje s Vysokou školou ekonomickou v Praze.

V minulosti byla NTK odpovědná za distribuci šedé literatury za Českou republiku do evropského systému pro šedou literaturu SIGLE (System for Information on Grey Literature in Europe), který provozovala Evropská asociace pro využívání šedé literatury (European Association for Grey Literature Exploitation - EAGLE). Po zániku tohoto systému v roce 2005 nedochází ke koordinované evidenci šedé literatury v České republice.

Hlavním cílem projektu je návrh na vytvoření Národního úložiště šedé literatury, které bude sloužit jako centrální rozhraní pro vyhledávání tzv. nepublikovaných či polopublikovaných dokumentů. Projekt bude zaměřen na získávání metadat a případně plných textů z oblasti vzdělávání vědy a výzkumu.

V současné době jsou řešeny otázky typologie dokumentů šedé literatury, metadatové formáty, persistentní identifikátory, softwarová a hardwarová podpora, legislativní otázky atd.

Velký důraz je kladen na vytvoření sítě spolupracujících institucí, které byly osloveny v rámci dotazníkového šetření.

Úzká spolupráce byla navázána s reprezentanty vysokých škol v České republice - Asociace knihoven vysokých škol (eVŠKP) a Knihovnou Akademie věd České republiky.

PROJEKT NUŠL

- Aktuality
- O projektu
- Semináře
- Dotazníkové šetření
- Typologie
- Software
- Metadata
- Identifikátory
- Prezentace
- Spolupráce
- Kontakty

ŠEDÁ LITERATURA

- Definice
- Legislativa v ČR
- Výzkum a vývoj
- Informační zdroje
- Producenti v ČR
- Mezinárodní organizace

Obrázek 4: Náhled úvodní stránky webu projektu

Webové stránky jsou dostupné na adrese <http://nusl.techlib.cz/>. Pro vytvoření webových stránek byla použita aplikace Media Wiki a grafický styl WordPress, modifikovaný podle grafického manuálu NTK. Aplikace Media Wiki podporuje sdílenou tvorbu obsahu v rámci pracovní skupiny a otevírání diskusních skupin k jednotlivým tématům i s účastí veřejnosti. Od roku 2009 mají webové stránky svoji anglickou verzi, která je kopií české verze.

Informovat a podporovat odbornou diskusi z oblasti šedé literatury má za úkol *Seminář ke zpřístupňování šedé literatury*, který je pořádán jedenkrát ročně od roku 2008. Na semináři jsou prezentovány aktuální výsledky projektu NUŠL a reflektována témata související s šedou literaturou, např. stav systémů pro šedou literaturu v tuzemsku i ve světě, formáty a standardy pro dlouhodobé archivování, standardizace pro popis a výměnu zdrojů a autorská práva. Všechny prezentace a plné texty přednášek jsou k dispozici v on-line sborníku na webových stránkách projektu.⁴⁴ Semináře jsou pořádány česko-anglicky s anglickým sborníkem. Důvodem je snaha přiblížit se zahraniční komunitě šedé literatury, která je prozatím velice malá a každý výstup ve všem srozumitelném jazyce je tedy velmi cenný.

Od konce roku 2009 je k dispozici centrální rozhraní pro vyhledávání šedé literatury v České republice. Toto centrální vyhledávací rozhraní nabízí díky vizualizaci dat a kontextové dynamické navigaci komfortní práci s vyhledávanými daty. Do tohoto rozhraní se postupně připojují všechny instituce partnerské sítě NUŠL. Rozhraní je dostupné z webové adresy www.nusl.cz, jeho podrobný popis naleznete v deváté kapitole.

Cílem sběru šedé literatury v České republice je systematicky shromažďovat, dlouhodobě archivovat a především zpřístupňovat tuto literaturu uživatelům z oblasti vědy, výzkumu a vzdělávání. Při sběru šedé literatury se proto klade důraz na důvěryhodnost informací, která je zajišťována spoluprací s ověřenými producenty. z jejich činnosti vznikají druhy šedé literatury v souladu s výše zmíněnou typologií. Předmětem sběru tedy není neověřená šedá literatura typu blogů, telefonních rozhovorů, diskusí, e-mailů apod.

Vzhledem k úzké problematice, kterou šedá literatura představuje, je důležitá spolupráce se zahraničím. Přebírání zkušeností a poskytování řešení v rámci mezinárodní komunity je jedinou možností jak úspěšně budovat přístup k šedé literatuře. NTK proto úzce spolupracuje s mezinárodní organizací pro šedou literaturu GreyNet a správcem databáze OpenSIGLE INIST. Zároveň se NTK snaží veškeré své výstupy z oblasti šedé literatury poskytovat v anglickém jazyce, aby byly srozumitelné mezinárodní komunitě. V roce 2010 hostí NTK GL12.

⁴⁴ <http://nusl.techlib.cz/sbornik/>

LITERATURA:

NOVÁK, Petr. *Šedá literatura v STK: nové aktivity, nové obzory*. Státní technická knihovna. 2008. 22 s.

ELEKTRONICKÉ ZDROJE:

DRIVER [online]. Digital Repository Infrastructure Vision for European Research. [cit. 2010-06-10]. Dostupné z www: <<http://www.driver-community.eu/>>

DSpace [online]. [cit. 2010-06-10] Available from www: <<http://www.dspace.org/>>

GreyNet International [online]. [cit. 2010-06-10]. Dostupné z www: <<http://www.greynet.org/>>

INIST-CNRS [online]. [cit. 2010-06-10]. Dostupné z www: <<http://international.inist.fr/>>

OpenSIGLE [online]. INIST. [cit. 2010-06-10]. Dostupné z www: <<http://opensigle.inist.fr/>>

Portál Národní digitální knihovny [online]. 09. října 2006 [cit. 2010-04-22]. Česká digitální knihovna - diagram - velký. Dostupné z www: http://www.ndk.cz/obrazky/dk_diagram_velky.jpg/view

Seminář ke zpřístupňování šedé literatury [online]. 2008. Praha: Státní technická knihovna, 2008- [cit. 2010-03-15]. Dostupné z www: <http://nusl.techlib.cz/sbornik/>. ISSN 1803-6015.

SVOBODA, Martin. *Šedá literatura v ČR* [online]. 28. 6. 2005, STK, Praha. [cit. 2010-08-31]. Dostupné z www: <http://invenio.ntkcz.cz/record/74>

Textrelease [online]. [cit. 2010-04-22]. Dostupné z www: <http://www.textrelease.com/textreleasehome.html>

Vysokoškolské kvalifikační práce [online]. Fakulta informatiky Masarykovy univerzity. Aktualizováno 10. 6. 2010. [cit. 2010-06-10]. Dostupné z www: <<http://www.theses.cz/>>

World wide science [online]. [cit. 2010-04-22]. World wide science aliance. Aktualizováno 10. 6. 2010. [cit. 2010-06-10]. Dostupné z www: <<http://worldwidescience.org/>>

4. Partnerská síť NUŠL

Iveta Fürstová

Budování partnerské sítě NUŠL

V souvislosti s budováním digitálního repozitáře **NUŠL** vzniká **partnerská síť** institucí, které produkují šedou literaturu. Cílem NUŠL je nabídnout svým uživatelům odborné dokumenty šedé literatury s originálním a hodnotným obsahem. K naplnění tohoto záměru je nezbytné navázat úzkou spoluprací se seriózními subjekty, z jejichž činnosti vzniká kvalitní, důvěryhodná a informačně cenná šedá literatura. Informační hodnota šedé literatury je rozebrána v první kapitole v části Rysy a význam šedé literatury.

Partnerem, který bude ukládat svou produkci šedé literatury do repozitáře NUŠL, se nemůže stát jakýkoliv subjekt. Vhodný partner musí poskytovat šedou literaturu obsahující hodnotné informace, které mohou přispět k dalšímu rozvoji v konkrétním oboru nebo k širší informovanosti veřejnosti. NUŠL v žádném případě neslouží ke sběru neověřené šedé literatury ani k šíření politických myšlenek a diskusí ani jako blog či sociální síť, ale zaměřuje se na odbornou šedou literaturu především z oblasti vědy, výzkumu a vzdělávání.

Důležitým aspektem, který přispívá k nutnosti vytvářet partnerskou síť, je dodržování platných českých právních norem, především práva autorského, práva ochrany osobních údajů a obecného smluvního práva (Polčák, 2009). Z **právní expertizy** pro NUŠL vyplývá, že v repozitáři lze volně zpřístupňovat metadata, ale pro zpřístupňování digitálních dokumentů šedé literatury je nezbytné uzavřít s producenty šedé literatury **licenční smlouvu**,⁴⁵ kde budou jasně a přesně stanoveny možnosti dalšího užití dokumentů. Šedé literatuře z pohledu českého práva se podrobně věnuje kapitola Právní aspekty šedé literatury.

V současnosti má šedá literatura velký potenciál a je důležitým informačním zdrojem. Prostřednictvím šedé literatury se výsledky výzkumů mohou šířit rychleji a efektivněji než v publikovaných zdrojích, kde informace posléze ztrácí na aktuálnosti. Informační možnosti šedé literatury jsou velké, přesto bývá často obtížné tento materiál vyhledat. Jak je již uvedeno v první kapitole, je potřeba využívat i další zdroje kromě on-line databází a při vyhledávání je nutné používat rozšířených vyhledávacích metod.

Při zpracovávání projektu NUŠL byla zjištěna potřeba získat informace o současném stavu a uchovávání šedé literatury v České republice

⁴⁵ Licenční smlouva je smlouva, kterou autor poskytuje nabyvateli oprávnění k výkonu práva dílo užít (Hulva, 2007).

a o potenciálních partnerech z vybraných oblastí. Před zahájením spolupráce byl uskutečněn průzkum formou dotazníkového šetření, v jehož rámci byli osloveni respondenti z AV ČR a z veřejných vysokých škol.⁴⁶ Na webové stránce byl připraven on-line dotazník skládající se z 16 otázek, které pomohly zjistit stav šedé literatury, poptávku po systému shromažďujícím šedou literaturu a zajistit kontakty na odpovědné osoby přímo v jednotlivých institucích majících danou problematiku ve své agendě. Dotazníkové šetření poskytlo první podklady pro zahájení spolupráce s producenty šedé literatury.

Online dotazník se dotazoval jen na několik typických druhů šedé literatury, ale existuje velké množství nejrůznějších materiálů, které je možné do šedé literatury zařadit. Při rozpoznávání toho, co je šedá literatura se vychází z nejrůznějších definic, s nimiž seznamuje již první kapitola Úvod do šedé literatury. Obvykle do oblasti šedé literatury řadíme informační zdroje, které neprocházejí standardním vydavatelským procesem a pocházejí z tzv. šedé zóny (Papík, 2000). Na základě dotazníkového šetření, cílů projektu a zohlednění českého prostředí byla vydefinována typologie NUŠL. Díky této typologii došlo k upřesnění potenciálních partnerů, kterými jsou zejména odborné a výzkumné instituce, školy, knihovny, organizace státní správy, vybrané firmy, ale i odborníci.

Možnosti spolupráce s NUŠL

NUŠL nabízí svým partnerům tři možné způsoby spolupráce. Pokud má spolupracující instituce vlastní databázi, tak se data do NUŠL přeberou z této databáze a importují podle konverzní tabulky. V případě, že instituce svou vlastní databázi nemá, může data vkládat přímo do digitálního repozitáře NUŠL. V tomto případě bude spolupracující instituce používat metadatový formát NUŠL,⁴⁷ který je implementován v šablonách pro vkládání dokumentů. Pro instituci bude vytvořena vlastní sbírka, ke které dostane přidělena příslušná administrátorská práva. Pro administraci sbírky v rámci NUŠL je instituci k dispozici návod v českém jazyce s odkazy na úplnou dokumentaci v anglickém jazyce.⁴⁸ Případně může správu sbírky převzít NTK, což by bylo ošetřeno v příslušné smlouvě mezi NTK a organizací. Třetí možnost může zvolit instituce, která nemá vlastní databázi a chce si nainstalovat systém CDS Invenio, přednastavený v rámci projektu NUŠL a popsáný v deváté kapitole jako lokální repozitář. V tomto případě bude instituce používat metadatový formát NUŠL. Data se pak budou z lokálního repozitáře importovat do digitálního repozitáře NUŠL přes protokol OAI-PMH přímo bez nutnosti konverze dat. Lokální instalace přednastaveného repozitáře a podrobný postup pro jeho instalaci, nastavení

⁴⁶ „V rámci dotazníkového šetření bylo osloveno celkem 77 respondentů, z toho 47 z nich (61%) dotazník vyplnilo. 18 dotazníků bylo získáno z vysokých škol, 29 z výzkumných ústavů AV ČR.“

⁴⁷ http://nusl.techlib.cz/images/Metada_NUSL_final_aktualizace2.pdf

⁴⁸ http://nusl.techlib.cz/index.php/CDS_Invenio_jako_lokalni_repozitar

a spuštění je k dispozici ke stažení na webových stránkách projektu NUŠL. Všechny formy spolupráce jsou a budou poskytovány bezúplatně.

Na podporu spolupráce byly vytvořeny smlouvy, ve kterých je též ošetřena ochrana osobních údajů, protože součástí metadat mohou být osobní údaje. Odstavec ve smlouvě s názvem Ochrana osobních údajů řeší problematiku nakládání s osobními údaji v rámci vzniklého smluvního vztahu. NTK požádala u Úřadu pro ochranu osobních údajů (dále jen ÚOOÚ) o registraci zpracování osobních údajů v databázi NUŠL. Po důkladném šetření bylo zjištěno, že podle ustanovení § 18 odst. 1 písm. b) *Zákona č. 101/2000 Sb., o ochraně osobních údajů a o změně některých zákonů* (dále jen zákon o ochraně osobních údajů) se oznamovací povinnost nevztahuje na zpracování osobních údajů, které správci ukládá zvláštní zákon, nebo je takových údajů třeba k uplatnění práv a povinností vyplývajících ze zvláštního zákona (např. Zákon č. 257/2001 Sb). z tohoto důvodu není nutné plnit oznamovací povinnost a registrovat NUŠL u ÚOOÚ. Příslušný odstavec ve smlouvě zůstal zachován pro ochranu poskytovatele i nabyvatele.

Partneři NUŠL

Akademie věd České republiky

Významným producentem kvalitní šedé literatury v oblasti vědy a výzkumu je AV ČR. „Naším posláním a naší prací je věda“.⁴⁹ Posláním AV ČR a jejích 54 pracovišť je uskutečňovat základní výzkum v širokém spektru přírodních, technických, humanitních a sociálních věd. Informace o tom, co a proč se zkoumá, o výsledcích českých vědců a o rozvoji a financování vědy jsou veřejnou záležitostí, protože AV ČR je financována ze státního rozpočtu. Proto by měla mít jak odborná tak laická veřejnost co nejširší možnosti, jak výsledky vědecké práce vyhledat a využít. Vědečtí pracovníci ústavů AV ČR se snaží velkou většinu svých výstupů publikovat v odborných domácích i zahraničních periodikách, sbornících a monografiích, ale během jejich vědecké práce vzniká ještě další materiál, který může sloužit jako hodnotný zdroj aktuálních informací, tj. šedá literatura. NTK ve spolupráci s Knihovnou AV ČR má zájem shromažďovat a zpřístupňovat šedou literaturu, která vzniká v rámci AV ČR a tím přispět k lepší dostupnosti tohoto informačního zdroje a zvýšení viditelnosti výsledků jednotlivých pracovišť.

Dlouhodobá archivace a možnost snadného vyhledání odborné šedé literatury v sobě nese nemalé výhody, na které upozornila na druhém ročníku *Semináře ke zpřístupňování šedé literatury* Sophia Jones z Univerzity of Nottingham (Jones, 2009). Pro vědeckého pracovníka poskytuje přínos v podobě centrálního archivu jeho prací, zlepšuje dostupnost jeho prací a zvyšuje šíření výsledků

⁴⁹ http://www.cas.cz/o_avcr/poslani/

z jeho výzkumů. Pro instituci je nespornou výhodou zvýšení viditelnosti a prestiže před veřejností a financujícími orgány. Výhodou dlouhodobé archivace je v neposlední řadě možnost přístupu veřejnosti k výsledkům výzkumů. Digitální repozitář NUŠL je právě takovým místem v digitálním prostoru, který trvale uchovává a zpřístupňuje intelektuální produkci vědců i instituce.

Při psaní projektu a tvorbě licenčních smluv se předpokládalo, že ústavy AV ČR budou mít své vlastní institucionální databáze, ale v drtivé většině případů ústavy AV ČR žádnou databázi šedé literatury nevedou a šedá literatura není součástí ani jiné databáze. V praxi budou vkládat jednotlivé druhy dokumentů přes systém **CDS Invenio** nebo přes systém **ASEP**.

Ústavy AV ČR produkují různé množství a druhy šedé literatury. Nejčastějším druhem šedé literatury z produkce pracovišť AV ČR jsou **výroční zprávy**, které obsahují zpravidla informace o vykonaných činnostech a hospodaření instituce za určité období a **výzkumné zprávy**, jež obsahují výsledky výzkumů a uvádějí údaje o řešení výzkumného úkolu. Výroční zprávy příslušný ústav obvykle zveřejňuje na svých webových stránkách.⁵⁰ Výroční a výzkumné zprávy patří mezi typické druhy šedé literatury, které jsou přijímány do NUŠL. Pro uložení a zveřejnění tohoto digitálního dokumentu v NUŠL, i přesto, že je již vystavený na webových stránkách ústavu, musí být uzavřena s ústavem licenční smlouva. Na základě této smlouvy pak může být dokument šedé literatury uložen a zveřejněn v NUŠL.

Z výsledků dotazníkového šetření⁵¹ vyplynulo, že pracoviště AV ČR produkují nejvíce **zprávy o ukončených projektech** a **VŠKP**.⁵² Při osobních jednáních se ukázalo, že pracoviště AV ČR uskutečňují doktorské studijní programy ve spolupráci s vysokými školami, s nimiž mají uzavřenou dohodu o spolupráci. **Disertační práce**, které píší studenti v rámci svého doktorského studia na vysoké škole ve spolupráci s pracovištěm AV ČR, posléze obhájí na příslušné fakultě své školy, kde také záznam a samotný digitální dokument jejich disertační práce zůstane uložen. Právo předat jej do NUŠL má tedy pouze vysoká škola, nikoliv pracoviště AV ČR, přestože s ním doktorand úzce spolupracoval. Text disertační práce je pak prezenčně dostupný v knihovně vysoké školy případně i ústavu. AV ČR rovněž uděluje na základě *Směrnice akademické rady AV ČR č. 3/2006* vědecký titul „doktor věd“ ve zkratce DSc. Podmínkou pro nabytí vědeckého titulu DSc. na **AV ČR** je vytvoření a obhájení disertační práce. Tyto práce jsou společně s tezemi archivovány a prezenčně dostupné v **Knihovně AV ČR** a mohou být AV ČR poskytnuty do NUŠL. Bližší průzkum produkce šedé literatury na jednotlivých pracovištích AV ČR také ukázal, že **zprávy o ukončených projektech** nebo výzkumných záměrech uchovává jen několik ústavů. Většina z nich tyto zprávy posílá příslušné grantové agentuře a dále je nearchivuje, což

⁵⁰ Výroční zprávy jsou na webových stránkách ústavů AV ČR z 90% zveřejněny, ale výzkumné zprávy jen u některých ústavů.

⁵¹ http://nusl.techlib.cz/images/Vyhodnoceni_dotazniku.pdf

⁵² V rámci ústavů AV ČR jsou nejvíce zastoupeny VŠKP a závěrečné zprávy o ukončených projektech, a to u 24 ústavů z 29. 15 ústavů uvedlo, že produkují výzkumné zprávy a 6 technické.

potvrdilo, že dalším neopomenutelným partnerem NUŠL jsou **grantové agentury**.

Mezi druhy dokumentů šedé literatury, které se často objevují v produkci ústavů AV ČR, patří také materiály z domácích i zahraničních konferencí a seminářů. Jedná se o sborníky, **prezentace**, **postery**, programy, přednášky, informační a další doprovodný konferenční materiál. **Sborník** je druh šedé literatury balancující na hraně mezi publikovanou a nepublikovanou literaturou. Je to neperiodická publikace obsahující soubor samostatných, tematicky rámcově příbuzných textů, který může mít i ISBN, ale přesto patří do šedé literatury. Je tzv. **literaturou polopublikovanou**, tedy ne vždy snadno dostupnou jak v tištěné tak v elektronické podobě. **Prezentace**⁵³ a **postery**⁵⁴ jsou po přednesení na konferenci nejčastěji vystaveny, ale ne archivovány na webových stránkách pořádající instituce. Další informační a doprovodné **konferenční materiály** jsou většinou nedostupné a nedohledatelné pro ty, kteří se konference neúčastnili. Zde se naskytuje příležitost zajistit zpřístupnění tohoto typu dokumentů prostřednictvím NUŠL.

Na pracovištích AV ČR vzniká také velké množství autorských prací, což potvrzuje i sdělení na webových stránkách AV ČR, že „v AV ČR pracuje téměř 7 000 zaměstnanců, z nichž více než polovina jsou badatelé s vysokoškolským vzděláním“.⁵⁵ Tyto autorské práce mohou jejich autoři zveřejnit v NUŠL jako **referáty** či preprinty. **Preprint** je z hlediska autora termín pro článek před recenzním řízením a z pohledu vydavatele je preprint obsahově konečná podoba článku s dokončeným recenzním řízením, ale bez konečného formování pro tisk (Rygelová, 2009). Pojetí preprintu vydavateli je v prostředí databáze již chápáno jako **postprint**. Souhrnně a pro zjednodušení se pro pojmy preprint a postprint užívá termín **e-print** (Harnad, 2003). Do NUŠL může autor vložit pouze preprint, tedy článek nebo studii, na kterou ještě neuzavřel s vydavatelem **výhradní licenci**.⁵⁶ Jak zdůrazňuje kapitola Právní aspekty šedé literatury, je nutné rozlišovat, kdy vykonává majetková práva autorská sám autor a kdy je jejich vykonavatelem jiný subjekt.

⁵³ Prezentace je výstup před publikem, kdy řečník pro názornější předávání informací a zvýšení pozornosti posluchačů při své řeči ukazuje přehledné podpůrné materiály, které si předem připravil. Jednotlivým stránkám - listům těchto materiálů říkáme „snímky“ (slides), na nichž mohou být texty, obrázky, tabulky, grafy, diagramy atd.

⁵⁴ Poster je jedním z důležitých prostředků pro prezentaci na konferencích, který by měl shrnout celý vědcův záměr a jeho myšlenky srozumitelně reprodukovat. Je zpravidla 120 cm široký a 180 cm vysoký.

⁵⁵ http://www.cas.cz/o_avcr/poslani/

⁵⁶ Výhradní nebo také exkluzivní licence je podle autorského zákona taková licenční smlouva, podle které autor nesmí poskytnout licenci třetí osobě a je povinen, není-li sjednáno jinak, se i sám zdržet výkonu práva užít dílo způsobem, ke kterému licenci udělil.

Vysoké školy

Z akademického prostředí jsou předními partnery pro NUŠL veřejné, státní a soukromé **vysoké školy** v České republice, které patří k nejvýznamnějším vzdělávacím a vědeckým institucím.⁵⁷ Všeobecné poslání vysokých škol vyjádřila ve svém statusu výstižně Karlova univerzita: „Posláním univerzity je šíření vzdělanosti a ochrana poznání, pěstování svobodného myšlení, nezávislého vědeckého bádání i svébytné umělecké tvorby a všestranná podpora tvůrčího ducha lidské společnosti“.⁵⁸ Vysoké školy patří k významným a plodným producentům šedé literatury jak v oblasti vzdělávání, tak v oblasti vědy a výzkumu.

V oblasti vzdělávání jsou nedůležitějším druhem šedé literatury **VŠKP**, tedy práce **bakalářské**, **diplomové**, **disertační**, **rigorózní** a **habilitační**, které jsou podrobněji popsány v kapitole Zpřístupnění vysokoškolských kvalifikačních prací. Pro vysoké školy platí stejně jako pro AV ČR skutečnost, že veřejnost by měla mít volný přístup k výsledkům obsaženým ve VŠKP, protože vysoké školy jsou financovány z veřejných prostředků, což velkou měrou přispěje k větší transparentnosti a možnosti kontroly kvality výsledků studia. V současnosti, kdy VŠKP vznikají primárně v elektronické podobě, poskytuje možnost zpřístupnění v elektronickém prostředí snadný a rychlý přístup k informacím.

Právní kvalifikace pojmu VŠKP je ve srovnání s ostatními druhy šedé literatury odlišná. Na VŠKP se vztahuje *Zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů* (dále jen autorský zákon), který toto dílo označuje za tzv. **školní dílo**. Vedle toho *Zákon č. 111/1998 Sb., o vysokých školách a o změně a doplnění některých dalších zákonů* (dále jen zákon o vysokých školách) upravuje způsob zveřejnění VŠKP prostřednictvím **databáze**. Všechny tyto legislativní zásahy představují pro vysoké školy i studenty změny, které je nutné v souladu se zákonem uvést do praxe. Z právního hlediska jsou VŠKP rozebrány v kapitole Právní aspekty šedé literatury v oddílu Školní díla jako šedá literatura. Právní analýzu vztahů plynoucích z elektronického zpracování a publikace VŠKP včetně metadat zahrnutých v databázích vysokých škol poskytuje **právní expertiza** vytvořená pro NUŠL. Tato expertiza má za úkol sloužit jak vysokým školám tak studentům, jako soubor právních doporučení, jak řešit zveřejnění VŠKP (Polčák, 2009).

V současné době najdeme v registru archivů e-VŠKP na webových stránkách **Komise eVŠKP** 14 fungujících repozitářů vysokých škol (Registr archivů, c2004-2007). Nejrozšířenějším systémem je *Informační systém studijní agendy* (dále jen **STAG**) (IS STAG, c2007), který vyvinula *Západočeská univerzita v Plzni* (dále

⁵⁷ V České republice je 26 státních a 56 soukromých vysokých škol.

⁵⁸ <http://certik.ruk.cuni.cz/asuk/statutarni/predpisy/registrovane/statut.html>

jen ZČU) a od roku 2000 jej používá 12 vysokých škol.⁵⁹ Mezi nimi je *Univerzita Tomáše Bati ve Zlíně* (dále jen UTB), která uchovává již pouze eVŠKP a zpřístupňuje základní informace o VŠKP a pokud dal autor souhlas, jsou k dispozici i plné texty (Budínský, 2007, 2009).

Z dostupných typů softwarových systémů je pro tvorbu digitálních knihoven často používán software DSpace, v němž budují své digitální knihovny VŠKP *Vysoká škola báňská - Technická univerzita Ostrava*⁶⁰ (dále jen VŠB-TUO), *Univerzita Jana Evangelisty Purkyně v Ústí nad Labem*⁶¹ (dále jen UJEP), *Akademie múzických umění v Praze*⁶² (dále jen AMU), *Univerzita Pardubice*⁶³ (dále jen UPa) a v testovacím provozu *Masarykova univerzita v Brně* (dále jen MU).⁶⁴ MU zpřístupnila již 1. 9. 2006 VŠKP v prostředí Internetu v úplném rozsahu prostřednictvím *Archivu závěrečných prací MU v Informačním systému MU*.⁶⁵

Univerzita Karlova v Praze (dále jen *Univerzita Karlova*) zpřístupnila v prosinci 2009 *Digitální univerzitní repozitář Univerzity Karlovy v Praze*,⁶⁶ který je provozován v systému DigiTool společnosti ExLibris a umožňuje dlouhodobou archivaci, správu a zpřístupňování elektronických dokumentů. Systém DigiTool implementuje také *Vysoké učení technické v Brně* (dále jen VUT) pro budování *Digitální knihovny VUT v Brně*, jehož využití bude jak pro ukládání a zveřejňování VŠKP, tak jako institucionální repozitář (Šimová, 2009).

Mendelova zemědělská a lesnická univerzita v Brně (dále jen MENDELU) realizuje archiv VŠKP v rámci jednotného *Univerzitního informačního systému*, kde je zpřístupněn veřejný registr VŠKP. Tento archiv používá metadatový soubor požadovaný Komisí eVŠKP. *Janáčkova akademie múzických umění v Brně* (dále jen JAMU) vede celouniverzitní repozitář, který je přístupný z webových stránek knihovny a je určen k ukládání a zpřístupňování VŠKP. Přístup k bibliografickým údajům a anotacím je bez omezení a plné texty jsou přístupné pouze pro studenty a pracovníky JAMU.

Ostravská univerzita v Ostravě (dále jen OU) provozuje od roku 2003 *Databázi vysokoškolských kvalifikačních prací OU*, kde vyhledávání uložených prací je dostupné na *Portálu OU*. Zde studenti také ukládají informace o svých pracích.

⁵⁹ Informační systém STAG používá např. Technická univerzita v Liberci, Univerzita Jana Evangelisty Purkyně, Akademie výtvarných umění Praha, Univerzita Pardubice, Univerzita Palackého v Olomouci, Slezská univerzita v Opavě, Ostravská univerzita v Ostravě, Veterinární a farmaceutická univerzita Brno, Jihočeská univerzita v Českých Budějovicích atd.

⁶⁰ <http://dspace.vsb.cz/>

⁶¹ <http://edu.ujep.cz/main.php?idwebu=dspace>

⁶² <http://dspace.amu.cz/>

⁶³ <http://dspace.upce.cz/>. Zde STAG zůstal prostředím pro kompletní ukládání metadat.

⁶⁴ <http://dspace.muni.cz/>

⁶⁵ Ekonomicko-správní fakulta MU vede Elektronický archiv akademických závěrečných prací <http://zp.svi.econ.muni.cz/>. V archivu nabízí bakalářské a diplomové práce z let 1994-2005, celkově je zde přes 3 500 dokumentů, z toho přes 1 200 z nich je volně přístupných i veřejnosti. Práce od roku 2006 jsou celouniverzitně shromažďovány v rámci IS MU.

⁶⁶ <http://repozitar.cuni.cz/>

VŠKP jsou k dispozici ve studovnách a půjčují se pouze prezenčně podle vyhledané signatury v databázi VŠKP na Portále OU nebo přímo na webových stránkách.⁶⁷ *Technická univerzita v Liberci* (dále jen **TUL**) má v rámci celouniverzitního systému v odkazu Absolventské práce umístěna metadata k **VŠKP**.

Pro takto vytvořené a funkční univerzitní databáze, kde si školy sami archivují VŠKP, slouží projekt NUŠL jako **centrální vyhledávací prostředí**, kde uživatel na jednom místě nalezne metadata a odkazy na plné texty ze všech spolupracujících vysokých škol v České republice. Pomůže tak zajistit lepší dostupnost a transparentnost VŠKP. Další aspekty a informace o repozitářích VŠKP v České republice i v zahraničí přináší následující kapitola Zpřístupnění vysokoškolských kvalifikačních prací.

Z pohledu vědy a výzkumu jsou vysoké školy také producentem dalších druhů šedé literatury, obdobně jako ústavy AV ČR. Zaměstnanci vysokých škol zpracovávají **studijní materiály** (např. studijní opory a skripta, záznamy a prezentace z přednášek, okruhy ke zkouškám, sbírky příkladů, učební osnovy aj.), **zprávy** (např. výroční, z výzkumů, z projektů atd.) a **konferenční materiály**. Ve větší míře vznikají také **autorské práce** jako jsou **preprinty** či **referáty**. Výše uvedené varianty spolupráce s NUŠL pro zajištění ukládání, archivace a zpřístupnění těchto druhů literatury jsou shodné jako u ústavů AV ČR.

Grantové agentury

Součástí **partnerské sítě** NUŠL by se měly stát již výše zmíněné **grantové agentury**. Nejvýznamnější grantové agentury v České republice jsou *Grantová agentura ČR* (dále jen **GA ČR**) a *Grantová agentura AV* (dále jen **GA AV**), které zabezpečují přípravu a realizaci grantových projektů a poskytují účelovou podporu v souladu s Národní politikou výzkumu, experimentálního vývoje a inovací České republiky. Většina vysokých škol má také své grantové agentury a kromě toho funguje v České republice řada resortních grantových agentur (př. FRVŠ Ministerstva školství, mládeže a tělovýchovy, zdravotnictví - IGA MZ, zemědělství, životního prostředí, vnitra a dalších), grantové agentury krajské a místní samosprávy (kultura, sociální péče) a zejména zahraničních nadací. Digitální repozitář NUŠL je vhodným řešením pro zajištění dlouhodobé archivace a zpřístupnění **závěrečných zpráv z projektů**, které grantové agentury financují.

⁶⁷ <https://portal.osu.cz/wps/portal/dipl>

Státní správa a podniky

Důležitým článkem partnerské sítě NUŠL jsou instituce státní správy - v první fázi především **ministerstva**. V současné době oslovujeme MK ČR, které kromě projektů VaV zajišťuje i projekty *Veřejné informační služby knihoven* (dále jen VISK) a Ministerstvo školství mládeže a tělovýchovy, s nimiž bychom rádi dojednali ukládání závěrečných zpráv z projektů, ale i další šedé literatury do NUŠL. Předpokládáme rovněž spolupráci s českými **podniky** (např. sklárny, pivovary, průmyslové podniky atd.), které jsou z pohledu NUŠL zajímavé především vydáváním výročních zpráv, katalogů, analýz, statistik, průzkumů atd.

Osobní archivy

Zajímavou součástí NUŠL by měla být spolupráce s osobnostmi z nejrůznějších oblastí vědy a výzkumu, které by měli zájem v NUŠL dlouhodobě archivovat a zpřístupnit svůj **osobní archiv**. S NUŠL v současné době aktivně spolupracuje Ing. arch. Jan Moučka, který během své praxe v oboru architektury shromáždil řadu informací a dokumentů, které by rád uložil do repozitáře NUŠL a tím umožnil přístup k dokumentům především studentům a vědeckým pracovníkům z oboru.

Posláním projektu NUŠL, který vzniká a bude dále fungovat pod záštitou MŠMT, je vytvořit centrální prostředí pro šedou literaturu, které dokáže sjednotit roztržité institucionální databáze a shromáždí jednotlivé dokumenty šedé literatury z nejrůznějších oblastí vědy, výzkumu a vzdělávání na národní úrovni. Dostupnost šedé literatury v České republice je v současné době na velmi nízké úrovni. Neutěšená situace je v oblasti zpráv, konferenčních materiálů, analýz, statistik, autorských prací atd., kde vědecké a správní instituce nevedou žádnou interní databázi. Lepší **dostupnost**, ale zato nejednotnost, panuje v případě zpřístupňování VŠKP, kde stále chybí možnost prohledávat napříč databázemi všech vysokých škol. Digitální repozitář NUŠL pomůže ke sjednocení šedé literatury a ke snadnějšímu a rychlejšímu vyhledání informací.

LITERATURA:

BUDÍNSKÝ, Lukáš. Zpřístupňování e-VŠKP na Univerzitě Tomáše Bati ve Zlíně. *Konference Digitální knižnice* [online]. 1. - 3. 10. 2007, Jasná pod Chopkem, 2007. [cit. 2010-03-15]. Dostupné z [www: <http://invenio.ntkcz.cz/record/22/files/CONFERENCE-PRESENTATION-2009-022.pdf>](http://invenio.ntkcz.cz/record/22/files/CONFERENCE-PRESENTATION-2009-022.pdf)

BUDÍNSKÝ, Lukáš. DSpace na UTB - na půl cesty k cíli. 3. setkání českých uživatelů DSpace [online]. 18. - 19. 5. 2010, Ostrava. 2010. [cit. 2010-06-02]. Dostupné z [www: <http://dspace.vsb.cz/handle/10084/78295>](http://dspace.vsb.cz/handle/10084/78295)

HARNAD, Stevan. Eprints. Electronic preprints and postprints. In DRAKE, Miriam A. (ed.). *Encyclopedia of library and information science* [online]. New York: Marcel Dekker, c2003. [cit. 2010-06-02]. Dostupné z [www: <http://www.ecs.soton.ac.uk/~harnad/Temp/eprints.htm>](http://www.ecs.soton.ac.uk/~harnad/Temp/eprints.htm)

HULVA, Tomáš. *Právo duševního vlastnictví*. 1. vydání. Opava: Slezská univerzita v Opavě, 2007, s. 72.

JONES, Sophia. Open Access and digital repositories: the role of the DRIVER project [online]. 22. 10. 2009 Praha: Národní technická knihovna, 2009. [cit. 2010-03-12]. Dostupné z [www: <http://nysl.techlib.cz/sbornik/2009/](http://nysl.techlib.cz/sbornik/2009/) ISSN 1803-6015.

PAPÍK, Richard. Zdroje šedé literatury a jejich strategický potenciál pro vědu a obchod [online]. In. *INFORUM 2000. Praha 23. - 25. 5. 2000*. 2000 [cit. 2010-03-08]. Dostupné z [www: <http://www.inforum.cz/archiv/inforum2000/prednasky/zdrojesedelit.htm>](http://www.inforum.cz/archiv/inforum2000/prednasky/zdrojesedelit.htm)

POLČÁK, Radim. Digitální zpracování tzv. šedé literatury pro Národní úložiště šedé literatury [online]. 2009. [cit. 2010-03-10]. Dostupné z [www: <http://nysl.techlib.cz/index.php/Právní_expertiza>](http://nysl.techlib.cz/index.php/Právní_expertiza).

RYGELOVÁ, Pavla. Služba SHERPA/RoMEO - další život vědeckého článku. *Zpravodaj ÚVT MU* [online]. 2009. roč. 20, č. 2, s. 12 - 17 [cit. 2010-04-29]. Dostupné z [www: <http://www.ics.muni.cz/bulletin/clanky_tisk/630.pdf>](http://www.ics.muni.cz/bulletin/clanky_tisk/630.pdf) ISSN 1212-0901.

ŠÍMOVÁ, Barbora. Budování Digitální knihovny VUT v Brně. *Systémy pro zpřístupňování VŠKP: zkušenosti, možnosti, nabídky, potřeby: 4. ročník semináře konaného 21. 10. 2009 na VUT v Brně* [online]. Praha: Odborná komise pro otázky elektronického zpřístupňování VŠKP AKVŠ ČR. [cit. 2010-03-15]. Dostupné z [www: <http://www.evskp.cz/seminar/2009/>](http://www.evskp.cz/seminar/2009/). ISSN 1803-7003.

Zákon č. 101/2000 Sb., zákon o ochraně osobních údajů a o změně některých zákonů (o ochraně osobních údajů)

Zákon č. 111/1998 Sb., zákon o vysokých školách a o změně a doplnění některých dalších zákonů (o vysokých školách).

Zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), jak vyplývá z pozdějších změn.

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

ELEKTRONICKÉ ZDROJE:

I. úplné znění statutu Univerzity Karlovy v Praze ze dne 26. dubna 2004 [online], 2004. Univerzita Karlova v Praze. [cit. 2010-03-11]. Dostupné z [www: <http://certik.ruk.cuni.cz/asuk/statutarni/predpisy/registrovane/statut.html>](http://certik.ruk.cuni.cz/asuk/statutarni/predpisy/registrovane/statut.html)

Absolventské práce [online]. Univerzitní knihovna. Technická univerzita v Liberci. c2005 - 2010. [cit. 2010-06-10]. Dostupné z [www: <http://knihovna.tul.cz/cat/Absolventske-prace-45.php>](http://knihovna.tul.cz/cat/Absolventske-prace-45.php)

Registr archivů e-VŠKP [online]. Stránky Odborné komise pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací při Asociaci knihoven vysokých škol České republiky. c2004 - 2007. Aktualizováno 12. 10. 2007. [cit. 2010-03-15] Dostupné z [www: <http://www.evskp.cz/registry.php?sek=2&tsekce=1&id=10>](http://www.evskp.cz/registry.php?sek=2&tsekce=1&id=10)

Databáze VŠKP (vysokoškolských kvalifikačních prací). Portál Ostravské univerzity [online]. Centrum informačních technologií, c2004 - 2007. [cit. 2010-06-02]. Dostupné z [www: <https://portal.osu.cz/wps/portal/dipl>](https://portal.osu.cz/wps/portal/dipl)

Digitální knihovna Univerzity Pardubice [online]. Univerzitní knihovna UPa. Univerzita Pardubice, c2002 - 2009. [cit. 2010-06-02]. Dostupné z [www: <http://dspace.upce.cz/>](http://dspace.upce.cz/)

Digitální repozitář Masarykovy univerzity [online]. Masarykova univerzita. [cit. 2010-06-02]. Dostupné z [www: <http://dspace.muni.cz/>](http://dspace.muni.cz/)

Digitální repozitář Vysoké školy báňské - Technické univerzity Ostrava [online]. Vysoká škola báňská - technická univerzita Ostrava, c2002 - 2007. [cit. 2010-06-02]. Dostupné z [www: <http://dspace.vsb.cz/>](http://dspace.vsb.cz/)

Digitální univerzitní repozitář Univerzity Karlovy v Praze [online]. Univerzita Karlova v Praze. Ústav výpočetní techniky, c2007, [cit. 2010-06-02]. Dostupné z [www: <http://repozitar.cuni.cz/>](http://repozitar.cuni.cz/)

Doktorské studijní programy uskutečňované ve spolupráci s vysokými školami na pracovištích AV ČR [online]. Akademie věd České republiky. Středisko společných činností AV ČR, v.v.i. Aktualizováno 2. 2. 2010. [cit. 2010-03-11] Available from [www: <http://www.cas.cz/vzdelavani/doktorske_studijni_programy/prehled_akreditovanych_doktorskych_studijnich_programu.html>](http://www.cas.cz/vzdelavani/doktorske_studijni_programy/prehled_akreditovanych_doktorskych_studijnich_programu.html)

DSpace - AMU repository [online]. Akademie múzických umění v Praze, c2008. [cit. 2010-06-02] Available from [www: <http://dspace.amu.cz/>](http://dspace.amu.cz/)

DSpace - úložiště kval. prací UJEP [online]. CI Univerzity Jana Evangelisty Purkyně v Ústí nad Labem, c2008. [cit. 2010-06-02]. Dostupné z [www: <http://edu.ujep.cz/main.php?idwebu=dspace>](http://edu.ujep.cz/main.php?idwebu=dspace)

Fond rozvoje vysokých škol [online]. Aktualizováno 16. 4. 2010. [cit. 2010-05-30]. Dostupné z [www: <http://www.frvs.cz/>](http://www.frvs.cz/)

IS STAG. Informační systém studijní agendy [online]. CIV-SIS ZČU, c2009. Aktualizováno 20. 5. 2010. [cit. 2010-03-15]. Dostupné z [www: <http://is-stag.zcu.cz/zajemci/>](http://is-stag.zcu.cz/zajemci/)

Knihovna [online]. Janáčkova akademie múzických umění v Brně. [cit. 2010-06-02]. Dostupné z [www: <http://www.jamu.cz/organizacni-struktura/knihovna/>](http://www.jamu.cz/organizacni-struktura/knihovna/)

Národní úložiště šedé literatury [online]. Praha: Národní technická knihovna, 2008. Aktualizováno 23. 3. 2009, [cit. 2010-08-05]. Dostupné z [www: <http://nysl.techlib.cz/>](http://nysl.techlib.cz/)

Odborná komise pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací [online]. OK e-VŠKP AKVŠ ČR, c2004 - 2007. Aktualizováno 12. 10. 2007. [cit. 2010-03-12]. Dostupné z [www: <http://www.evskp.cz/>](http://www.evskp.cz/)

Poslání AV ČR. Akademie věd České republiky [online]. Středisko společných činností AV ČR, v.v.i., 2009. Aktualizováno 11. 11. 2009. [cit. 2010-06-03]. Dostupné z [www: <http://www.cas.cz/o_avcr/poslani/>](http://www.cas.cz/o_avcr/poslani/)

Směrnice akademické rady AV ČR č. 3/2006 - Pravidla pro udělování vědeckého titulu „doktor věd“ v AV ČR [online]. Akademie věd České republiky. 2006. [cit. 2010-03-15]. Dostupné z [www: <http://www.cas.cz/miranda2/export/sitesavcr/data.avcr.cz/vzdelavani/vedecky_titul_doktor_ved/files/IN06-08.pdf>](http://www.cas.cz/miranda2/export/sitesavcr/data.avcr.cz/vzdelavani/vedecky_titul_doktor_ved/files/IN06-08.pdf)

Výzkum možností spolupráce při šíření výsledků výzkumu a vývoje. Analýza výsledků dotazníkového šetření. NTK [online]. 2008. [cit. 2010-03-10]. Dostupné z [www: <http://nysl.techlib.cz/images/Vyhodnoceni_dotazniku.pdf>](http://nysl.techlib.cz/images/Vyhodnoceni_dotazniku.pdf)

Základní informace o asociaci [online]. Asociace knihoven vysokých škol České republiky. c2003 - 2010. Aktualizováno 18. 4. 2010. [cit. 2010-03-12]. Dostupné z [www: <http://www.akvs.cz/o-asociaci/index.html>](http://www.akvs.cz/o-asociaci/index.html)

Závěrečné práce na MENDELU [online]. UIS MENDELU. Brno: Mendlova univerzita v Brně, [2010]. [cit. 2010-06-02]. Dostupné z [www: <http://is.mendelu.cz/zp/>](http://is.mendelu.cz/zp/)

5. Zpřístupnění vysokoškolských kvalifikačních prací

Jan Mach

Vysoké školy jsou nejvyšším článkem ve vzdělávací soustavě České republiky, jsou vrcholnými centry vzdělanosti, nezávislého poznání a tvůrčí činnosti a mají klíčovou úlohu ve vědeckém, kulturním, sociálním a ekonomickém rozvoji. Významnou součástí šedé literatury tvoří VŠKP, jejichž zpracováním student při obhajobě prokazuje nabyté zkušenosti během studia a které jsou definovány v zákonu o vysokých školách. Vysoké školy jsou významnými partnery pro NUŠL, protože VŠKP obsahují ověřené informace, které jsou zpracovávány pod dohledem zkušených pedagogů a procházejí obhajobou. Téma jejich zpřístupňování v rámci zákona o vysokých školách je v současnosti aktuální a intenzivně diskutované. Vzhledem ke své odlišnosti v rámci ostatních druhů šedé literatury je problematice VŠKP věnována samostatná kapitola.

V České republice je na vysokých školách možnost studovat v bakalářském, magisterském a doktorském studijním programu, který je každý zakončen specifickou závěrečnou prací a její obhajobou. V případě bakalářského studijního programu se studium ukončuje státní závěrečnou zkouškou, jejíž součástí je zpravidla obhajoba bakalářské práce. V případě magisterského studijního programu se studium ukončuje státní závěrečnou zkouškou, jejíž součástí je obhajoba diplomové práce. V oblasti lékařství a veterinárního lékařství a hygieny se studium řádně ukončuje státní rigorózní zkouškou. Absolventi magisterských studijních programů, kteří získali akademický titul magistr, mohou vykonat v téže oblasti studia státní rigorózní zkoušku, jejíž součástí je obhajoba rigorózní práce. Doktorský studijní program se zakončuje doktorskou zkouškou a obhajobou disertační práce.

Zatímco pro práce bakalářské, diplomové a rigorózní není v zákoně o vysokých školách přesněji stanoveno, jakou mají mít formu a obsah, v případě doktorské zkoušky a disertační práce musí student prokázat „schopnost a připravenost k samostatné činnosti v oblasti výzkumu nebo vývoje nebo k samostatné teoretické a tvůrčí umělecké činnosti. Disertační práce musí obsahovat původní a uveřejněné výsledky nebo výsledky přijaté k uveřejnění“ (Zákon č. 111/1998 Sb.).

Mezi VŠKP řadíme i habilitační práce, které předkládají žadatelé při habilitačním řízení za účelem získání titulu docent. V tomto případě se tedy nejedná o studenty předkládající práci v rámci studijního programu školy. Zákon o vysokých školách v § 72 definuje, že:

„(1) V habilitačním řízení se ověřuje vědecká nebo umělecká kvalifikace uchazeče, a to zejména na základě habilitační práce a její obhajoby a dalších vědeckých, odborných nebo uměleckých prací, a jeho pedagogická způsobilost na základě hodnocení habilitační přednášky a předcházející pedagogické praxe.

(3) Habilitační práci se rozumí:

- a) písemná práce, která přináší nové vědecké poznatky, nebo
- b) soubor uveřejněných vědeckých prací nebo inženýrských prací doplněný komentářem, nebo
- c) tiskem vydaná monografie, která přináší nové vědecké poznatky, nebo
- d) umělecké dílo nebo umělecký výkon nebo jejich soubor, kterým je například vynikající veřejná umělecká činnost“ (Zákon č. 111/1998 Sb.).

VŠKP, především práce disertační a habilitační, obsahují výsledky vědecké, výzkumné, vývojové nebo umělecké činnosti a procházejí obhajobou, která představuje recenzní řízení. Tvoří tak významnou část šedé literatury a existuje zde oprávněný zájem na jejich zpřístupnění veřejnosti. Přestože jsou do jisté míry publikovány, nejsou většinou jednoduše dostupné. Z tohoto důvodu je potřeba zajistit jejich zpřístupnění. Vzhledem k možnostem moderních informačních technologií se nabízí především jejich zpřístupnění elektronickou formou.

Do roku 2006 bylo běžnou praxí půjčovat práce pouze v tištěné podobě prezenčně v knihovně pouze studentům a zaměstnancům dané vysoké školy na základě školní licence dle autorského zákona. V případě, že škola podepisovala se studenty licenční smlouvu, bylo možné půjčovat tyto práce i širší veřejnosti.

Zákon 216/2006 Sb., kterým se mění autorský zákon, mimo jiné vymezil nové situace, kdy užitím díla nedochází k zásahu do autorského práva. VŠKP se týká § 37 Knihovní licence, podle kterého:

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

(d) půjčuje-li originály nebo rozmnoženiny obhájených diplomových, rigorózních, disertačních a habilitačních prací na místě samém, a to výhradně pro účely výzkumu nebo soukromého studia, pokud takové užití autor nevyloučil“ (Zákon č. 216/2006 Sb.).

Toto užití však ještě neumožňovalo zpřístupnit práce volně prostřednictvím Internetu.

Velmi významnou legislativní změnou byl Zákon č. 552/2005 Sb. vkládající do zákona o vysokých školách § 47 b). Ten školám dává povinnost nevýdělečně zveřejňovat disertační, diplomové, bakalářské a rigorózní práce, u kterých proběhla obhajoba, včetně posudků oponentů a výsledku obhajoby. Tyto práce musí být též nejméně pět dní před obhajobou dostupné veřejnosti, je možné si z nich pořizovat výpisy, opisy nebo rozmnoženiny. Způsob, jakým jsou práce zveřejněny v databázi kvalifikačních prací, má stanovit vnitřní předpis vysoké školy. Autor odevzdáním své práce souhlasí s takovýmto zveřejněním bez ohledu na výsledek obhajoby. Zákon o vysokých školách neupravuje zveřejňování prací habilitačních.

Novela zákona o vysokých školách však situaci s volným zpřístupněním prací prostřednictvím internetu plně nevyjasnila. Nejvíce problematickým je pojem

použitý v novele „zveřejnění“ závěrečných prací, který není jasně stanoven ani v zákoně o vysokých školách, ani v autorském zákoně. Autorský zákon používá odlišný termín „sdělování díla veřejnosti“. Z tohoto důvodu byla praxe na jednotlivých školách odlišná.

Některé školy práce vybíraly i nadále pouze v tištěné podobě a zpřístupňovaly je v rámci knihoven jednotlivých fakult. Některé se rozhodly VŠKP zpřístupňovat v rámci školy nebo knihovny s odkazem na licenci danou autorským zákonem. U těchto škol byl zájemce většinou nucen stát se registrovaným čtenářem příslušné **vysokoškolské knihovny**, pokud chtěl k pracím získat přístup. Třetí kategorie škol práce zpřístupnila volně prostřednictvím internetu, a to buď s odkazem na souhlas autora daný dle zákona o vysokých školách odevzdáním práce, nebo pro tyto účely i nadále získávaly licence od studentů.

Po několika letech účinnosti novely zákona o vysokých školách tato rozdílná praxe vysokých škol ve zpřístupňování VŠKP i nadále trvá. Přesto přibývá škol přiklánějících se k povinnosti dané zákonem o vysokých školách a práce zpřístupňují veřejnosti prostřednictvím Internetu, které je ovšem často podmíněno registrací uživatele.

Podobně jako zpřístupnění VŠKP na jednotlivých vysokých školách, tak i jejich archivování se velmi liší. Každá vysoká škola si pomocí vnitřní směrnice stanovila skartační řád pro tištěné VŠKP. Různorodost těchto skartačních řádů dokládá příloha č. 3, která rozebírá praxi na pěti vybraných vysokých školách.

Do způsobu archivace se promítl i přechod od sběru tištěných VŠKP ke sběru elektronických verzí. Díky legislativním změnám nyní školy nemusí uchovávat práce v tištěné podobě v kapacitně omezených knihovnách a archivech, ale mohou pracovat a dlouhodobě uchovávat pouze dokument digitální. Např. VŠE považuje za primární dokument elektronickou verzi a vybírá a archivuje práce v elektronické podobě. Listinné verze (pokud jsou vůbec požadovány k obhajobě) se vrací autorovi práce nebo je pracoviště, na němž se práce obhájí, skartuje ve vlastní režii a v souladu se **skartačním řádem**.

Při budování repozitářů šedé literatury je nutné zajistit požadavek **dlouhodobého uložení** VŠKP, jenž představuje archivaci elektronické formy. Dlouhodobé uložení bez odstraňování z repozitáře, které u tištěných materiálů představovala skartace, je opodstatněné též z důvodu nedávných kauz, ve kterých bylo často i po mnoha letech po dostudování prokázáno **plagiátorství** v obhájených VŠKP.

Organizace zabývající se zveřejňováním VŠKP

Koncept eVŠKP byl poprvé diskutován v roce 1987 na setkání v Michiganu a dále rozpracován na začátku 90. let minulého století na *Virginia Polytechnic Institute and State University* (dále jen Virginia Tech) pod vedením profesora Edwarda A. Foxe, který se stal spolupředsedou pracovní skupiny pro diplomové práce, technické zprávy a disertační práce. Účastníci workshopu *Southeastern*

*Universities Research Association*⁶⁸ (dále jen SURA) v roce 1996 zvolili Adobe Portable Document Format (dále jen PDF) a Standard Generalized Markup Language (dále jen SGML) pro reprezentaci a archivaci eVŠKP.

Na základě proběhlých odborných diskusí a workshopů vznikl ve stejném roce na Virginia Tech software ETD database⁶⁹ (dále jen ETD db), zajišťující kompletní řešení pro odevzdávání, zpracování, archivaci a zpřístupnění eVŠKP (NDLTD, c2010). Program byl uvolněn pro bezplatné užívání na mezinárodní úrovni. Virginia Tech koordinovala vývoj a implementaci systému distribuované digitální knihovny, která shromažďovala data od jednotlivých spolupracujících institucí. Systém umožňoval prohlížení a vyhledávání podle instituce, data vytvoření, autora, názvu, klíčových slov a plného textu práce. Plné texty kvalifikačních prací byly zpřístupněny celosvětově s možností stažení, uložení a tisku.

V roce 1996 byla v USA založena skupina *The National Digital Library of Theses and Dissertations*. Po rozšíření své působnosti na mezinárodní úroveň byla přejmenována na současný název NDLTD. V roce 2003 se z ní stala nezisková charitativní organizace, která v současnosti sdružuje již stovky vysokých škol z celého světa a partnerských organizací.

Nejvýznamnějším dokumentem NDLTD je standard „ETD-MS: an Interoperability Metadata Standard for Electronic Theses and Dissertations“ (ETD-MS, 2008), který popisuje VŠKP pomocí 13 prvků Dublin Core a jednoho vlastního prvku thesis.degree, obsahujícího podprvky name, level, discipline a grantor. Aktuální verze 1.00 revize 2 z 21. 3. 2006 obsahuje, kromě ukázek zápisu metadat ve Vanilla kódování,⁷⁰ i specifikaci převodu do MARC-21. Standard ETD-MS sloužil jako východisko pro návrh dalších zahraničních standardů včetně českého EVSKP-MS.

Dodržování standardu ETD-MS a zpřístupnění metadat protokolem OAI-PMH umožňuje efektivnější budování registrů eVŠKP, např. vyhledávací systém SCIRIUS ETD Search společnosti Elsevier nebo souborná metadatová databáze XTCat NDLTD Union Catalog. Nejnovější systém VTLIS Visualiser⁷¹ nabízí uživatelsky velmi přívětivé rozhraní s možností dodatečného zpřesňování nalezených výsledků.

V současné době se zahraničí problematikou eVŠKP zabývá řada institucí a projektů na národní či mezinárodní úrovni. Pro Českou republiku je významný především nadnárodní projekt *DART-Europe E-theses Portal* (dále jen DART-Europe) zpřístupňující na svém portálu elektronické disertace z vybraných evropských knihoven (DART-Europe, c1999-2010). Metadata jsou do portálu automatizovaně sklízena z lokálních repozitářů zapojených vysokých škol. Podmínkou pro zapojení je zpřístupnění metadat disertačních prací protokolem

⁶⁸ <http://www.sura.org/home/index.html>

⁶⁹ <http://scholar.lib.vt.edu/ETD-db/index.shtml>

⁷⁰ <http://www.evskp.cz/Seminar3/seminar3-BratkovaMach.pdf>

⁷¹ <http://www.vtls.com/products/visualizer>

OAI-PMH ve formátu Dublin Core a vystavení plných textů volně prostřednictvím Internetu.

V České republice prokázaly zájem knihovníků o zveřejňování eVŠKP prezentace a následná diskuse na Celostátní poradě vysokoškolských knihoven v Hradci Králové 19. - 20. 11. 2003 (Prezentace, 2003). Z iniciativy Klubu vysokoškolských knihovníků při Svazu knihovníků a informačních pracovníků (dále jen SKIP) byla následně v roce 2004 založena Komise eVŠKP. V rámci Komise eVŠKP byli zastoupeni knihovníci z většiny veřejných vysokých škol a z NTK, která řeší zpřístupnění šedé literatury na národní úrovni.

Komise eVŠKP měla za úkol navázat na zkušenosti ze zahraničí a usilovat o vytvoření koncepce systému na národní úrovni, sjednotit různorodou praxi na jednotlivých školách včetně předpisů, organizačně-administrativních postupů, formu eVŠKP a nakládání s nimi. Komise eVŠKP si byla vědoma nutnosti zvyšovat počítačovou, publikační, právní a informační gramotnost samotných autorů VŠKP a nutnosti řešit související autorsko-právní problematiku. Vzhledem k různosti metodických a technologických procesů na školách se ukázala jako nezbytná příprava a implementace standardů včetně aplikace mezinárodně doporučovaných standardů pro evidenci a zpřístupňování VŠKP (K otázkám, 2004).

Během své činnosti komise připravila *Soubor doporučení pro vysokoškolské knihovny*, zahrnující doporučený postup kroků při zavádění registrů eVŠKP na školách, workflow sběru a zpřístupňování prací, doporučení pro vysokoškolské předpisy, metodické pokyny a standardy pro popis souvisejících metadat. Díky tomu byly vysoké školy připraveny po roce 2006 pružně reagovat na změny dané novelami zákona o vysokých školách a autorského zákona.

Jedním ze zásadních dokumentů vzniklých na půdě Komise eVŠKP byl *Metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR* (dále jen EVSKP-MS), který vznikl jako návrh ve verzi 0.1 v roce 2005 a definoval národní soubor metadatových prvků pro popis eVŠKP obhájovaných na vysokých školách v České republice a pro přenos souborů. Tvůrci českého standardu metadat EVSKP-MS se při práci inspirovali kromě ETD-MS především projektem *Digitale Dissertationen im Internet*⁷² (dále jen DissOnline) koordinovaným německou národní knihovnou. V rámci celonárodního projektu byl připraven standard XMetadiss obsahující, oproti ETD-MS, rozsáhlý německý národní metadatový set. Tento standard detailně popisuje disertační a habilitační práce včetně referencí na autory (MetaPers) s návodem na mapování na standard ETD-MS (Standard XMetadiss, 2006). Na základě německého standardu byly Komisi eVŠKP navrženy metadatové sety PersCZ pro popis fyzických osob (PersCZ, 2008) a CorpCZ pro popis korporací (CorpCZ, 2008) ve výměnných formátech, jako je např. metadatový set EVSKP-MS, příp. libovolný jiný. Návrh EVSKP-MS byl, i díky zastoupení většiny veřejných vysokých škol v Komisi eVŠKP, akademickou obcí široce přijat a na vysokých školách započal sběr eVŠKP včetně potřebné evidence popisných metadat podle tohoto standardu. Návrh

⁷² <http://www.dissonline.de/>

byl později doplněn o další především administrativní prvky do současné verze 1.1 (EVSKP-MS, 2008).

Komise eVŠKP ukončila svoji činnost v roce 2010 z důvodu úspěšného dokončení hlavních stanovených cílů. Komise eVŠKP doporučila následné řešení problematiky evidence, dlouhodobého uchování a podpory zpřístupňování výsledků vědy a výzkumu z produkce vysokých škol na nové platformě.

Repozitáře eVŠKP v České republice

Jak již bylo řečeno, vysoké školy vzhledem ke specifickým požadavkům na sběr eVŠKP včetně metadat a potřebě napojení na studijní agendu použily většinou pro lokální repozitáře samostatné řešení s následným exportem metadat do knihovních katalogů (On-line Public Access Catalog - dále jen **OPAC**). Jedny z prvních repozitářů vytvořila např. Akademie múzických umění s pomocí systému DSpace (OPAC Tinlib) a VŠE, která zvolila vlastní softwarové řešení v rámci webových stránek školy s exportem dat do knihovního systému Aleph.

Výhodou oddělení lokálního repozitáře od OPAC je vysoká flexibilita a možnost integrace do informačního systému vysoké školy. Nevýhodou je vynechání knihovníků z procesu odevzdávání eVŠKP. Důsledkem toho může docházet například k chybně vytvořeným anotacím psaných v první osobě nebo se subjektivním hodnocením a k nevhodné volbě klíčových slov. Vytváření předmětových hesel samotnými studenty není pro obtížnost doporučováno, a proto metadata **EVSKP-MS** počítají primárně s klíčovými slovy. Předmětová hesla podle různých řízených slovníků však mohou být dodatečně vložena knihovníky při importu metadat do OPAC. Pro ilustraci jsou v příloze č. 4 popsány výsledky průzkumů zveřejňování eVŠKP z let 2006 až 2009, které prováděla Komise eVŠKP.

V současné době většina českých vysokých škol používá pro sběr eVŠKP různé varianty řešení, kterými jsou informační studijní systém, knihovní katalog, systém **DSpace** nebo **Digitool** a odevzdávání eVŠKP přímo do systému **theses.cz**. Z lokálních repozitářů jsou eVŠKP dostupné většinou pouze v rámci školy nebo knihovny. Mezi preferovaný formát sběru dat patří **PDF** a **DOC**.

Nejrozsáhlejší databázi eVŠKP má MU, která již v roce 2000 začala s přípravou koncepce na bázi systému **NDLTD** v rámci Ústavu výpočetní techniky, ale projekt byl po roce pozastaven. Od roku 2004 informační systém MU poskytuje plnou podporu pro evidenci VŠKP a byla vůbec první, která začala eVŠKP plně zpřístupňovat veřejnosti v rámci svých webových stránek. Systém MU používá vlastní popis eVŠKP.

Komise eVŠKP iniciovala projekt vzniku **národního registru** eVŠKP, ale protože AKVŠ ani jí zřízené komise se nemohou účastnit dotačních programů jako celek, podali v roce 2006 žádost o financování projektu zájemci z řad vysokých škol samostatně. V té době však projekt nebyl odsouhlasen vzhledem k omezenému rozpočtu na centralizované rozvojové projekty. O rok později se MU na základě zájmu vysokých škol o své antiplagiátorské řešení rozhodla řešit projekt

podobný projektu národního registru eVŠKP připravovaného Komisí eVŠKP. Na základě doporučení došlo ke spojení sedmnácti vysokých škol a podání jednoho společného centralizovaného rozvojového projektu C1/2008 MŠMT *Národní registr vysokoškolských kvalifikačních prací a odhalování plagiátů*. V rámci projektu systém theses.cz v současnosti používá 29 vysokých škol. Zapojení a účast v systému theses.cz je i do budoucna na základě prohlášení řešitele a Komise eVŠKP pro české veřejné vysoké školy zdarma.

V současnosti systém theses.cz agreguje okolo 100 tisíc metadatových záznamů eVŠKP s odkazy na eVŠKP, které jsou dostupné buď přímo v systému theses.cz nebo v lokálních repozitářích vysokých škol. Kromě standardu EVSKP-MS a protokolu OAI-PMH akceptuje systém theses.cz i alternativní varianty sběru dat.

Nad metadaty v systému theses.cz je umožněno fulltextové vyhledávání a od roku 2010 i s podporou českého skloňování. Dle rozhodnutí jednotlivých škol může být eVŠKP dostupná v jednom z následujících režimů:

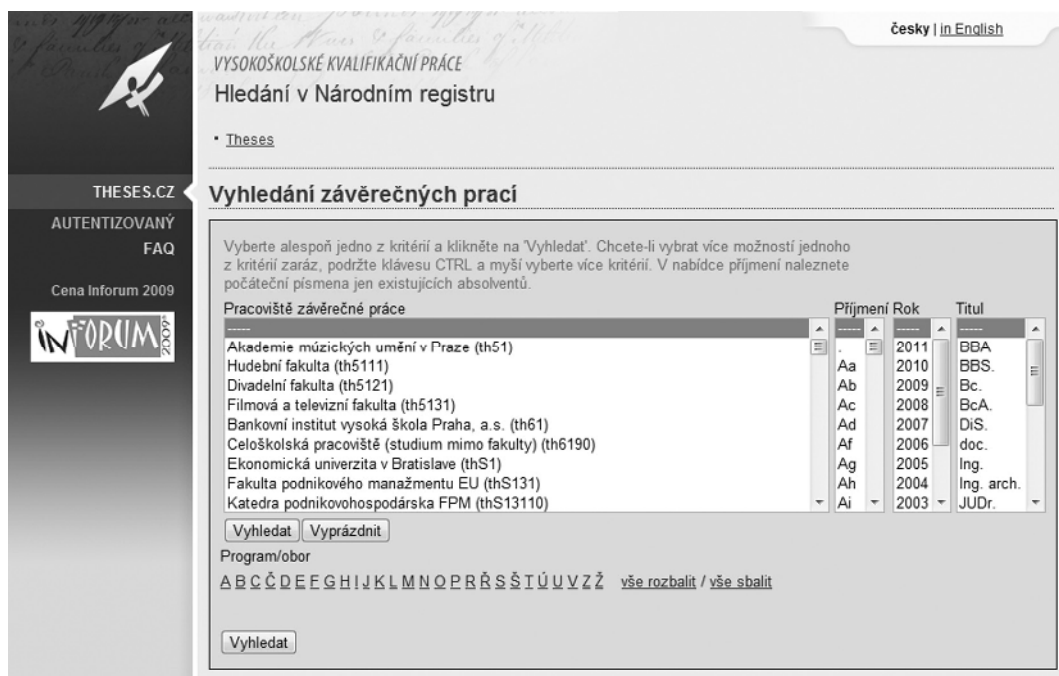
- práce je v theses.cz zveřejněna volně,
- práce je po přihlášení zpřístupněna v systému theses.cz libovolnému uživateli,
- práce je po přihlášení zpřístupněna v systému theses.cz pouze studentům a vyučujícím z příslušné školy,
- uživatel je přesměrován do lokálního repozitáře, kde přístup k plnému textu si řídí sama škola.

Systém theses.cz nemá v současné době⁷³ implementován protokol OAI-PMH na straně exportu, který by umožnil systému fungovat jako národní proxy server a automatizovaně předávat metadata ve sjednocené podobě do dalších národních či zahraničních systémů (např. NUŠL, DART-Europe aj.). Zájemci z řad vysokých škol, kteří chtějí zpřístupnit svá metadata v dalších repozitářích, musí nyní jednat s provozovateli těchto systémů samostatně a řešit konverzi a export metadat individuálně. Implementace otevřeného protokolu OAI-PMH v rámci systému theses.cz byla hlavnímu řešiteli projektu navržena spolupracujícími školami v rámci pokračování projektu.

Systém theses.cz pomáhá na základě porovnávání vložených textů eVŠKP, případně přednášek a seminárních prací, odhalovat projevy plagiátorství. Aplikace porovnává nově vložené práce s texty již uloženými v databázi a dokáže odhalit podobnosti i v případě, kdy došlo k drobné změně ze strany autora (např. záměna termínů nebo některé části textu apod.). K vložené práci je připraven v řádu několika hodin PDF dokument znázorňující shodné pasáže textu a odkaz na nalezené odpovídající dokumenty. Tyto dokumenty jsou dostupné buď veřejně nebo na vyžádání správcem příslušné vysoké školy, aby mohlo být posouzeno, zda nedošlo k plagiátorství. Rozhodnutí, zda se jedná

⁷³ Informace platná k polovině roku 2010.

o plagiátorství či řádnou citaci, je na konkrétním vyučujícím nebo správci dat za vysokou školu.



Obrázek 5: Národní registr vysokoškolských kvalifikačních prací, www.theses.cz

V roce 2009 byl programátorským týmem systému theses.cz v rámci centralizovaného rozvojového projektu vybudován obdobný systém Odevzdej.cz⁷⁴ pro odevzdávání [seminárních prací](#) a jejich kontrolu na projev plagiátorství. Vysoké školy mohou využít systém Odevzdej.cz přímo prostřednictvím webové stránky nebo implementovat napojení na systém Odevzdej.cz ve svém studijním systému. Na rok 2010 je plánováno rozšíření antiplagiátorského systému o porovnávání vůči textům na internetu. Toto rozšíření by významně pomohlo odhalit prohřešky studentů, kteří nejčastěji kopírují rozsáhlé pasáže z volně dostupných textů na internetu.

Tyto systémy z podstaty věci nemohou nalézt takové shody, kdy text byl autorem významně upraven či přeložen. Prevenci plagiátorství proto může velmi pomoci hrozba plagiátorům, že jejich text bude volně dostupný prostřednictvím internetu a jejich plagiátorství tak bude snáze prozrazeno. K volnému zpřístupnění eVŠKP prostřednictvím internetu se zatím přiklonilo pouze několik škol, jedna z posledních je UK požadující pouze on-line registraci čtenáře. Zjednodušení přístupu k eVŠKP, jejich integrace v centrálním repozitáři a podpora vyhledávání shodných pasáží je argumentem proti obavám, že zveřejněním eVŠKP by došlo k rozšíření nelegálního kopírování z dokumentů. Naopak nedostupnost eVŠKP mezi školami usnadňuje situaci [plagiátorům](#), kteří se tak méně obávají, že jejich čin bude odhalen.

⁷⁴ www.odevzdej.cz

Závěrem je důležité zmínit, že eVŠKP jsou nedílnou součástí sběru šedé literatury na národní úrovni v rámci projektu NUŠL. Prvními spolupracujícími vysokými školami s NUŠL jsou VŠE a UK. Pro všechny přístupující vysoké školy jsou připraveny na výběr tři formy spolupráce, přičemž výběr formy spolupráce je plně na vysokých školách. První dvě formy spolupráce nabízejí dlouhodobou archivaci a zpřístupnění eVŠKP. Pro naplnění těchto funkcí je nutné uložení dat přímo do centrálního digitálního repozitáře NUŠL, a to buď přímým vložením nebo předáním dat z repozitáře vysoké školy. Třetí formou spolupráce je zapojení repozitáře vysoké školy do centrálního vyhledávacího rozhraní NUŠL, které zajistí vyhledání dokumentů, ale již nezajistí dlouhodobou archivaci a zpřístupnění. Všechny tři formy spolupráce jsou rozebrány ve čtvrté kapitole. NUŠL zajistí vysokým školám nejen vyhledatelnost, zpřístupnění a dlouhodobou archivaci dat, ale také jejich začlenění do mezinárodních registrů šedé literatury. V rámci projektu NUŠL byla spolupráce s vysokými školami plánována od začátku řešení, proto, jak se dočtete v sedmé kapitole, byl metadatový formát VSKP-MS jedním z hlavních výchozích formátů pro metadatový formát NUŠL.

Sběr a zpřístupnění VŠKP se liší od ostatních druhů šedé literatury také z pohledu českého práva, proto je VŠKP věnována v následující kapitole celá část Školní díla jako šedá literatura.

LITERATURA:

GRULICH, Petr. *Vysokoškolské kvalifikační práce jako specifický typ archiválie a jejich digitalizace na Univerzitě Hradec Králové*. [online]. 2005. [cit. 2010-06-10]. Dostupné z:

<<http://www.cesarch.cz/nastenka/2005/chrudim/grulich.doc>>

HOROVÁ, Iva. *Aktuální stav zpřístupňování VŠKP v ČR. Výsledky průzkumu z prosince 2007*. [on-line] (PDF - 246.9kB) 6. 3. 2008. [cit. 2010-06-10].

Dostupné z [www: <http://www.evskp.cz/Dokumentyver/pruzkum2007-080306104121.pdf>](http://www.evskp.cz/Dokumentyver/pruzkum2007-080306104121.pdf)

HOROVÁ, Iva - KRKOŠKOVÁ, Jarmila. *Aktuální stav zpřístupňování VŠKP v ČR. Výsledky výzkumu*. [on-line] (PDF - 266.8kB) 14. 3. 2007. [cit. 2010-06-10].

Dostupné z [www:](http://www.evskp.cz/Dokumentyver/vyzkum-070622150034.pdf)

<<http://www.evskp.cz/Dokumentyver/vyzkum-070622150034.pdf>>

K otázkám budování systémů pro zpřístupňování elektronických vysokoškolských kvalifikačních prací v České republice. Stanovisko Asociace knihoven vysokých škol. [on-line] (PDF) 31. 3. 2004. [cit. 2010-06-10]. Dostupné z [www: <http://www.akvs.cz/pdf/vskpshrnuti.pdf>](http://www.akvs.cz/pdf/vskpshrnuti.pdf)

VYČÍTALOVÁ, Lucie. *Aktuální stav zpřístupňování vysokoškolských kvalifikačních prací v ČR. Výsledky průzkumu z října 2009*. [on-line] (PDF)

Duben 2010. [cit. 2010-06-10]. Dostupné z [www:](http://www.evskp.cz/Dokumentyver/pruzkum2009-100430133025.pdf)

<http://www.evskp.cz/Dokumentyver/pruzkum2009-100430133025.pdf>>

Zákon č. 111/1998 Sb., zákon o vysokých školách a o změně a doplnění některých dalších zákonů (o vysokých školách).

Zákon č. 552/2005 Sb., kterým se mění zákon č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů, a některé další zákony.

Zákon č. 216/2006 Sb., kterým se mění zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, a některé další zákony.

ELEKTRONICKÉ ZDROJE:

CorpCZ: metadatový soubor pro popis korporací [online]. Zpracovatelé Eva Bratková, Jan Mach. Verze 1.0. Praha: Odborná komise pro otázky elektronického zpřístupňování VŠKP AKVŠ ČR. [cit. 2008-05-05]. 13 s. Dostupné z www: <<http://www.evskp.cz/standardy/corpcz/1.0/>>.

DART-Europe E-theses Portal [online]. The DART-Europe E-theses Portal is endorsed by LIBER. c1999-2010. [cit. 2010-06-10]. Dostupné z www: <<http://www.dart-europe.eu/basic-search.php>>

DissOnline [online]. Digitale Dissertationen im Internet. Aktualizováno 12. 3. 2009. [cit. 2010-08-05]. Dostupné z www: <<http://www.dissonline.de/>>

ETD - db [online]. Digital library and archives. [cit. 2010-08-10]. Dostupné z www: <<http://scholar.lib.vt.edu/ETD-db/index.shtml>>

ETD-MS: an Interoperability Metadata Standard for Electronic Theses and Dissertations [online]. Aktualizováno 25. 6. 2008. [cit. 2010-06-10] Dostupné z www: <<http://www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.html>>

EVSKP-MS: metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR [online]. Zpracovatelé Eva Bratková, Jan Mach. Verze 1.1. Praha: Odborná komise pro otázky elektronického zpřístupňování VŠKP AKVŠ ČR. [cit. 2008-07-15]. Dostupné z www: <<http://www.evskp.cz/standardy/evskp/1.1/>>.

Metadatový standard EVSKP-MS, v. 1.1 pro popis VŠKP a standardy související [on-line]. Eva Bratková, Jan Mach. *Systémy pro zpřístupňování VŠKP: zkušenosti, možnosti, nabídky, potřeby: 3. ročník semináře konaného 7. 10. 2008 na VUT v Brně* [on-line]. Praha: Odborná komise pro otázky elektronického zpřístupňování VŠKP AKVŠ ČR, 2008-12-22 [cit. 2010-08-10]. Dostupný z www: <<http://www.evskp.cz/Seminar3/seminar3-BratkovaMach.pdf>>

PersCZ: metadatový soubor pro popis fyzických osob [online]. Zpracovatelé Eva Bratková, Jan Mach. Verze 1.0. Praha: Odborná komise pro otázky elektronického zpřístupňování VŠKP AKVŠ ČR. [cit. 2008-05-05]. 1 s. Dostupné z www: <<http://www.evskp.cz/standardy/perscz/1.0/>>.

Prezentace. Celostátní porada vysokoškolských knihoven [online]. Hradec Králové, 19. - 20. 11. 2003. [cit. 2010-06-10]. Dostupné z www: <<http://lib.uhk.cz/cpvsk2003/index.php?id=prezentace>>

RIV. Rejstřík informací o výsledcích [online]. Rada pro výzkum, vývoj a inovace. Aktualizace 10. 5. 2010. c2010. [cit. 2010-08-27]. Dostupné z www: <http://www.vyzkum.cz/FrontClanek.aspx?idsekce=956>

Standard XMetadiss. Verze 1.3. [online]. Metadata Set of the German National Library for Online. 22. 8. 2006. [cit. 2010-06-10]. Dostupné z www: http://www.d-nb.de/eng/standards/pdf/ref_xmetadiss_v1-3.pdf

NDLTD [online]. The National Digital Library of Theses and Dissertations. Aktualizováno 15. 3. 2010. [cit. 2010-08-10] Dostupné z www: <<http://www.ndltd.org/about/history>>

VTLS Visualizer [online]. VTLS. c2008. [cit. 2010-08-05]. Dostupné z www: <http://www.vtls.com/products/visualizer>

6. Právní aspekty šedé literatury

Radim Polčák

Pojem šedé literatury z pohledu práva

Pojem šedé literatury je z právního hlediska relativně složitý, neboť platné **právo** se mu specificky nevěnuje. Už samo užití výrazu „šedý“ vyvolává v právních souvislostech spíše představu něčeho nekalého nebo alespoň pochybného, podobně jako v případě výrazů „šedý dovoz,“ nebo třeba „šedá ekonomika.“ Ve skutečnosti však šedá barva v tomto případě neznačí přechodnou fázi mezi legálním a ilegálním, ale spíše naznačuje specifický objekt právního zájmu vykazující na jedné straně typické znaky standardní literatury, na straně druhé však mající i znaky a účely, které jsou pro právo relativně nové a netypické. Šedivost tak je z pohledu právního v tomto případě třeba vnímat nikoli jako znak nebezpečí, ale spíše jako indikátor potřeby pozorné právní úvahy.

Standardní právní model tvorby a publikace literatury (můžeme ji označit třeba jako „bílou literaturu“) je založen na kreativní činnosti autora, který pak následně své dílo licencí postupuje vydavateli a ten jej komerčním způsobem uvádí na trh. Takové **literární dílo** se standardně šíří tiskem a zákazníci za jeho rozmnoženiny platí v knihkupectvích kupní cenu. Z ní pak autor zpravidla inkasuje příslušný podíl, přičemž je z ní pochopitelně financována i činnost vydavatele a knihkupce.

Příchod informačních a komunikačních technologií tento klasický model v poslední době čím dál tím více mění, a to nejen co do samotného média (setkáváme se běžně s tzv. elektronickými knihami),⁷⁵ ale i co do ekonomických či obecně transakčních mechanismů tvorby a šíření literárních děl. Kromě literárních děl, u nichž je primárním záměrem autora „prodat“ je vydavateli, respektive zajistit jim co největší komerční odbyt, se tak stále častěji setkáváme s díly, jejichž primárním účelem je nekomerčním způsobem šířit určitou informaci,⁷⁶ dostát právní nebo akademické povinnosti, nebo dokonce taková, jejichž smysl a účel můžeme vidět i jen v tom, že prostě jen vzniknou.

Máme-li tedy poměrně široký a značně metaforický pojem šedé literatury poněkud upřesnit pro potřebu jeho následné právní reflexe, můžeme tak učinit negativně - namísto pozitivního vymezení toho, co je šedou literaturou, tak

⁷⁵ Ke změnám souvisejícím s možnostmi elektronické publikace knih viz např. PICKER, R. Mediated Book, John M. Olin Law and Economics working Paper No. 463. Chicago: The University of Chicago, 2009.

⁷⁶ K tomu srov. např. AUGER, Ch., P. Information sources in grey literature, 2nd edition. London: Bowker-Saur: New York, 1989.

můžeme naopak říci, že šedou literaturu tvoří taková literární díla, jejichž smysl a účel je odlišný od standardní literatury šířené nakladatelským způsobem. Jakékoli bližší vymezení šedé literatury či i jen pokus o její pozitivní definici musejí totiž za současného bouřlivého vývoje v tomto oboru nutně skončit neúspěchem. Vždyť ani sami knihovníci, které můžeme považovat za hlavní teoretiky konceptu šedé literatury, nemají v této věci úplně jasno.⁷⁷

Z hlediska právního každopádně není třeba mít stoprocentní jistotu v tom, co všechno může být považováno za šedou literaturu. Spíše je nutné zaměřit se na otázku, jaké právní režimy mohou mít nejrozumnější šedé literární výstupy. K tomu je třeba provést klasifikaci typů šedé literatury nikoli dle jednotlivých jejích forem, ale ideálně dle účelů, k jejichž naplnění vznikají.⁷⁸ Prozatím se jako dostačující jeví uvažovat o šedé literatuře jako o literárních dílech plnících primárně některý z následujících účelů:

1. plnění studijních či kvalifikačních povinností,
2. vykazování akademických aktivit (nejčastěji ve výzkumu a vývoji pro potřeby poskytovatelů dotací),
3. sdílení myšlenek k odborné diskusi (typicky při konferencích, v procesu psaní rozsáhlejších publikací, při výuce apod.),
4. vytváření technických standardů,
5. plnění právních povinností (typicky v oblasti účetnictví, veřejných zakázek apod.).

Z hlediska subjektů a právních režimů tvorby a šíření šedé literatury pak můžeme k výše uvedeným kategoriím přiřadit následující typické situace:

1. literární dílo vytváří student nebo uchazeč (typicky sem patří práce seminární, diplomové, habilitační apod.) - většinou se jedná o díla školní, respektive o díla se zvláštním režimem užití dle zákona o vysokých školách,
2. literární dílo vytváří výzkumník nebo výzkumný tým pro svého zaměstnavatele, resp. pro poskytovatele dotace, případně toto dílo (typicky u závěrečných zpráv) vytváří najatá agentura - většinou jde o díla zaměstnanecká nebo o díla vytvořená na zakázku,
3. literární dílo vytváří nezávislý autor a sám jej pak i publikuje - jde o standardní individuální autorská díla,

⁷⁷ Jako jeden z neúspěšných pokusů vymezit šedou literaturu jako pojem viz např. GOKHALE, P. Grey Literature Varieties - Definitional Problems, in Third International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information. 13. - 14. 11.1997, Amsterdam: GreyNet, 1998, s. 259-273.

⁷⁸ Účel představuje v platném právu jednu z centrálních kategorií. Přestože nebývá nutně v textech právních předpisů vyjádřen, je ve struktuře platného práva obsažen implicitně a umožňuje adekvátní aplikaci platného práva i na případy, s nimiž sám právo tvůrce v době vzniku příslušné právní úpravy ani nepočítal. K tomu srov. např. HOLLÄNDER, P. Filosofie práva. Plzeň: Aleš Čeněk, 2006, s. 74-93.

4. dokument vytváří úředník nebo zaměstnanec profesní organizace či profesního sdružení - v typických případech není nutné řešit režim ochrany, neboť se zpravidla nejedná o autorská díla,
5. dokument ve formě výkazu nebo dokladu vytváří zpravidla zaměstnanec pro svého soukromoprávního nebo veřejnoprávního zaměstnavatele - dtto.

Přestože se o výše uvedených formách šedé literatury někdy hovoří jako o literatuře nepublikované, z hlediska právního se ve valné většině případů jedná o literární díla zveřejněná, byť často nikoli nakladatelsky a tiskem. Ke zveřejňování šedé literatury tak dochází jednak specifickými formami dle typu díla (například zveřejněním závěrečné zprávy na webu, veřejnou obhajobou **VŠKP** apod.) a dále pak stále častější obecnou formou, tj. publikací v národních nebo mezinárodních databázích a registrech.

Právě tato standardní forma zpracování šedé literatury se z právního hlediska jeví jako nejvíce problematická. Provozovatel příslušné databáze zde v tomto případě sice na jedné straně fakticky zveřejňuje příslušné literární dílo, na straně druhé však nemá (a ani nechce mít) právní postavení srovnatelné s nakladatelem. Smysl a účel databází šedé literatury je tedy diametrálně odlišný od účelu různých forem nakladatelských publikací a v návaznosti na to se ve značné míře liší od nakladatelského modelu i právní a faktické postavení zúčastněných stran.

Obecné právní otázky zpracování šedé literatury

Z právních souvislostí tvorby a publikace šedé literatury budou probírány, jak naznačeno shora, především otázky autorskoprávní. Literární dílo je v tomto směru chráněno již od svého vytvoření, tj. od jeho zachycení v objektivně vnímatelné podobě. **Autorská práva** k dílu tak vznikají a jsou chráněna ze zákona, a to dokonce bez ohledu na vůli samotného autora.⁷⁹ Tato práva zahrnují osobnostní a majetkovou složku (§ 10 autorského zákona).

Osobnostní složka autorských práv se váže bezprostředně k osobě autora a zahrnuje právo osobovat si **autorství**, tj. být uveden jako autor (§ 10 autorského zákona), právo na integritu díla (§ 11 odst. 3 autorského zákona), tj. ochranu před zásahy do struktury a obsahu díla a právo rozhodnout o tom, zda bude dílo zveřejněno či nikoli (§ 11 odst. 1 autorského zákona). Pro pořizování a vedení databází šedé literatury je však mnohem důležitější majetková složka autorských práv, která zahrnuje práva příslušné dílo užívat. Katalog různých forem užití je relativně rozsáhlý a obsahuje prakticky všechny

⁷⁹ Autor tak nemusí své dílo nikde registrovat ani se jinak aktivně o jeho ochranu zasazovat. Autor dokonce ani nemá de iure možnost jednostranným jednáním ochranu svého díla vyloučit nebo omezit - viz dále.

myslitelné typy manipulace s autorským dílem včetně zveřejnění,⁸⁰ šíření tiskem, kopírování apod.

Zahrnutí autorského díla v uzavřené nebo veřejné databázi šedé literatury je nepochybně možné považovat za jeho užití ve smyslu zákona. K tomu, aby mohl pořizovatel nebo provozovatel databáze určité literární dílo tímto způsobem užít, je třeba, aby k tomu disponoval příslušným oprávněním. To může vyplynout přímo ze zákona, kdy hovoříme o volných užitích, resp. zákonných licencích (oddíl 2 autorského zákona), nebo může být uděleno **autorem** či jiným vykonavatelem majetkových práv autorských ve formě smluvní licence.

V uvedených souvislostech je důležité si uvědomit, že nekryje-li možnost užití autorského díla **zákonná licence** nebo **volné užití**, lze dílo užít jen na základě dohody se subjektem vykonávajícím k dílu **majetková autorská práva**. Za takovou dohodu však nelze považovat jednostranné prohlášení nebo například skutečnost, že své dílo vystavil autor na volně dostupném webu - licence tak musí mít povahu dvoustranného právního úkonu. V případě projektu NUŠL to znamená, že NTK (jako pořizovatel databáze) uzavírá licenční smlouvy se všemi producenty dat nezávisle na způsobu získání díla, jak je zmíněno ve čtvrté kapitole.

Poněkud problematická situace, kdy autor vlastně nemůže dát jednostranný souhlas s dalším užíváním svého díla,⁸¹ se negativně projevuje právě u šedé literatury. Zejména u individuálních autorských děl mají jejich autoři eminentní zájem na tom, aby se jejich dílo co nejvíce rozšířilo a chtějí k tomu využít i typicky masivních nepřímých **síťových efektů** internetu. Dát neurčitému okruhu zájemců jednostranně možnost dílo kopírovat a dál šířit však dle platného práva nelze.

Přirozeným, i když jen nepřímým řešením tohoto problému, jsou takzvané **volné licence**.⁸² Technicky se jedná o veřejnou nabídku na uzavření licenční smlouvy adresovanou neurčitému okruhu zájemců. Skutečnost, že je příslušné dílo nabízeno pod volnou licenci, je v díle samotném nebo v jeho blízkosti oznámena⁸³ a zájemce o kopírování a další šíření díla je tak informován o tom, že jeho užitím přistupuje konkludentně na podmínky licence. Záleží pak na typu licenčního ujednání, zda autor (nebo vykonavatel majetkových práv autorských) umožní jen další volné nevýdělečné šíření díla (tento typ ujednání se používá standardně) nebo zda například svolí i s jeho výdělečným užitím, změnou, se zpracováním do cizího díla apod. Přestože je s využitím institutu

⁸⁰ Zde je třeba rozlišovat mezi rozhodnutím o zveřejnění (patří mezi osobnostní práva) a samotným zveřejněním (různé formy spadají pod majetková práva).

⁸¹ V tomto případě jde o jeden z paradoxů práva duševního vlastnictví, k jejichž odkrytí dochází s postupným rozvojem informační společnosti - k tomu srov. např. LESSIG, L. Free Culture. New York: The Penguin Press, 2004.

⁸² K podstatě volných licencí viz např. GUADAMUZ, A. Open Science: Open Source Licences in Scientific Research. North Carolina Journal of Law and Technology, roč. 7, č. 2, s. 321-366.

⁸³ Pro tento účel se používají standardizované formulace, ikony či piktogramy s odkazem na www stránky s plným zněním licenční smlouvy.

volných licencí spojena řada právních problémů,⁸⁴ získal si tento instrument značnou oblibu, a to nejprve v USA a krátce na to i v Evropě, Českou republiku nevýmaje.⁸⁵ Značná popularita volných licencí a jejich mediální známost společně s neznalostí právních souvislostí však často vedou k situacím, kdy dílo pod volnou licenci nabízí subjekt, který však nedisponuje odpovídajícím katalogem práv.

Ať už má být příslušné literární dílo užíváno na základě volné nebo standardní licence, je třeba rozlišovat, kdy vykonává majetková práva autorská sám autor a kdy je jejich vykonavatelem jiný subjekt. Například tam, kde bylo dílo vytvořeno jako zaměstnanecké (viz dále), nebo tam, kde autor již dříve dohodnul s někým výhradní licenci, nemá sám autor, byť to může znít paradoxně, právo s dílem dále nakládat. Autor, který publikoval článek v časopise a uzavřel s jeho vydavatelem exkluzivní licenční smlouvu, už není oprávněn poskytnout poté své dílo například ke zpracování v databázi šedé literatury (s takovým užitím by musel souhlasit vydavatel časopisu, s nímž autor dříve uzavřel **exkluzivní licenci**)⁸⁶ nebo jej dát k dispozici pod volnou licenci.

Z právě uvedeného vyplývá, že zařadit určité autorské dílo do databáze šedé literatury lze pouze na základě práva vyplývajícího přímo ze zákona (typicky u zákonných licencí, zaměstnaneckých děl apod.) nebo na základě licenční smlouvy. Nejedná-li se ani o jeden z právě uvedených případů, jde o protiprávní užití autorského díla a pořizovatel příslušné databáze se vystavuje možnosti právního postihu (díl 5 autorského zákona).

Zásah do majetkových práv autorských formou protiprávního užití se hodnotí na objektivní bázi, tj. za užití objektivních faktických a právních kritérií.⁸⁷ Znamená to, že se nezkoumá míra zavinění na straně rušitele (v našem případě pořizovatele databáze šedé literatury), ale pouze skutečnost, zda k zásahu do majetkových práv autorských objektivně došlo.⁸⁸ Nehraje tedy roli například skutečnost, že pořizovatel databáze jednal v dobré víře, tj. například se mylně domníval, že autor poskytující do databáze svoje odborné články k nim vykonává majetková práva autorská. V takovém případě nezbývá pořizovateli

⁸⁴ Volné licence, které mají ambici pokrývat záraz různé státní jurisdikce, tak narážejí na rozdíly v národních právních řádech, na kogentní ustanovení rigorózně chránící kolektivní správu autorských práv apod. - k tomu srov. např. MARACKÉ, C. Creative Commons International The International License Porting Project. *Journal of Intellectual Property. Information Technology and E-Commerce Law*, č. 1, s. 4-18.

⁸⁵ Po delší prodlevě byla do českého právního prostředí nyní importována i nejznámější z volných licencí pro literární a jiná mediální díla, Creative Commons. Informace o licenci a stručný návod k použití lze nalézt na www.creativecommons.cz.

⁸⁶ Opačný postup však možný je, tj. autor nejprve poskytne svůj manuskript do databáze šedé literatury a pak (typicky poté, co shromáždí ohlasy a text dopracuje) se dohodne na jeho exkluzivní publikaci v časopise nebo formou knihy. Nevýhradní licence, na jejímž základě je původní verze článku dostupná v databázi šedé literatury, přitom stále zůstává v platnosti. K tomu srov. § 47 odst. 4 autorského zákona.

⁸⁷ K pojmu objektivní právní odpovědnosti viz např. HAVRÁNEK, J. et al. *Teorie práva*. Plzeň: Aleš Čeněk, 2008, s. 392.

⁸⁸ K tomu srov. TŮMA, P. *Autorský zákon - Komentář*. Praha: C.H. Beck, 2007, s. 421.

databáze šedé literatury než odškodnit poškozeného (tj. legitimního vykonavatele autorských práv) a regresně se pak domáhat náhrady na tom, kdo celou situaci svým jednáním způsobil (tj. na autorovi nebo na jiné osobě, která dokument neoprávněně do databáze nahrála). **Odpovědnosti** vůči třetím osobám se přitom pořizovatel databáze nemůže zříci ani ji nijak jednostranně omezit.⁸⁹

V souvislosti s právě uvedeným je třeba připomenout i prastarou zásadu soukromého práva, že totiž nikdo nemůže na jiného převést více práv, než kolika sám disponuje. Subjekt, který nahraje do databáze šedé literatury autorské dílo aniž by k němu disponoval odpovídajícími majetkovými právy, tedy logicky nemůže *de iure* udělit oprávnění k užití díla v databázi jejímu pořizovateli. Přestože tedy může **pořizovatel databáze** žít v domněnku, že mu byla udělena platná licence, ve skutečnosti zpracovává a zveřejňuje příslušné dílo v rozporu se zákonem a poškozené třetí osobě (tj. legitimnímu vykonavateli majetkových práv autorských) za to přímo odpovídá.

Jedinou podstatnou výjimkou z právě uvedeného je situace, kdy je repozitář šedé literatury provozován jen jako veřejné úložiště, tj. jako služba, která uživatelům pouze umožňuje ukládání jejich dat, aniž by s nimi dále jakkoli manipulovala. Je-li pak příslušné dílo nahráno uživatelem do takového úložiště, je odpovědnost poskytovatele této služby omezena na případy, kdy o protiprávnosti užití díla sám ví, nebo vědět má a může.⁹⁰

Školní díla jako šedá literatura

Žákovské nebo studentské práce tvoří významnou součást kmene šedé literatury. Jedná se o literární díla, která vznikají v průběhu studia a jejichž účelem je prokázat splnění studijních povinností nebo dosažení dostatečné úrovně odbornosti k udělení akademického titulu. Patří sem tedy nejrůznější formy průběžných studijních plnění, jako jsou např. seminární nebo semestrální práce, ale též VŠKP, jakými jsou práce bakalářské, diplomové apod., které jsou podrobně popsány v páté kapitole.

Z hlediska režimu užití žákovských nebo studentských prací v databázích šedé literatury je v první řadě rozhodné, zda jde o **školní dílo** bez dalšího, nebo zda se jedná o školní dílo spadající též pod rozsah pojmu závěrečné práce dle zákona o vysokých školách.⁹¹ Na školní díla, která nenaplňují zákonnou definici závěrečné práce, tj. např. na práce seminární, se vztahuje autorskoprávní

⁸⁹ Zde je třeba připomenout, že nejrůznější tzv. „disclaimers“ jednostranně oznamující, že provozovatel služby neodpovídá za porušení práv třetích osob, jsou v evropském právu ve valné většině případů právně irelevantní.

⁹⁰ Provozovatel repozitáře tedy v tomto případě není pořizovatelem databáze, ale poskytovatelem služby ukládání uživatelských dat ve smyslu § 5 zákona č. 480/2004 Sb. K tomuto typu odpovědnosti podrobněji viz POLČÁK, R. Právo na internetu - spam a odpovědnost ISP. Brno: Computer Press, 2007, s. 68 a násl.

⁹¹ Závěrečnými pracemi se v tomto směru rozumí bakalářské, diplomové, rigorózní a disertační práce (srov. § 47b odst. 1 zákona o vysokých školách).

režim školního díla dle autorského zákona. Student či žák v tomto případě vykonává v plném rozsahu osobnostní i majetková autorská práva,⁹² přičemž škola může takové dílo užít, stručně řečeno, pro vlastní potřebu.⁹³ Právo školy užít školní dílo svého žáka nebo studenta však nezahrnuje oprávnění jej zveřejnit nebo jej jinak dál šířit. K tomu, aby mohla škola například zahrnout takové dílo do své veřejně přístupné databáze šedé literatury nebo jej předat ke zpracování do jiného repozitáře, je tedy potřeba dobrovolného licenčního ujednání.⁹⁴

U školních děl spadajících i pod rozsah pojmu VŠKP zákona o vysokých školách, tj. u prací bakalářských, magisterských, rigorózních a disertačních, je režim výkonu majetkových autorských práv mírně modifikován. Změnu v tomto směru přinesla novela zákona o vysokých školách, která zavedla vysokým školám povinnost zveřejňovat VŠKP prostřednictvím knihovny, resp. prostřednictvím veřejně přístupné databáze. Pro potřeby zpracování VŠKP jako šedé literatury je důležité především ustanovení nového § 47b odst. 1 následujícího znění:

„Vysoká škola nevýdělečně zveřejňuje disertační, diplomové, bakalářské a rigorózní práce, u kterých proběhla obhajoba, včetně posudků oponentů a výsledku obhajoby prostřednictvím databáze kvalifikačních prací, kterou spravuje. Způsob zveřejnění stanoví vnitřní předpis vysoké školy.“

Tato povinnost zveřejňovat po obhajobě (úspěšné či neúspěšné) VŠKP prostřednictvím veřejně dostupné databáze v sobě obsahuje i implicitní dovolu (licenci) pro vysokou školu k pořízení veřejné databáze VŠKP a zpracování jednotlivých školních děl. Rozsah možností užití školních děl se tedy tímto ustanovením rozšiřuje pro VŠKP též o zákonnou licenci k jejich zahrnutí do databáze a následnému jejich zveřejnění touto formou.

Implicitní licence k výše uvedenému užití VŠKP v sobě samozřejmě nese i možnost pro vysokou školu, neodporuje-li to jejímu vnitřnímu předpisu poskytnout příslušné kvalifikační práce do veřejných databází šedé literatury vedených třetími osobami. V takovém případě je však třeba pamatovat na princip zmíněný na konci předchozí podkapitoly, totiž že nelze dát třetí osobě více práv, než kolik jich má sám převodce. Vysoká škola tak není oprávněna například bez souhlasu autora poskytnout VŠKP do výdělečných databází nebo je výdělečně či nevýdělečně vydávat tiskem.

⁹² V české akademické obci se v tomto směru traduje mýtus ohledně spoluautorství učitelů nebo dokonce škol - vzhledem k tomu, že studentské práce jsou (musí být) vždy původními výtvary samotných studentů, nemůže být o spoluautorství učitelů dozirajících na tvorbu prací řeč. Zasluky učitele o vytvoření díla tedy mohou být ohodnoceny například poděkováním ze strany žáka či studenta, nikoli však podílem učitele na výkonu osobnostních či majetkových práv autorských.

⁹³ Meze užití jsou v tomto případě dány § 35 odst. 3 autorského zákona. Mezi dovolené způsoby užití v tomto případě patří například užití díla při výuce nebo jeho zahrnutí do referenční databáze pro kontrolu plagiátů. Podrobněji k tomu srov. např. TELEČ, I. - TŮMA, P. Autorský zákon - Komentář. Praha: C.H. Beck, 2007, s. 380 a násl.

⁹⁴ Škola má při sjednávání licence sice zvláštní postavení založené § 60 odst. 1 autorského zákona, to ji ale samozřejmě neopravňuje k přímému užití díla.

Krátkou poznámku je v tomto směru třeba věnovat ještě dvěma problematickým momentům, které běžně zaznamenáváme v každodenní vysokoškolské praxi, a to zveřejňování specificky chráněných informací a uzavírání licenčních smluv se studenty. První jmenovaný problém je palčivý především v oborech, kde VŠKP vznikají někdy i za užití informací se zvláštními režimy ochrany - může se tak příkladně jednat o **osobní údaje**, **obchodní tajemství**, know-how, **utajované skutečnosti** apod. Přestože zákon z povinnosti zveřejňovat VŠKP nečiní výjimky, je zřejmé, že u prací s výskytem specificky chráněných informací by tupé trvání na naplnění textu zákona vedlo k nesmyslným následkům. Požadavek na zveřejňování všech prací bez výjimky by tak zprostředkovaně vedl k poškození studentů a samotných vysokých škol, neboť by odrazoval nejruznější subjekty, s nimiž školy a jejich studenti na přípravě VŠKP spolupracují.

Proporcionální interpretace zákona o vysokých školách nás tedy v této otázce vede k závěru, že ve specificky odůvodněných případech je možné přijmout výjimku z požadavku na zveřejnění VŠKP, a to v rozsahu nezbytně nutném k ochraně informací ve zvláštním režimu. Je přitom vhodné upravit postup žádosti o vynětí práce z povinného režimu zveřejnění, resp. zahrnutí do databáze šedé literatury vnitřním předpisem příslušné školy a technologií příslušné dostupné databáze, a to tak, aby každá žádost byla individuálně odůvodněna⁹⁵ a aby k omezení veřejného přístupu k VŠKP došlo jen v nezbytně nutné míře.

Druhý výše zmíněný problém související s užíváním školních děl se týká uzavírání licenčních smluv mezi školami a jejich studenty. Na některých vysokých školách se tak vyskytují praktiky, kdy je všem studentům předkládána k podpisu licenční smlouva opravňující vysokou školu dále bez omezení užívat příslušné VŠKP. V takových případech však lze mít důvodné pochybnosti o perfekci smluv, a to především co do splnění požadavku na **autonomii vůle** smluvních stran. Studenti v takových případech nemají, zjednodušeně řečeno, na výběr a předložené smlouvy podepisují, to však nikoli z důvodu vůle k poskytnutí svého díla škole, ale kvůli obavě o úspěšné ukončení studia. Smlouvy pak v takových případech mohou být právně vadné a užití VŠKP se zde řídí výše zmíněnou zákonnou licencí.

Zaměstnanecká díla jako šedá literatura

Podstatnou a důležitou součástí šedé literatury jsou nejruznější pravidelné nebo mimořádné zprávy z činnosti akademických institucí, projektových týmů nebo jednotlivých výzkumníků. Výkaznictví je v akademické praxi důležité

⁹⁵ Nabízí se například procedura individuálně odůvodněné žádosti podané vedoucím práce a posuzované vedoucím ústavu nebo příslušným proděkanem. Nemá však smysl uvažovat o vyřazení takových prací z databází, ale pouze o omezení veřejného přístupu - přítomnost práce v databázi, její zpřístupnění na individuálně odůvodněnou žádost nebo její zahrnutí do systému pro kontrolu plagiátů tím není třeba omezovat.

nejen kvůli motivaci a dokládání finančních operací, ale slouží též v odborné rovině k zachycení výsledků vědecké a výzkumné činnosti.

Výkazy a zprávy jsou ve valné většině případů vytvářeny pracovníky akademických institucí, případně též soukromoprávních korporací čerpajících veřejné prostředky na vědu a výzkum, a následně jejich zaměstnavatelem předkládány poskytovatelům finančních prostředků nebo přímo veřejnosti. Autorskoprávní režim takových dokumentů pak většinou spadá pod rozsah zvláštní kategorie takzvaných zaměstnaneckých autorských děl (§ 58 autorského zákona).

Přestože i v případě zaměstnaneckého díla náleží veškerá osobnostní i majetková práva autorovi, dává zákon zaměstnavateli ve vztahu k takovému dílu výlučné právní postavení. Zaměstnavatel tak nejen že exkluzivně vykonává majetková práva autorská, ale má též právo zasahovat do práv osobnostních, tj. příslušné dílo měnit, doplňovat a zveřejňovat pod vlastním jménem (nikoli tedy nutně pod jménem samotného autora). Výroční zpráva vysoké školy, byť je sepsána společně konkrétními akademickými a neakademickými pracovníky, tak může být zveřejněna s autorskou doložkou označující za autora pouze vysokou školu.

Autor zaměstnaneckého díla, jak vyplývá z právě uvedeného, není oprávněn sám o užití svého díla rozhodovat a nemůže jej tedy nabízet k publikaci či jej bez vědomí zaměstnavatele poskytnout například do databáze šedé literatury. Pokud tedy autor nakládá se zaměstnaneckým dílem, lidově řečeno, na vlastní pěst, vystavuje sebe i případné další subjekty manipulující s tímto dílem nebezpečí právního postihu ze strany vykonavatele majetkových autorských práv, tj. zaměstnavatele.

V akademické praxi nebývá obvykle problém s užíváním zaměstnaneckých děl v podobě výročních nebo grantových zpráv, ale spíše se stanovením režimu u literárních děl, jež vznikají jako následné produkty akademických činností. Nejde tedy v tomto případě o výkazy, ale o takřikajíc meritorní literární díla, jakými mohou být články, učebnice či monografie vzniklé v přímé vazbě k pedagogickému nebo vědeckému působení autora na příslušné akademické instituci.

Pokud bychom přijali tezi, že všechno, co **zaměstnanec** akademické instituce vyprodukuje v oboru, pro který je zde zaměstnán, je zaměstnaneckým autorským dílem, jednalo by se zřejmě o neadekvátně extenzivní interpretaci § 58 odst. 1 autorského zákona. Skutečnost, že jim **zaměstnavatel** „sebere“ majetková práva autorská ke všemu, co ve svém oboru kreativně vytvoří, by byla zřejmě navíc pro potenciální autory značně demotivující.

Přestože nelze formulovat jednoznačnou hranici pro klasifikaci určitých literárních děl jako děl zaměstnaneckých, v praxi se jako rozumné řešení prosazuje přístup praktikovaný standardně našimi nejvýznamnějšími vysokými školami, tj. posuzování literárních děl jako zaměstnaneckých pouze v případě, jsou-li vytvářena na přímý pokyn zaměstnavatele. Do této kategorie tak mohou spadat například sylaby předmětů, interaktivní výukové materiály pro e-learning nebo texty vytvořené v rámci úvazku na přímý pokyn nadřízeného.

Naopak do této kategorie nepatří učebnice a monografie vytvořené díky individuální motivaci zaměstnance. Nadále ale platí, že tato otázka představuje pro vztah akademických institucí a jejich pracovníků velmi citlivou a komplikovanou záležitost vyžadující konstruktivní přístup obou zúčastněných stran při zohlednění akademických zvyklostí zachovávaných v jednotlivých vědních oborech.

Jak plyne z právě uvedeného, je u zaměstnaneckých děl jejich zařazení do databází šedé literatury záležitostí rozhodnutí příslušného zaměstnavatele. I ten je však v některých případech v nakládání s těmito literárními díly omezen, typicky pokud se k jejich tvorbě zavázal externímu poskytovateli finančních prostředků. K této situaci dochází tehdy, pokud je příslušné dílo, například vědecká studie, předmětem grantové smlouvy uzavřené mezi zaměstnavatelem (vysokou školou) a poskytovatelem dotace a ten si k výsledku vyhradil výkon majetkových práv autorských. Ve vztahu mezi vysokou školou a autorem tak jde o dílo zaměstnanecké, vysoká škola je však při dispozici s ním vázána ještě dohodou s poskytovatelem dotace zachycenou v dotační smlouvě.⁹⁶ Pokud v tomto případě užije například dílo bez souhlasu sám autor příkladně tím, že jej nahraje do databáze šedé literatury, odpovídá za zásah do majetkových práv autorských svému zaměstnavateli. Současně však zaměstnavatel odpovídá za nedovolené užití díla poskytovateli dotace, s nímž v dotační smlouvě dříve dohodl **exkluzivní licenci**. Z tohoto důvodu je v rámci projektu NUŠL budována partnerská síť institucí, které drží majetková autorská práva k zaměstnaneckým dílům.

Tato kapitola stručně nastínila právní otázky související s tvorbou a zpracováním šedé literatury. Jako klíčový právní problém bylo identifikováno především oprávnění k zařazení literárního díla do příslušné databáze, přičemž pořizovatel databáze často není s to ověřit, zda subjekt, který mu dokument do databáze poskytl, k němu skutečně disponuje příslušnými právy. Jedinou, byť nikoli ideální obranou provozovatele databáze šedé literatury před nároky třetích osob, je pak možnost následného regresu.⁹⁷

Pořizovatelům databází šedé literatury usnadňuje v řadě případů situaci skutečnost, že majetkovými právy autorskými k šedé literatuře často nedisponují jednotlivci ale instituce, konkrétně u školních děl příslušné školy a u děl zaměstnaneckých pak zaměstnavatelé jejich autorů. Pořizovatel databáze šedé literatury tak může v těchto případech kontrahovat užití autorských děl hromadně a relativní stabilita příslušných institucí, s nimiž uzavírá licenční smlouvy, mu dává i lepší pozici pro případné regresní nároky v situaci, kdy bude dílo zařazeno do databáze neoprávněně.

⁹⁶ Autorský zákon označuje tento režim výkonu majetkových práv autorských jako dílo vytvořené na objednávku a dává stranám možnost dohodnout vzájemný poměr práv k takovému dílu prakticky libovolně (§ 61 odst. 1 autorského zákona).

⁹⁷ V této situaci se nenacházejí ty databáze šedé literatury, které technicky fungují pouze jako uživatelská úložiště - provozovatelé těchto databází jsou kryti omezením odpovědnosti poskytovatelů služeb informační společnosti dle § 5 zákona č. 480/2004 Sb.

Tam, kde **databáze šedé literatury** umožňuje aktivní přístup jednotlivcům, lze z právního hlediska doporučit omezení jejích funkcionalit na prosté uživatelské úložiště dat. Provozovatel takové databáze tak sice nemá postavení pořizovatele ve smyslu § 88 a násl. autorského zákona a nemůže s databází ani s dokumenty tvořícími její obsah nijak nakládat (tj. ani je například nabídnout k využití partnerským databázím), je však díky omezení odpovědnosti poskytovatelů služeb informační společnosti v relativním právním bezpečí proti možnému postihu za nezaviněný zásah do autorských práv třetích osob.

Krátkou zmínku jsme v této kapitole věnovali i relativně novému fenoménu volných licencí, který s šedou literaturou velmi úzce souvisí. Jedná se o standardizované nabídky na uzavření bezúplatných licenčních smluv, které umožňují autorům praeter legem poskytnout jejich díla široké veřejnosti a využít k jejich masovému rozšíření mohutných přímých a nepřímých síťových efektů celosvětové informační sítě. Nelze v této souvislosti nevidět, že masová obliba volných licencí a mohutně rostoucí fenomén šedé literatury stále více tlačí na změnu základní filozofie a orientace současné rigidní a v mnoha směrech problematické autorskoprávní ochrany.

LITERATURA:

AUGER, Ch., P. *Information sources in grey literature, 2nd edition*. London: Bowker-Saur: New York, 1989.

GOKHALE, P. *Grey Literature Varieties - Definitional Problems, in Third International Conference on Grey Literature: Perspectives on the Design and Transfer of Scientific and Technical Information*. 13. -14. 11. 1997, Amsterdam: GreyNet, 1998, s. 259-273.

GUADAMUZ, A. *Open Science: Open Source Licences in Scientific Research*. North Carolina Journal of Law and Technology, roč. 7, č. 2, s. 321-366.

HAVRÁNEK, J. et al. *Teorie práva*. Plzeň: Aleš Čeněk, 2008, s. 392.

HOLLÄNDER, P. *Filosofie práva*. Plzeň: Aleš Čeněk, 2006, s. 74-93.

LESSIG, L. *Free Culture*. New York: The Penguin Press, 2004.

MARACKE, C. *Creative Commons International The International License Porting Project*. Journal of Intellectual Properte. Information Technology and E-Commerce Law, č. 1, s. 4-18.

PICKER, R. *Mediated Book*. John M. Olin Law and Economics working Paper No. 463. Chicago: The University of Chicago, 2009.

POLČÁK, R. *Právo na internetu - spam a odpovědnost ISP*. Brno: Computer Press, 2007, s. 68 a násl.

TŮMA, P. *Autorský zákon - Komentář*. Praha: C.H. Beck, 2007, s. 421.

TELEC, I. - TŮMA, P. *Autorský zákon - Komentář*. Praha: C.H. Beck, 2007, s. 380 a násl.

Zákon č. 111/1998 Sb., zákon o vysokých školách a o změně a doplnění některých dalších zákonů (o vysokých školách).

Zákon č. 121/200 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), jak vyplývá z pozdějších změn.

Zákon č. 480/2004 Sb., o některých službách informační společnosti a o změně některých zákonů (zákon o některých službách informační společnosti).

ELEKTRONICKÉ ZDROJE:

Creative commons Česká republika. [online]. c2010. [cit. 2010-06-10]. Dostupné z [www: <www.creativecommons.cz>](http://www.creativecommons.cz)

7. Popisné údaje pro šedou literaturu

Bohdana Frantíková

Formát pro uložení metadat je obecně nezbytnou součástí budování repozitářů. Je tedy neopomenutelný též při budování repozitářů šedé literatury. Metadata charakterizuje *Česká terminologická databáze knihovnictví a informační vědy* (dále jen TDKIV) takto:

„Strukturovaná data, která nesou informace o primárních datech. Pojem metadat je používán především v souvislosti s elektronickými zdroji a vztahuje se k datům v nejširším smyslu slova (datové soubory, textové informace, obrazové informace, hudba aj.). Funkce metadat je popisná, selekční a archivační. V souvislosti s těmito funkcemi se rozlišují metadata pro účely popisu, správy, právních nároků, technické funkčnosti, užití a archivace. Údaje se obvykle vkládají přímo do zdroje (umísťují se např. v záhlaví dokumentu HTML)“ (KTD, 2003).

Ačkoli existuje několik standardů metadatových schémat, lze konstatovat, že v prostředí šedé literatury existuje jejich velká variabilita. Tato různorodost nespočívá pouze v různých formátech, ale i ve způsobu jejich interpretace. Ta vychází z absence instrukcí a pravidel, jak konkrétní pole vyplňovat. Nelze samozřejmě tvrdit, že se jedná o popis zcela bez pravidel, ovšem zkoumaná metadatová schémata (viz níže) postrádají instrukce tak podrobné jako jsou např. manuály pro MARC21. V knihovnickém prostředí se řídí interpretace dat katalogizačními pravidly (např. v České republice Anglo-americká katalogizační pravidla). Díky tomu, navzdory všem kritickým připomínkám proti formátu MARC21 i AACR2, je pro toho, kdo oba uvedené standardy zná, metadatový formát přesným zachycením popisovaného objektu. Problémem analyzovaných metadatových formátů je právě absence směřování k místním pravidlům/dokumentaci, podle kterých byla data vytvářena. Jednou z cest, jak tento problém eliminovat, je způsob, kdy každý metadatový záznam obsahuje i link na přístupnou dokumentaci, která pomůže porozumět sémantice záznamu. Tato pravidla musí mimo trvalou dostupnost splňovat i podmínku možnosti aktualizace a doplňování o specifická užití. V červnu 2010 byla vydána katalogizační pravidla *Resource Description and Access* (dále jen RDA). Tato pravidla byla připomínkována i speciální pracovní skupinou *Dublin Core Metadata Initiative*.⁹⁸ Bude zajímavé, jak se tato pravidla promítnou v oblasti zpracování záznamů pro metadata.

Mimo níže uvedená schémata se lze v prostředí digitálních repozitářů setkat často i s *Metadata Object Description Schema* (dále jen MODS), *Metadata Encoding and Transmission Standard* (dále jen METS), *Digital Item Declaration*

⁹⁸ <http://dublincore.org/groups/libraries/rda/>

Language (dále jen DIDL) (Vyčítalová, 2009). Tyto formáty však nebyly pro tvorbu formátu NUŠL výchozími a z tohoto důvodu nejsou dále popisovány.

Základními požadavky na formát NUŠL jsou co největší jednoduchost formátu a jeho kompatibilita se standardem **Dublin Core**. Návrh formátu NUŠL byl tedy koncipován tak, aby oba požadavky splňoval.

Z formátů specializovaných na šedou literaturu byl hlavním vzorem formát **EVSKP-MS**, který byl navržen pro české prostředí vysokých škol a který ze standardu Dublin Core také vychází. Tento formát byl vytvořen pro eVŠKP, které jsou pouze jedním druhem šedé literatury. Bylo nutné rozšířit tento formát o další specifické prvky, aby bylo možné kvalitně popisovat i ostatní druhy šedé literatury. Vzhledem k významnosti formátu **EVSKP-MS** je jeho stručný popis součástí tohoto dokumentu. Vzory pro další specifické prvky popisující šedou literaturu byly báze **ASEP** a lokální báze s metadatovým formátem pro **Rejstřík informací o výsledcích** (dále jen **RIV**).⁹⁹

Do formátu NUŠL se promítají prvky z mezinárodních formátů Dublin Core a Dublin Core Terms, ETD-MS a MARC21, které jsou popsány níže. Záznamy z báze NUŠL budou dodávány do báze OpenSIGLE, a proto je uvedena i její krátká charakteristika.

Výchozí formáty pro NUŠL

Dublin Core

Dublin Core je nejznámější mezinárodně uznávaný metadatový formát především pro popis elektronických zdrojů. Základy tohoto formátu byly položeny na semináři, který v roce 1995 pořádala organizace v Dublinu v Ohio *Online Computer Library Center* (dále jen **OCLC**).

Formát Dublin Core původně vznikl jako reakce na potřebu zlepšení vyhledávání informačních zdrojů v prostředí internetu. Klasické knihovnické popisy typu MARC byly pro toto prostředí složité jednak z hlediska předpokládaných budoucích tvůrců záznamů, tak i jejich struktura a sémantika přestala vyhovovat novému prostředí. „Formáty typu MARC jsou jenom ‚strojem čitelné‘, nikoliv však stroji srozumitelné, tj. jejich struktura i sémantika nevyhovují požadavkům na budoucí architekturu komunikace informací včetně informací o informacích (metadat) na www“ (Bratková, 1999).

Dublin Core je soubor metadatových prvků. Původní formát obsahoval třináct prvků, ale byl posléze doplněn a dnes obsahuje 15 prvků. Později byl Dublin Core rozšířen a obohacen o další prvky a vznikl tak formát Dublin Core Terms.

⁹⁹ <http://www.vyzkum.cz/FrontClanek.aspx?idsekce=956>

Současná verze formátu „DCMI Metadata Terms“ odráží změny schválené a platné od prosince 2006.¹⁰⁰

Dublin Core nepředepisuje žádnou závaznou syntaxi (jedná se o sémantický standard).

Formát Dublin Core je rozšiřitelný v konkrétních lokálních aplikacích, všechny prvky jsou volitelné a opakovatelné, vybrané údaje lze modifikovat, zejména pro potřeby speciálních systémů přes tzv. kvalifikátory (schémata/modely hodnot údajů a typy údajů) (Bratková, 1999).

V roce 1998 byla metadata **Dublin Core** uznána internetovou normou IETF RFC 5013. V září 2001 ratifikoval *American National Standards Institute* (dále jen ANSI) verzi 1.1 jako národní americký standard ANSI Z39.85-2007. V roce 2003 byl soubor Dublin Core schválen *International Standard Organization* (dále jen ISO) jako mezinárodní norma ISO 15836:2003 (Synková, 2008).

Formát Dublin Core byl přeložen cca do 25 jazyků a je použit v řadě projektů a systémů po celém světě. Některé projekty ho převzaly v jeho původní podobě, další ho podobně jako projekt NUŠL využily jako základ pro vytvoření vlastního formátu.

Na rok 2010 připadá 15. výroční založení Iniciativy Dublin Core, kdy proběhne již 10. ročník mezinárodní konference.¹⁰¹ Programem konference je reflexe minulosti i pohled do budoucnosti: konceptuální modely a rámce (např. RDF, DCAM, OAIS), aplikační profily, domény metadat (např. pro podniky, kulturní instituce, školství, státní správu aj.), bibliografické standardy (RDA, FRBR, předmětová hesla), dostupnost metadat a také metadata pro vědecká data, vyhledávače a metadata, znalostní systémy a *Simple Knowledge Organization System* (dále jen SKOS) aj.

OpenSIGLE

Na vývoj systému SIGLE, který je popsán ve druhé kapitole, se mezi lety 1980 až 2005 podílelo šestnáct zemí. V roce 2005 obsahovala databáze 855 260 záznamů výzkumných zpráv, VŠKP, konferenčních sborníků, datových souborů a překladů z různých vědeckých oblastí.

Organizace INIST, francouzský člen EAGLE, vyvinula na základě softwaru **MIT Dspace** repozitář OpenSIGLE a z databáze SIGLE do ní uložila ve zjednodušeném formátu XML záznamy bývalých členů EAGLE, kteří vyjádřili souhlas. Od konce roku 2008 jsou do repozitáře ukládány plné texty šedé literatury.

V repozitáři OpenSIGLE je nyní otevřený přístup k více než 200 konferenčních příspěvků z mezinárodní konference o šedé literatuře (GL1 - GL10).¹⁰²

¹⁰⁰ Více o těchto změnách naleznete na <http://dublincore.org/usage/decisions/2008/dcterms-changes/#sect-1>.

¹⁰¹ <http://www.asis.org/Conferences/DC2010/>

¹⁰² Tyto kolekce jsou dostupné na <http://opensigle.inist.fr/handle/10068/697753>.

Další informace o OpenSIGLE jsou uvedeny v kapitole Vývoj šedé literatury v České republice.

ETD-MS

Metadatový formát ETD-MS definuje soubor prvků pro popis elektronických disertačních prací. Tento formát není určen k tomu, aby nahradil stávající lokální metadatové formáty vyvinuté pro jednotlivé instituce, ale aby byl používán pro sdílení dat mezi jednotlivými systémy. Základem tohoto formátu jsou prvky Dublin Core, doplněné specifické prvky pro eVŠKP, které jsme využili v našem formátu. Formát ETD-MS je zde zmíněn, protože byl jedním z výchozích formátů při tvorbě českého formátu EVSKP-MS, ze kterého formát NUŠL vychází.

MARC 21

Formát MARC21 je spravován *Library of Congress* ve spolupráci s různými skupinami uživatelů. Je standardem pro popis a reprezentaci bibliografických informací ve strojově čitelné podobě, specifikuje pole údajů, návěští polí (tagy), indikátory a kódy podpolí. Stanovuje povinné údaje pro výměnu, pořadí údajů, jejich odpovídající kódová označení a pravidla použití.

MARC 21 byl navržen tak, aby se usnadnila výměna bibliografických a dalších souvisejících informací mezi institucemi.

Jeho krátkou charakteristiku zde uvádíme proto, že je nativním formátem systému CDS Invenio, který je využíván v NTK jako digitální repozitář. Proto bylo nutné vytvořit mezi formátem NUŠL a MARC21 konverzní tabulku, která je v příloze *Konverzní tabulka mezi formátem NUŠL a MARC 21* uvedená v závěru této kapitoly.

Metadatový standard eVŠKP

Formát EVSKP-MS je určen pro popis eVŠKP. Byl vytvořen a spravován Komisí eVŠKP obdobně jako metadatové formáty PersCz pro popis fyzických osob a CorpCz pro popis korporací. Tyto formáty byly využity jako dílčí v rámci EVSKP-MS, slouží pro přesnější zápis např. jmen autorů či detailnější identifikaci instituce. Při ukládání metadat je nutno mít na paměti, že tyto údaje se mohou v průběhu let měnit, což lze demonstrovat např. na titulu autora bakalářské práce, který odpovídající titul Bc. získává až po obhajobě práce.

Formát EVSKP-MS vychází z formátu Dublin Core. Na základě změn, které přinesla verze definic termínů *Dublin Core Metadata Initiative* (dále jen DCMI) publikovaná v lednu 2008, byl upraven i formát EVSKP-MS. Kromě toho obsahuje formát také vybrané prvky z nejvýznamnějšího mezinárodního standardu ETD-MS, který využívá mezinárodní databáze NDLTD. Při tvorbě formátu byl

také brán zřetel na německý národní metadatový standard XmetaDiss a francouzskou normu pro metadata pro elektronické disertace TEF 2.0 (EVSKP-MS, 2008).¹⁰³

ASEP

Systém ASEP obsahuje bibliografické záznamy o výsledcích vědeckého výzkumu v ústavech AV ČR od roku 1985, v úplnosti od roku 1993. Záznamy jsou vytvářeny on-line prostřednictvím webových formulářů.

Záznamy ze systému ASEP se používají také pro zasílání do databáze RIV. ASEP má však širší rozsah. Protože v pokynech pro předávání dat do RIV dochází každoročně k úpravám, je také v systému ASEP pozměňována struktura záznamů. Změny se týkají především přidávání nových polí a definování jejich obsahu, případně zařazování nových typů dokumentů.

Formát ASEP je formátem typu MARC s doplněnými vlastními poli. Data jsou uchovávána v systému ARL. Informace k databázi ASEP včetně popisu formátu jsou dostupné na stránkách AV ČR.¹⁰⁴

Formát NUŠL

Metadatový formát NUŠL byl speciálně vytvořený pro zpracování záznamů o digitálních dokumentech šedé literatury.

Metadatový formát NUŠL používá prvky z formátů Dublin Core, Dublin Core Terms, EVSKP-MS, EDT-MS a vlastní prvky.

Formát NUŠL je otevřený formát, který může být v budoucnu doplňován o elementy, které budou nutné pro provoz digitální knihovny šedé literatury a spolupráci jak s lokálními dodavateli dat, tak s mezinárodními registry.

Ke změnám a zjednodušení původní koncepce metadatového formátu NUŠL došlo již po výběru softwaru digitálního repozitáře NUŠL, kterým se stal systém **CDS Invenio** (více informací o tomto systému naleznete v deváté kapitole).

Předpokládá se, že pokud pilotní projekt nepotvrdí potřebnost některého elementu, bude tento element vyřazen.

Formát NUŠL se skládá z povinných a nepovinných prvků. V této kapitole jsou podrobně popsány povinné prvky metadatového formátu NUŠL a specifické prvky pro šedou literaturu. S celým formátem NUŠL, tedy i s nepovinnými prvky, se můžete seznámit na http://nusl.techlib.cz/images/Metada_NUSL_final_aktualizace2.pdf.

¹⁰³ Popis formátu je na stránkách Komise eVŠKP na adrese <http://www.evskp.cz/standardy/evskp/>.

¹⁰⁴ <http://www.lib.cas.cz/cs/ASEP>

Povinné prvky

Metadatový formát obsahuje kromě atributů pro jednotlivé prvky také obecný atribut `xml:lang=[kód jazyka]` společný pro většinu prvků. Obsahem tohoto atributu je vždy kód jazyka vztahující se k danému prvku, který nemá nic společného se samostatným prvkem `dc:language`.

nusl:metadata

Nově vytvořený prvek pro formát NUŠL.

Kořenový element metadatového záznamu o zdroji. Všechny ostatní elementy jsou vnořené v tomto elementu.

dc:identifier

Tento prvek je do formátu NUŠL převzatý z Dublin Core (`dc:identifier`).

Jednoznačný a trvalý identifikátor zdroje.

Původním záměrem bylo, že v tomto poli bude uveden trvalý identifikátor typu **URN:NBN**, **Handle** apod. V současné době však v České republice nefunguje resolver URN:NBN, který by zprostředkoval a zajišťoval trvalé propojení mezi identifikátorem a odpovídajícím URL odkazem. Identifikátor na záznam URI bude tedy generován přímo v systému CDS Invenio ve formátu: **www.nusl.cz/ntk/nusl-ID**.

Identifikátor reprezentuje pořadové číslo přidělené systémem. Pokud bude v budoucnosti vyřešena problematika přidělování trvalých identifikátorů URN:NBN v rámci České republiky, budou tyto identifikátory do záznamů doplněny dodatečně.

Také harvestovaným záznamům bude systémem CDS Invenio přidělen identifikátor stejného formátu. V opakovaném výskytu pole `dc:identifier` budou uvedeny identifikační čísla přidělená lokálními repozitáři, popř. systémové číslo.

dc:title

Tento prvek je do formátu NUŠL převzatý z Dublin Core.

Hlavní název zdroje, který zdroji přidělil tvůrce.

dc:creator

Tento prvek je do formátu NUŠL převzatý z Dublin Core.

Fyzická osoba, která je primárně odpovědná za vytvoření obsahu zdroje, tj. autor VŠKP, hlavní řešitel projektu, vedoucí výzkumného záměru apod., příp. korporace, která nese primární autorskou odpovědnost za dílo.

dcterms:dateCreated

Tento prvek je do formátu NUŠL převzatý z Dublin Core Terms. Datum zveřejnění zdroje.

dc:subject

Tento prvek je do formátu NUŠL převzatý z Dublin Core.

Obsahem tohoto prvku je věcný popis šedé literatury vyjádřený buď ve formě termínů z *Polytematického strukturovaného hesláře* (dále jen **PSH**), nebo ve formě volně tvořených klíčových slov, a to v českém nebo anglickém jazyce. Podrobněji je věcné třídění v rámci digitálního repozitáře NUŠL popsáno na konci této kapitoly.

dc:type

Tento prvek je do formátu NUŠL převzatý z Dublin Core.

Prvek je v rámci NUŠL formátu opakovatelný, v 1. výskytu je obsahem typologie dokumentu NUŠL, ve 2. výskytu typu dokumentu podle DCMI.¹⁰⁵

Obsahem 1. výskytu - typ dokumentu NUŠL jsou typy dokumentů podle typologie dokumentů NUŠL, která je popsána ve třetí kapitole. Zde se nikdy nesmí vyskytovat slovní řetězec (pojmenování typu dokumentu), který není uveden v typologii dokumentů NUŠL.

Obsahem 2. výskytu - typ dokumentu podle typologie DCMI. Zápis tohoto prvku je pro formát NUŠL nepovinný a opakovatelný. V souvislosti s funkcemi systému CDS Invenio je ideální, pokud spolupracující repozitáře respektují typologii dokumentu NUŠL. U starších záznamů a při přebírání záznamů z jiných databází, které nemohou tomuto požadavku vyhovět, se vytvoří harmonizační tabulky mezi typologiemi.

dc:language

Tento prvek je do formátu NUŠL převzatý z Dublin Core.

Jazyk intelektuálního obsahu zdroje.

dc:rights

Tento prvek je do formátu NUŠL převzatý z Dublin Core.

Textová informace o právech k využívání. Specifikují se podmínky, na základě kterých může být práce distribuována, reprodukována atd.

V rámci formátu NUŠL se využívá opakovatelnost tohoto pole. V 1. povinném výskytu tohoto pole je vždy uvedeno, že plný text dokumentu je chráněn autorským právem (Polčák, 2009). Ve 2. výskytu tohoto pole je možné uvést některou z licencí Creative Commons, podle které lze s dokumentem zacházet. V případě využití licence Creative Commons se pole Omezení přístupu, resp. `nusl:transfer (accessRights)`, automaticky vyplní obsahem, že dokument je public (tj. veřejný). Pokud pole `dc:rights` nebude vyplněno, je nutné vyplnit následující pole o Omezení přístupu a v něm zvolit některou z možností zpřístupnění dokumentu.

¹⁰⁵ <http://dublincore.org/documents/1998/10/23/type-element>

nusl:transfer (accessRights)

Nově vytvořený prvek pro NUŠL.

V tomto poli se zapisuje tzv. omezení přístupu ke zdroji, které popisuje způsob zveřejnění dokumentu. Pole je potřeba vyplnit obsahem Public, tj. soubor je veřejně přístupný přes síť internet online bez nutnosti registrace. Způsob vyplnění tohoto pole vychází z pravidel, podle kterých vyplývá, že ostatní instituce mohou být dokumenty bezplatně archivovány pouze v případě, že jsou dokumenty veřejně přístupné. Další zvažované možnosti zpřístupnění digitálních dokumentů, např. domain, restrictedAccess, eduoram aj., nejsou v tuto chvíli předmětné.

Specifické prvky pro šedou literaturu*Elementy pro zpracování zpráv*

Elementy týkající se zpracování zpráv jsou pro NUŠL nově vytvořené a inspirací pro jejich vznik byl především systém ASEP. Prvky byly do formátu doplněny nově proto, že v existujících metadatových formátech nebyly nalezeny takové, které by vyhovovaly specifickým požadavkům na zpracování dokumentů typu zprávy.

nusl:resultID

Nově vytvořený prvek pro NUŠL.

Obsahem opakovatelného pole je označení grantu (grantů), v jehož rámci byla vypracována hlášená publikace. Obsahem pole může být také označení výzkumného záměru pracoviště, v jehož rámci byla vypracována hlášená publikace.

nusl:resultGrantGarant

Nově vytvořený prvek pro NUŠL.

Obsahem prvku je název poskytovatele projektu.

Elementy pro zpracování konferenčních materiálů

Oblast popisu konferenčních materiálů nebyla v žádném z formátů, které autoři NUŠL považovali za zdrojové nebo výchozí pro tvorbu formátu NUŠL, podrobněji zpracována. Rozhodli proto pro vytvoření vlastních prvků. V prvním návrhu formátu byl tento prvek navržen jako jeden hlavní s prvky vnořenými.

Vzhledem k systému, který byl pro digitální repozitář vybrán, a tomu, že jeho vnitřním formátem je MARC 21, byla vytvořena konverzní tabulka mezi formátem MARC 21 a formátem NUŠL (viz příloha *Konverzní tabulka*). Při vytváření konverzní tabulky byl tento prvek zjednodušen a vznikly tak následující tři prvky: nusl:event, nusl:eventPlace, nusl:eventDates (podrobněji

jsou tyto prvky rozepsané níže). Prvky, tak jak jsou nyní používány, jsou také vhodné pro spolupráci se systémem OpenSIGLE, protože v podstatě odpovídají tomu, jak jsou v tomto systému zpracovávány konferenční materiály. S tímto systémem byla rovněž sladěna pravidla pro vyplnění jednotlivých polí.

nusl:event

Nový element vytvořený pro formát NUŠL.

Název konference nebo jiné akce. Oficiální název konference nebo akce se zapisuje v originálním jazyce v nezkrácené podobě.

nusl:eventPlace

Nový element vytvořený pro formát NUŠL.

Stát (místo) konání konference nebo jiné akce. Zapisuje se ve formě město, v závorkách uvedený stát konání konference.

nusl:eventDates

Nový element vytvořený pro formát NUŠL.

Datum nebo rozmezí dat konání konference nebo jiné akce, zápis dle normy ISO 8101.

Elementy pro zpracování eVŠKP

Následující skupina prvků pro eVŠKP a studijní materiály byla převzata z formátu EVŠKP-MS, resp. ETD-MS: an Interoperability Metadata Standard for Electronic Theses and Dissertations. Prvky se týkají údajů o akademickém titulu nebo vědecko-pedagogické hodnosti a instituci, která ho přiděluje, a jsou vnořené do prvku thesis:degree.

thesis:name

Tento prvek je do formátu NUŠL převzatý z formátu EVŠKP-MS.

Zkratka jména akademického titulu nebo vědecko-pedagogické hodnosti přidělované vysokou školou (například Ph.D., doc., JUDr., Mgr., Ing. aj.).

Údaje o akademickém titulu nebo vědecko-pedagogické hodnosti získané v souvislosti se zpracovávaným dokumentem.

thesis:level

Tento prvek je do formátu NUŠL převzatý z formátu EVŠKP-MS.

Typ studijního programu, v jehož rámci byla připravena VŠKP (doktorský, magisterský, bakalářský aj.).

thesis:discipline

Tento prvek je do formátu NUŠL převzatý z formátu EVŠKP-MS.

Studijní program a studijní obor akreditovaný příslušnou komisí MŠMT, v jejichž rámci byla připravena a obhájena VŠKP.

thesis:grantor

Tento prvek je do formátu NUŠL převzatý z formátu EVŠKP-MS.

Instituce (škola, ústav apod.) garantující přidělení titulu spojené s vytvořením VŠKP.

Technická a administrativní metadata

Kromě bibliografických metadat systém CDS Invenio archivuje další technické a administrativní metadata, která nejsou přímo součástí metadatového formátu, ale slouží jako parametry různých funkcionalit a k uchování administrativních dat. Tyto administrativně/technické údaje jsou obsaženy v různých souborech, které nejsou podstatné pro samotný popis digitálního dokumentu. Ovšem jsou důležité pro administrátory repozitářů, proto jsou zde uvedena.

Následuje výčet (relevantních) technických a administrativních dat:

soubor PID

Identifikátor je trvalým (perzistentním) jménem daného záznamu se syntaxí: nusl-ID, kde ID = ID záznamu.

soubor SN

ID je identifikační číslo přidělené systémem CDS Invenio, odkazuje k pořadí vložení daného záznamu.

soubor doctype

Název souboru Typ dokumentu je v našem případě shodný s názvem sbírky. Typ dokumentu definuje sbírku, pod kterou je daný záznam vložen. Typ je definován v rámci typologie NUŠL.

soubor comboPODSBIRKA

Název podsbírký odpovídá typu dokumentu daného typologií NUŠL, kam byl daný záznam vložen.

soubor FILE_INPUT

Původní název digitálního dokumentu, který je přiložen k danému metadatovému záznamu.

soubor FILE_INPUT_RENAMED

Systémem nově přidělený název digitálního dokumentu, který nahrazuje název původní.

soubor indir

Název adresáře, ve kterém je uložen celý záznam spolu se všemi přiloženými digitálními dokumenty včetně těchto technických informací a dat.

Metadatový záznam zapsaný v XML

Metadatová pole jsou také kromě uložení v databázi (v našem případě MySQL - soubor recmysql) ukládána jako soubor v XML zápisu.

soubor SuE

Záznamy mohou vkládat knihovníci z různých institucí, součástí administrativních metadat je také email vkladatele, který daného knihovníka identifikuje.

Věcné třídění

Při řešení problematiky věcného třídění autoři formátu NUŠL zvažovali tyto možnosti: použití věcného třídění systému OpenSIGLE, hesláře PSH a volně tvořených klíčových slov.

Věcné třídění OpenSIGLE

Věcné třídění OpenSIGLE odpovídá *SIGLE Subject Category List* (SIGLE Manual, 1999), který byl naposledy aktualizován v roce 1999. Na základě pracovní analýzy systému bylo zjištěno, že věcné třídění OpenSIGLE je pouze v textové podobě, některé kategorie by bylo potřeba aktualizovat, ale v blízkém období nelze očekávat žádné rozšiřování tohoto třídění. Třídění OpenSIGLE je rozděleno do 20 základních kategorií, které obsahují další podhesla. Jejich počty se u jednotlivých kategorií liší, nejmenší počet hesel jsou 3 (kategorie 120 Mathematical sciences, general), nejvyšší počet 26 (pro kategorii 060 Biological and medical sciences, general).

PSH

Polytematický strukturovaný heslář je česko-anglický řízený slovník rozdělený do 44 tématických skupin, ve kterých jsou hesla hierarchicky členěna. Je určen k předmětové indexaci dokumentů a jejich následnému vyhledávání. PSH je standardně zveřejněn ve formátu SKOS, který umožňuje přístup k celému hesláři ve strojově čitelné podobě.

Pro využití slovníku PSH v rámci formátu NUŠL se sahá z těchto důvodů: je vyvíjen v rámci NTK, je dvojjazyčný a je přizpůsobován pro věcný popis elektronických informačních zdrojů.

Slovník PSH se v rámci formátu NUŠL používá do druhé úrovně hesel, implementace celého slovníku PSH do systému CDS Invenio se připravuje.

Je plánována automatická indexace plných textů hesel ze slovníku PSH pomocí Maui Indexer. „Přiřazování hesel z řízeného slovníku funguje tak, že je nejprve vytvořen model, který na vstupu vyžaduje plné texty intelektuálně opatřené termíny z použitého slovníku, na jejichž základě se Maui Indexer naučí

správnému indexování. Pokud je tento model k dispozici, lze začít zpracovávat nové dokumenty, které nejsou indexovány. Maui Indexer jim podle analýzy jejich obsahu a vztahů a pravděpodobností zachycených v modelu přiřadí co možná nejvhodnější indexační výrazy pocházející z daného slovníku“ (Mynarz, 2009).

Mapování PSH na třídění OpenSIGLE je prováděno mimo repozitář při předávání exportů do této databáze.

Volně tvořená klíčová slova

Vzhledem k tomu, že slovník PSH je implementován pouze do druhé úrovně hesel a že přebírané záznamy obsahují různé druhy popisu, je zde ponechána též varianta vložení volně tvořených klíčových slov nebo věcných termínů z jiných systémů.

LITERATURA:

BRATKOVÁ, Eva. Metadata jako nový nástroj pro komunikaci webovských informačních zdrojů. *Národní knihovna Knihovnická revue* [online]. 1999, roč. 10, č. 4 [cit. 2010-04-19] Dostupné z www: <<http://full.nkp.cz/nkkr/pdf/9904/9904178.pdf>. ISSN 1214-0678.

KRČÁL, Martin. o Komisi: základní informace [online]. Asociace knihoven vysokých škol ČR, Odborná komise e-VŠKP. Aktualizováno 14. 4. 2008 [cit. 2008-05-27]. Dostupné z www: <<http://www.evskp.cz/komise.php?tsekce=6&sek=>>.

MYNARZ, Jindřich. Jak lze prakticky využít Polytematický strukturovaný heslář pro věcný popis elektronických zdrojů. *Ikaros* [online]. 2009, roč. 13, č. 12. [cit. 2010-08-06]. Dostupné z www: <<http://www.ikaros.cz/node/5872>>. URN-NBN:cz-ik5872. ISSN 1212-5075.

POLČÁK, Radim. Digitální zpracování tzv. šedé literatury pro Národní úložiště šedé literatury [online]. 2009. [cit. 2010-03-10]. Dostupné z www: <http://nysl.techlib.cz/index.php/Právní_expertiza>.

SCHÖPFEL, Joachim - STOCK, Christiane - HENROT, Nathalie. From SIGLE to OpenSIGLE: An in-depth look on resource migration in the European context. In *Eighth International Conference on Grey Literature : Harnessing the Power of Grey, New Orleans, United States, 4-5 December 2006*. (Conference Program and Abstracts: GL-conference series, ISSN 1385-2308; No. 8). Preprint konferenčního příspěvku. Dostupné z www: <<http://opensigle.inist.fr/bitstream/10068/697774/2/GL8,%20Schopf%20et%20al.,%202007,%20Conference%20Preprint.pdf>>.

SIGLE Manual. Part 2. System for Information on Grey Literature in Europe. European Association for Grey Literature Exploitation. EAGLE Technical Committee: April 1999.

SYNKOVÁ, Veronika. Vybrané aplikace metadatového formátu Dublin Core. *Ikaros* [online]. 2008, roč. 12, č. 5. [cit. 2010-04-19]. Dostupné z [www: <http://www.ikaros.cz/node/4715>](http://www.ikaros.cz/node/4715). URN-NBN:cz-ik4715. ISSN 1212-5075.

VYČÍTALOVÁ, Hana. Zahraniční digitální archivy vědeckovýzkumných a technických zpráv a perspektivy jejich integrace [Foreign digital archives of research and technical reports and perspectives of their integration]. Praha, 2009. 112 s., V s. příl. Diplomová práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví 2009. Vedoucí diplomové práce PhDr. Eva Bratková PhD.

Zákon č. 130/2002 Sb. o podpoře výzkumu a vývoje z veřejných prostředků a o změně některých souvisejících zákonů (zákon o podpoře výzkumu a vývoje).

ELEKTRONICKÉ ZDROJE:

Dublin Core Czech [online]. Masarykova univerzita v Brně. Aktualizováno 20. 11. 2006. [cit. 2010-06-10]. Dostupné z [www: <http://www.ics.muni.cz/dublin_core/index.html>](http://www.ics.muni.cz/dublin_core/index.html).

Dublin Core Metadata Initiative. 2008. *DCMI Metadata Terms* [online]. Dublin Core Metadata Initiative, [cit. 2008-05-27].] Dostupné z [www: <http://dublincore.org/documents/dcml-terms/>](http://dublincore.org/documents/dcml-terms/).

EVSKP-MS: metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR [online]. Zpracovatelé Eva Bratková, Jan Mach. Verze 1.1. Praha: Odborná komise pro otázky elektronického zpřístupňování VŠKP AKVŠ ČR, 15. 7. 2008 [cit. 2008-09-05]. Dostupné z [www: <http://www.evskp.cz/standardy/evskp/1.0/>](http://www.evskp.cz/standardy/evskp/1.0/).

KTD : Česká terminologická databáze knihovnictví a informační vědy (TDKIV) [online]. Praha: Národní knihovna České republiky, 2003. Dostupné z [www: <http://sigma.nkp.cz/cze/ktl>](http://sigma.nkp.cz/cze/ktl)

Knihovna AV ČR. 2007. Evidence publikační činnosti v AV ČR: Manuál [online]. Praha: Knihovna AV ČR, [c2007]. [cit. 2008-05-23]. Dostupné z [www: <http://www.iach.cz/knav/manual/manual.htm>](http://www.iach.cz/knav/manual/manual.htm).

Metadatový formát NUŠL. Zpracovatel Veronika Synková, Radka Římanová a Bohdana Frantíková. Verze Beta 0.3. Praha: Národní technická knihovna, 2009.

RIV. Rejstřík informací o výsledcích [online]. Rada pro výzkum, vývoj a inovace. Aktualizace 10. 5. 2010. c2010. [cit. 2010-08-27]. Dostupné z [www: <http://www.vyzkum.cz/FrontClanek.aspx?idsekce=956>](http://www.vyzkum.cz/FrontClanek.aspx?idsekce=956)

Příloha Konverzní tabulka mezi formátem NUŠL a MARC 21

V tomto dokumentu je zpracována konverzní tabulka mezi prvky formátu NUŠL a formátem MARC 21. Potřeba zpracování konverzní tabulky v rámci tohoto projektu vznikla v souvislosti s výběrem softwaru pro NUŠL, kterým je systém CDS Invenio, jehož nativním formátem je právě MARC 21.

Východiskem při zpracování konverzní tabulky byla konverze *Library of Congress* (Network Development and MARC Standards Office) Dublin Core do MARC 21, která je vystavena na <http://www.loc.gov/marc/dccross.html>. Prvky, které jsou do formátu NUŠL přebrány z metadatového standardu ETD-MS, jsme konvertovali tak, jak je uvedeno na stránkách NDLTD, <http://www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.html#thesis.degree>, pokud to umožňoval systém CDS Invenio.

Pro prvky, které nejsou z Dublin Core, jsme se snažili sami najít v MARC odpovídající pole, pokud to bylo možné.

NUŠL prvek	MARC 21- identifikátor, pole, podpole	Poznámka
nusl:metadata (Metadatový záznam NUŠL)		nepřiřazuje se
dc:identifier (identifikátor zdroje)	856 40\$u	
dc:title	245 00\$a	
dc:title (nusl:typeTranslated)	245 00\$b	
dcterms:alternative	246 33\$a	
dcterms:alternative (nusl:typeTranslated)	246 33\$b	
dcterms:alternative (nusl:titleType=volTitle)	246 30\$p	
dcterms:alternative (nusl:titleType=volNumber)	246 3#\$n	
dc:creator	720 1#\$a (osoba) 720 2#\$a (korporace) 720 ##\$a (nespecifikován)	Konvertujeme do pole 720, protože nevytváříme hlavní ani vedlejší záhlaví pro korporace a osoby.
dc:contributor	720 1#\$i (osoba) 720 2#\$i (korporace) 720 ##\$i (nespecifikován)	Konvertujeme do pole 720, protože nevytváříme hlavní ani vedlejší záhlaví pro korporace a osoby.
dc:contributor.role	720 \$e	
nusl:event	711 2#\$a	
nusl:eventPlace	711 2#\$c	
nusl:eventDates	711 2#\$d	
thesis.degree.name	502 \$b	

thesis.degree.level	502 \$g	
thesis.degree.discipline	656 #7\$a	
	656 \$2	(podle číselníku AKVO)
thesis.degree.grantor	502 \$c	
nusl:resultID	C12\$a číslo projektu; C13\$a výzkumný záměr (kód záměru)	
nusl:resultGrantGarant	C12\$b poskytovatel CEP; C12\$c poskytovatel ostatní	
dc:publisher	260 ##\$b	
dcterms:dateCreated	046 ##\$k	Special Coded Dates/Date Created
dcterms:dateSubmitted	502 ##\$a	(Dissertation Note) with initial label „Date Submitted“
dcterms:dateAccepted	502 ##\$a	(Dissertation Note) with initial label „DateAccepted“
dcterms:modified	046 ##\$j	Special Coded Dates/Date Modified
dc:subject	650 #7\$a	
	650 \$2	PSH
dcterms:abstract	520 ##\$a (Summary etc. note)	
dc:type (nusl:typeType)	655 #7\$a	
	655 \$2	NUŠL typ dokumentu
dc:type (DCMIType)	655 #7\$a	
	655 \$2	DCMI type
dcterms:medium	340 ##\$a (Physical Medium)	
dc:language	041 07\$a	
dcterms:BibliograficCitation	500 ##\$a	
dc:rights	540 ##\$a (Terms Governing Use and Reproduction Note)	
dcterms:extent	300 ##\$a (Physical Description)	
dcterms:available	307 ##\$a (Hours, Etc.)	
nusl:dateDelivered	008 _00-05	
nusl:modified	005	
nusl:transfer (accessRights)	506##\$a (Restrictions on Access Note)	

8. Analýza a podklady pro výběr softwarového řešení pro NUŠL

Petr Karlach

Jedním z důležitých kroků při realizaci projektu NUŠL byl výběr softwarového řešení pro ukládání obsahu, jeho správu a zpřístupňování v rámci úložiště šedé literatury. Zadání výběru nicméně nezahrnovalo jen úložiště pro šedou literaturu, cílový software byl vybírán s tím, aby jej bylo možno použít i pro spolupracující **digitální úložiště** v rámci NUŠL.

Výběr softwarového řešení pro digitální knihovnu se v zásadě neliší od výběru softwaru pro jakoukoliv jinou činnost. Jakkoliv se může z digitální povahy úložiště zdát, že softwarový systém je hlavní součástí plánovaného úložiště, je to nakonec pouze nástroj, který umožní úložiště realizovat. Nejdůležitější přípravnou fází pro **výběr softwaru** je tedy dobrý popis samotného úložiště, jeho cílů, hlavních funkcí, procesů, činností, obsahu a vazeb na okolí. Z dobré definice úložiště pak lze snadno odvodit hlavní požadavky na softwarový systém a kritéria jeho výběru. Určitým specifickým rysem v oblasti softwaru pro digitální knihovny, jež se provozují převážně v neziskovém sektoru, je například v porovnání s komerční sférou dostupnost programového vybavení z oblasti **Open Source** (software s otevřeným zdrojovým kódem, dostupný zpravidla bez licenčních poplatků), jehož několik nejznámějších představitelů má za sebou úspěšné instalace, z nichž se lze poučit.

Tato kapitola je rozdělena do dvou částí. První část se věnuje definování požadavků na funkcionality úložiště šedé literatury, které byly podkladem pro výběrové řízení na softwarové řešení pro NUŠL, a v druhé části jsou popsány a analyzovány vybrané Open Source systémy podle stanovených kritérií. Zpracovaná analýza je vystavena na webových stránkách projektu NUŠL a byla inspirací pro potenciální dodavatele v rámci výběrového řízení na softwarové řešení pro NUŠL. Výsledek výběrového řízení je popsán v úvodu následující kapitoly.

Pokud je hotová kvalitní definice samotného úložiště, je možné z ní odvodit kritéria výběru. V případě podkladů pro výběr softwarového řešení pro NUŠL byla **kritéria** stanovena iterativně v několika krocích a ve fázi jejich prvotního návrhu rozdělena pro přehlednost do pěti skupin, do nichž byly bez nároku na naprosto přesné rozdělení a relevanci seskupeny hlavní požadavky:

Řízení obsahu digitální knihovny - toto kritérium se vztahuje k hlavním funkcím digitální knihovny, zahrnuje tvorbu, ukládání, kontrolu a verzování obsahu (např. při migraci souborů) digitální knihovny, jeho prohlížení lineární (rejstříky) i hierarchické, správu metadat, různé mechanismy pro získávání dat (např. harvesting - **OAI-PMH**, web services), interoperabilitu pro spolupráci s jinými digitálními knihovnami (např. Z39.50, OAI-PMH, SRU), podporu pro

velké objemy dat, vícejazyčné vyhledávání nad metadaty a soubory (fulltextové), logické a proximitní vyhledávání, podporu různých formátů ukládání souborů - textových (např. DOC, RTF, PDF), obrázkových (např. TIFF, JPEG, GIF, PNG), prezentačních (např. MS PPT, Adobe PDF a PostScript), strukturovaných (př. HTML a XML), audio a video (např. WAV, MP3, AVI, MPEG4, Real audio a video), podpora kódování (př. ASCII, UNICODE, UTF-8).

Uživatelské rozhraní - přístup přes www s podporou hlavních internetových prohlížečů, možnost přizpůsobení uživatelského rozhraní pro různé role uživatelů zvenku i zevnitř knihovny, vícejazyčné uživatelské rozhraní, vícejazyčná prezentace výsledků vyhledávání, funkce web 2.0.

Správa a řízení přístupu uživatelů - klasifikace a profily uživatelů na základě jejich rolí (např. RBAC), autentifikace uživatelů na základě uživatelského jména a hesla, řízení přístupu IP adres, přístup přes proxy, monitorování a reporting pro vyhodnocování chování uživatelů pro pozdější využití ke zlepšení služeb digitální knihovny, autorizace uživatelů (např. Shibboleth, Eduroam, OpenID), zaznamenávání aktivit uživatelů pro účely účtování.

Administrace systému - tento požadavek se vztahuje ke správě provozního prostředí digitální knihovny, která může být zejména u digitálních knihoven velkého rozsahu velice pracná a náročná. Spadají sem nástroje pro správu metadat, nastavování automatických mechanismů sběru dat, indexování, automatické generování klíčových slov, formátování výstupů apod.

Jiné požadavky - uživatelská a administrátorská dokumentace, helpdesk (podpora ze strany výrobce/implementátora systému), rozšiřující funkce pro uživatele jako jsou diskusní a komunikační skupiny, automatické upozorňování na různé události (výskyt nových dokumentů splňující určitá kritéria), RSS výstupy, uživatelské schránky a komentáře k dokumentům, požadavky na provozní hardware a software apod.

Po tomto kroku a vyhodnocení prvotního sběru kritérií jsme se znovu vrátili k funkční specifikaci NUŠL a se znalostí základních požadavků na software jsme pro účely dalšího postupu při stanovování výběrových kritérií zformulovali stručný popis hlavního procesu a hlavních činností NUŠL:

Získávání obsahu - poskytovatelé obsahu zpřístupní obsah (metadata, případně i vlastní soubory) různými způsoby, prostřednictvím web services, přes standard OAI-PMH, přes e-mail, předáním dat přes FTP, na nosiči, přímým vložením. V případě, že metadata získaná přes OAI-PMH obsahují URL na soubor(y) a dohoda s poskytovatelem to umožňuje, stahují se i samotné soubory.

Zpracování obsahu - takto poskytnutá data je nutno upravit, metadata převést do formátu NUŠL (software musí umožňovat definici vlastní struktury metadat) pomocí konverzních šablon, soubory zkonvertovat do prezentačního formátu (PDF/A), připravit dávku pro import do databáze s možností její kontroly, uložit do databáze metadata a prezentační, případně i originální podobu souborů v různých formátech, vygenerovat indexy a klíčová slova, zformátovat vyhledávací výstupy.

Zpřístupňování obsahu - uživatelé přistupují přes www rozhraní anonymně jako veřejnost, pokud přistupují z určitých institucionálních adres, systém je přihlásí jako anonymní institucionální uživatele a automaticky jim přiřadí příslušné role, nebo se hlásí jménem a heslem, které jim umožní získat knihovní systém, případně jejich autorizace probíhá přes systémy jako jsou Shibboleth, Eduroam, OpenID. Uživatelé mají možnost prohlížet obsah s přístupovými právy v závislosti na svých rolích podle různých kritérií lineárně v seznamech i hierarchicky v stromové struktuře, případně i formou Web 2.0, vyhledávat nad metadaty i obsahem souborů pomocí booleovských výrazů, fulltextově, proximálně. Vyhledávání a třídění musí zohledňovat specifika češtiny včetně nezávislosti na diakritice. Mají možnost ukládat výsledky vyhledávání do trvalých uživatelských schránek, (nákupní košík), psát komentáře k obsahu i formou diskusních skupin, vyžádat si zprávy e-mailem nebo do své pracovní plochy o událostech ve sbírkách (výskyt nového záznamu atd.). Systém umožňuje zaznamenávat aktivity uživatelů pro účely účtování, případně zpřístupňovat placené služby prostřednictvím rozhraní na jiný systém.

Uchování a ochrana obsahu - kontrola kvality obsahu z hlediska správce obsahu, duplicity dat, kvality metadat včetně integritních omezení, možnost formulovat dotazy nad databází, kontrola primárních souborů na udržitelnost obsahu atd.

Administrace systému - administrátorské nástroje pro řízení, konfiguraci a administraci celého systému včetně řízení operací a jejich časování nad centrální databází i s možností ručních zásahů (např. posun spuštěné časové náročné úlohy na pozdější dobu), správa uživatelských účtů, rolí a přístupových práv, automatické upozorňování na různé události, možnost formulovat administrátorské dotazy nad databází, systém zálohování dat.

Na základě tohoto stručného popisu NUŠL a na základě již shromážděných prvotních požadavků na software byla zformulována obecná **funkční specifikace** softwarového řešení pro digitální úložiště, kde byla požadována následující funkcionalita:

- systém a **databáze** pro ukládání velkého počtu knihovnických záznamů na základě vlastní definice metadat z různých interních i externích zdrojů a pro jejich rychlé a efektivní vyhledávání a poskytování koncovým uživatelům přes webové **uživatelské rozhraní** podporující interoperabilitu s jinými digitálními úložišti dokumentů
- automatické testování kvality a integrity **metadat** (např. délka polí, numerický/alfanumerický obsah, povinná/nepovinná/závislá pole) a vytváření zpráv o případných nedostatcích pro administrátory a katalogizátory
- automatická extrakce **klíčových slov** z textových dokumentů na základě frekvence výskytu výrazů z definovaného slovníku
- **konverze metadat** z proprietárních formátů do standardních formátů i vlastního formátu NUŠL, plná konfigurovatelnost formátu vstupů i výstupů, (systém popisu struktury vstupních a výstupních formátů,

podobně jako XSLT pro XML), součástí systému by měla být sada nastavení konverzí z/do obvyklých a známých formátů, možnost editace těchto popisů přes uživatelské webové rozhraní

- formátování bibliografických metadat pro různé účely, např. pro různé výstupy při vyhledávání a prezentaci dokumentů, oddělení administrace obsahu dat a administrace jejich výstupního vzhledu, možnost editace těchto popisů přes uživatelské webové rozhraní
- automatický mechanismus pro sběr dat (**harvesting**) podporující běžné standardy, založený na OAI-PMH (Open Archive Initiative Protocol for Metadata Harvesting) a podporující obousměrnou interoperabilitu s jinými digitálními knihovnami, web harvesting (stahování dat prostřednictvím webu), web submission (poskytování dat přes webové rozhraní od autorizovaných subjektů včetně konverzí fulltextových dokumentů z různých jiných textových i obrázkových formátů), e-mail upload (možnost nahrávání dat poslaných e-mailem), podpora vyhledávání v jiných digitálních knihovnách
- systém **indexování** metadat, odkazů a fulltextových souborů a mapování těchto indexů na tagy metadat pro rychlé vyhledávání v databázi, správa definice indexů
- systém klasifikace záznamů v databázi podle různých kritérií pro pozdější vyhledávání (např. frekvence výskytu slov, hodnota tagu, počet zobrazení záznamu, počet vyžádání dokumentu apod.)
- kontrola vstupních dat proti existující databázi s nastavitelnými kritérii, např. pro zamezení duplicitního uložení dat
- administrátorské nástroje pro řízení, konfiguraci a administraci celého systému včetně řízení operací a jejich časování nad centrální databází i s možností ručních zásahů (např. posun spuštěné časově náročné úlohy na pozdější dobu), řízení přístupu k systému na základě rolí (RBAC - Role Based Access Control)
- hromadný vstup předem naformátovaných dat (XML) do centrální databáze včetně kontroly konzistence dat (bulk upload)
- uživatelské webové rozhraní pro vyhledávání, prezentaci a poskytování dokumentů s podporou různých nadstavbových funkcionalit, jako jsou osobní uživatelské schránky pro ukládání dokumentů, (obdoba nákupních košíků), nastavování různých upozornění na výskyt dokumentů s definovanými klíčovými slovy, diskusní uživatelské skupiny, komunikační uživatelské nástroje (nástěnky, uživatelské komentáře a hodnocení k dokumentům apod.)
- vyhledávací nástroj s webovým rozhraním umožňující vyhledávání podle slov i frází včetně složitějších dotazů s booleovskými kritérii, strukturovaná prezentace výsledků vyhledávání (třídění podle typu dokumentů), v případě nenalezení podle zadaných kritérií návrh alternativ

Na základě této obecné funkční specifikace a na základě prvotně shromážděných hlavních požadavků na software bylo pro účely **výběru softwarového** řešení pro digitální úložiště a zároveň pro účely vyhodnocení výběrového řízení stanoveno jemnější členění **kritérií** do třinácti skupin:

1. **Řízení obsahu** - nástroje a postupy podporující ukládání obsahu do digitální knihovny a řízení procesu tohoto ukládání, verzování
2. **Získávání obsahu** - import a export obsahu, podpora různých formátů dokumentů
3. **Metadata** - podpora různých metadatových formátů je důležitá pro indexování, ukládání, zpřístupňování a ochranu obsahu
4. **Podpora vyhledávání** - týká se celé řady vyhledávacích a prohlížečích funkcí, vyhledávání v metadatach, fulltextové vyhledávání, hierarchické prohlížení apod.
5. **Řízení přístupu a ochrana soukromí** - řízení přístupu a ochrana soukromí zahrnuje administraci hesel, uživatelských účtů s přístupovými právy včetně možnosti získat zapomenuté heslo atd.
6. **Podpora výstupů a dotazování** - toto kritérium souvisí s možností vyhodnocovat využití digitální knihovny a odhalovat vzorce chování uživatelů pro vylepšování poskytovaných služeb, zaznamenávání aktivit uživatelů pro účely účtování
7. **Udržitelnost, ochrana dat** - ochrana metadat, konzistence a integrity datové základny, zálohování, podpora případné migrace dat
8. **Interoperabilita** - možnost obousměrné integrace s jinými distribuovanými systémy na úrovni metadat, prohledávání i získávání a poskytování dokumentů, podpora OAI-PMH, Z39.50
9. **Uživatelské rozhraní** - tato kategorie se týká podpory více jazyků a schopnosti přizpůsobit uživatelské rozhraní různým potřebám různých uživatelů či rozdílných implementací
10. **Podpora standardů** - standardy jsou důležité pro sdílení a dlouhodobé uchování digitálního obsahu. Jedná se o celé spektrum oblastí standardů od metadat přes interoperabilitu až po formáty uložených dokumentů
11. **Nástroje pro automatizaci** - tato kategorie se týká nástrojů pro automatizované získávání obsahu, harvesting, generování metadat, údržbových činností atd.
12. **Podpora, služby** - důležitý aspekt u všech softwarových systémů. Mnoho dobrých systémů digitálních knihoven se mj. nachází v oblasti Open Source, kde je třeba na toto hledisko dbát zejména. Důležitá je dokumentace, helpdesk, sběr požadavků na vylepšení, případně diskusní fóra atd.

13. Hardware a provozní software - hardwarové nároky na úložiště dat, provoz a zálohování systému, zabezpečení jeho dostupnosti, podporované operační a databázové systémy atd.

Skupiny kritérií 1 až 9 a 11 až 12 byly podrobně rozpracovány do vyhodnocovací tabulky s uvedením vah pro jednotlivé skupiny kritérií, případně pro některá zvláště významná jednotlivá kritéria. Skupina 10 - Podpora standardů byla v tabulce rozpuštěna do jednotlivých relevantních skupin kritérií a skupina 13 - Hardware a provozní software byla ponechána mimo tabulku s možností popisu podle zadání. Tabulka zde pro omezený rozsah publikace není uvedena.

Oblast digitálních knihoven je velice dobře pokryta softwarovými systémy z oblasti Open Source - z oblasti s otevřeným zdrojovým kódem. Tato otevřenost při dodržení určitých podmínek znamená technickou i legální dostupnost zdrojového kódu pro uživatele a tedy ekonomicky výhodnou možnost získat kvalitní software. Open Source ovšem neznamená software zdarma, i zde je třeba při výběru hledět na určitou stabilitu vývoje, dostupnost služeb podpory a rozvoje softwaru atd. Druhou podpůrnou větví výběru softwarového řešení pro NUŠL se tedy stalo vyhodnocení a porovnání nejznámějších vybraných programů z této oblasti.

Hodnoceny a porovnávány byly systémy CDS Invenio, DSpace, Eprints, Fedora a Greenstone. Tyto systémy byly vybrány víceméně intuitivně na základě jejich rozšíření a známosti ve světě i v České republice. Porovnání bylo zaměřeno na základní vlastnosti vybraných systémů a provedeno na poměrně vysoké úrovni abstrakce pro informativní podporu výběrového řízení.

CDS Invenio

CDS Invenio¹⁰⁶ (dřívější název CDSWare) je vysoce modulární systém pro uchování a zpřístupnění velkého počtu knihovnických záznamů (Veselý et al., 2004). Vznikl a dále se vyvíjí ve švýcarském *European Organization for Nuclear Research* (dále jen CERN). Vývoj začal v roce 1993 primárně pro interní potřeby této proslulé vědecké instituce jako institucionální repozitář, od roku 2000 zahrnuje podporu multimédií, OAI-PMH atd. a šíří se pod licencí GNU GPL i mimo CERN. Nyní představuje vyzrálé řešení vyvíjené ve spolupráci s *École Polytechnique Fédérale de Lausanne* (dále jen EPFL) pro všeobecnou správu dokumentů, institucionální repozitář nebo rozsáhlý knihovní systém. Je lokalizován do 18 jazyků včetně češtiny, ruštiny a dalších. Umožňuje vlastní definici metadatového schématu, interní jmenná konvence je podle MARC21 a podporuje standard OAI-PMH.

Systém CDS Invenio je velice komplexní, propracovaný a flexibilní, lze jej nastavit mnoha různými způsoby pro nejrůznější použití, platí se za to ovšem jeho relativní složitostí při zprovoznění, údržbě a provozu. Ze systémů

¹⁰⁶ CERN Document Server

zahrnutých do tohoto hodnocení má ovšem nejvíce formalizovanou podporu při instalaci i provozu poskytovanou za úplatu na základě smlouvy. Úroveň služeb a podpory se mu blíží systém Eprints. Kromě systému samotného lze počítat i s lokalizací některých služeb do češtiny, případně slovenštiny.

CDS Invenio se skládá z mnoha více či méně nezávislých modulů s přesně definovanou funkcionalitou, které dohromady tvoří vysoce funkční a výkonný celek, jenž lze dále doplňovat a rozšiřovat prostřednictvím definovaného rozhraní o další moduly jiných výrobců. Následuje jejich stručný popis, z kterého jsou zřejmé další vlastnosti systému.

BibCheck je nástroj pro administrátora a správce obsahu pro automatizaci různých testů kvality metadat jako jsou integritní pravidla, formáty a rozsahy polí atd.

BibClassify automaticky extrahuje klíčová slova z dokumentů na základě frekvence výrazů obsažených ve slovníku (thesaurus).

BibConvert umožňuje konverzi metadat z jakýchkoliv strukturovaných formátů do jiných formátů. Pracuje podobně jako procesor XSLT. Vstupní a výstupní formáty jsou plně konfigurovatelné, popisy běžných formátů jsou obsaženy v dodávce.

BibEdit umožňuje editaci metadat prostřednictvím webového rozhraní.

BibFormat umožňuje formátovat bibliografická data mnoha různými způsoby. Odděluje administraci obsahu od formátování výstupů, může pracovat na pozadí v reálném čase, nebo může předformátovat často používané výstupy.

BibHarvest je harvester kompatibilní s **OAI-PMH**, umožňuje získávat metadata z jiných úložišť podporujících OAI-PMH a správu úložiště. To je vytvořeno přímo v databázi a zahrnuje správcovský modul, který umožňuje administraci úložiště nezávisle na hlavní správě dat. Úložiště lze zcela nebo částečně otevřít pro harvesting v rozsahu protokolu OAI-PMH. Data jsou poskytována v nezpracované formě, jednotlivé tagy jsou pojmenovány podle konvence MARC21.

BibIndex indexuje metadata, odkazy (reference) a textové soubory. Vytváří dva druhy indexů - slovní a frázové. Lze definovat celou řadu logických indexů (autor, název atd.). Indexy se skládají ze dvou částí - dopředný a zpětný index. Toto indexování umožňuje velice rychlé prohledávání.

BibMatch filtruje vstupní soubory XML proti obsahu databáze a hledá záznamy podle různých kritérií, např. pro zamezení duplicitních záznamů.

BibRank umožňuje nastavit různá třídící/řadící kritéria pro pozdější použití při vyhledávání, např. podle frekvence výskytu určitého slova, podle počtu stažení dokumentu atd. BibRank je nezávislý na modulu BibIndex.

BibSched je časovač úloh, který jako centrální modul celého systému umožňuje ostatním modulům přístup k databázi kontrolovaným způsobem, zamezuje tak střetům při přístupu k databázi a zajišťuje konzistenci při aktualizaci databáze. Umožňuje administrátorovi monitorovat a řídit frontu úloh.

BibUpload umožňuje ukládat nová data do databáze. Vstup musí být dobře formátovaný soubor XML zohledňující zvolené metadatové schéma. Ve většině případů jsou tato vstupní data výsledkem činnosti modulu BibConvert.

ElmSubmit je brána pro automatické nahrávání dokumentů z ověřených zdrojů prostřednictvím e-mailu. Obvykle se upřednostňuje poskytování souborů nebo harvesting prostřednictvím www.

MiscUtil je sada příležitostných nástrojů, které používají jiné moduly.

WebAccess je modul, který zajišťuje přístup uživatelů k systému. Využívá se technika RBAC (Role-Based Access Control - přístup podle rolí), kdy se uživatelé dělí do několika skupin podle jejich rolí a oprávnění v systému. Každá skupina má povolené jen určité operace v rámci systému.

WebAlert zajišťuje upozorňování uživatelů systému na události v systému, např. na výskyt nového dokumentu splňujícího určitá kritéria. Tato kritéria odpovídají typickým uživatelským dotazům. Upozornění mohou být uživatelům zaslána e-mailem nebo uložena v jejich košíku.

WebBasket umožňuje uživatelům ukládat dokumenty do osobního košíku, podobně jako se v internetovém obchodě ukládá zboží. Jeden uživatel může mít více košíků. Košík může být osobní nebo veřejný. Veřejný košík umožňuje sdílení dokumentů ve skupině uživatelů.

WebComment je skupinový nástroj pro hodnocení dokumentů a pro sdílení komentářů k dokumentům mezi čtenáři. V integraci se skupinovými nástroji WebBasket, WebGroup, WebMessage tools a WebComment je středem nástrojů pro skupinovou komunikaci v systému CDS Invenio.

WebHelp reprezentuje globální dokumentaci systému CDS Invenio pro uživatele a administrátory. Dokumentace specifická pro jednotlivé moduly je zahrnuta v jednotlivých modulech.

WebMessage umožňuje komunikaci mezi (eventuelně anonymními) uživateli systému prostřednictvím webových nástěnek, např. pro pozvánky do diskusních skupin atd.

WebSearch je modul, prostřednictvím kterého uživatelé zadávají dotazy do systému pro vyhledání určitých slov nebo frází v databázi. Umožňuje dva druhy dotazů: vyhledávání slov nebo frází. Dále umožňuje komplexní booleovské dotazy, vyhledávání pomocí regulárních výrazů, kombinované vyhledávání v metadatech a ve vlastních souborech. Uživatelé mohou záznamy prohlížet také podle indexů. Pokud je dotaz neúspěšný, systém nabízí alternativy a pomoc při vyhledávání. Metadata ve sbírkách jsou přímo přístupná k prohlížení podobně jako v populárním konceptu *Web Directories*. Ortogonální pohledy na dokumenty jsou organizovány do virtuálních sbírek, dokumenty lze klasifikovat podle více kritérií. Tato flexibilní organizace umožňuje vytvořit velice přehledná navigační schémata pro uživatele.

WebSession je manažer relací a uživatelů, umožňuje rozlišovat uživatele a personalizovat uživatelské rozhraní jako jsou osobní košíky atd.

WebStat je konfigurovatelný systém, který umožňuje sběr údajů a statistik o systému jako takovém i o jeho užívání.

WebStyle je knihovna modulů, které definují design a vzhled systému CDS Invenio.

WebSubmit je komplexní systém, který autorizovaným uživatelům umožňuje ukládat do systému jednotlivé dokumenty. Systém poskytování obsahu zahrnuje kontrolní mechanismy, konverze různých formátů dokumentů a extrakci bibliografických údajů.

Více informací najdete na adrese <http://cdsware.cern.ch/invenio>.

DSpace

DSpace pochází jako většina systémů pro digitální knihovny z oblasti **Open Source** z univerzitního prostředí, vývoj byl zahájen knihovnou MIT a společností Hewlett-Packard, první verze pochází z roku 2002. Nyní na projektu spolupracují další vysoké školy a jiné organizace, celkem cca 500 organizací v 60 zemích. Vývoj zastřešuje nezisková organizace DSpace Foundation, Massachusetts. Vývoj a hlášení chyb je technologicky zajištěno prostřednictvím SourceForge,¹⁰⁷ pro vývoj jsou stanovena jasná pravidla, novými funkcemi může přispívat každý člen komunity.

DSpace je software určený pro tvorbu digitálních úložišť různých typů dokumentů pro nejrůznější typy organizací. V současné době podporuje metadatové schéma Dublin Core, pro určení perzistentních identifikátorů používá **CNRI Handle System**,¹⁰⁸ podporuje OAI-PMH 2.0 jako poskytovatel dat. Podporuje také standard OpenURL, umožňuje export dat do jednoduchého formátu XML nebo do formátu METS.

Data jsou v systému uložena v digitálních objektech s jednoznačným identifikátorem, které obsahují data i metadata. Dále existují v systému i struktury pro další logické členění záznamů do kolekcí a sbírek. Objekt může patřit do více logických struktur. DSpace podporuje tvorbu uživatelských účtů, seskupování uživatelů do skupin, přidělování přístupových práv a základní systém autentizace uživatelů.

Vyhledávat lze v systému DSpace prostřednictvím metadat i v samotných souborech, systém indexuje nejen prosté textové soubory, ale i některé další formáty, jako jsou PDF a DOC.

Poměrně kvalitní uživatelské webové rozhraní je tvořeno technologií kaskádových stylů a lze je tedy relativně snadno modifikovat. Poměrně čerstvou novinkou je systém Manakin, druhá verze uživatelského rozhraní systému DSpace, kde lze při modifikaci vzhledu pracovat se šablonami. Uživatelské

¹⁰⁷ <http://sourceforge.net/>

¹⁰⁸ <http://www.handle.net/>

rozhraní umožňuje všechny základní funkce, prohlížení a vyhledávání záznamů pro uživatele a ukládání záznamů pro správce obsahu.

Na úrovni vkládání obsahu lze definovat pracovní postupy, které zahrnují několik úrovní schvalování obsahu před jeho vlastním uložením do databáze.

Systém DSpace má poměrně širokou a živou komunitu uživatelů a lze jej relativně snadno zprovoznit, má však omezenou funkcionalitu, modifikace je nutno většinou programovat, má také specifické nároky na provozní systém UNIX a řada provozních komponent.

Více informací najdete na adresách

<http://www.dspace.org> a <http://www.ics.muni.cz/dspacecz>.

EPrints

Eprints je komplexní systém s mnoha instalacemi po celém světě pocházející opět z univerzitního prostředí, je vyvíjen v *University of Southampton* ve Velké Británii, k dispozici je pod licencí GNU. Celý systém je založen na webových technologiích, primárním účelem je tvorba institucionálních úložišť mnoha typů dokumentů od běžné literatury přes vědecké a vysokoškolské práce po nejruznější jiné typy dokumentů včetně multimédií s primárním zaměřením na vědecké informace. Přestože celý systém je značně konfigurovatelný a lze při implementaci vytvořit obecné úložiště, jeho primární orientace na vědeckou oblast může takové úsilí komplikovat a omezovat. Při instalaci a provozu lze využívat placené služby, jako jsou školení, řízení implementačního projektu, technická podpora atd., a to vše v angličtině.

V oblasti standardů podporuje systém EPrints OAI-PMH, metadata jsou reprezentována ve vlastním vnitřním formátu. Systém umožňuje import dat z formátu XML a z některých externích zdrojů (PubMed a dalších), export je možný v mnoha formátech XML, RSS, DublinCore, METS.

Vnitřní architektura nepoužívá pojem digitální objekt, ale Item (položka). Položka v sobě zapouzdřuje metadata i vlastní soubory. V rámci jedné instalace lze vytvořit více úložišť i pro odlišné organizace, případně více oddělených sbírek pro jednu organizaci s odlišným vzhledem i strukturou.

Systém Eprints podporuje správu uživatelských účtů, přidělování práv je méně propracované. Systém rozlišuje tři typy uživatelských účtů - administrátor, editor a běžný uživatel. I běžný uživatel může do systému vkládat data, a to velice podrobná, zejména na úrovni metadat.

Systém EPrints má podobně jako systém DSpace komplexní uživatelské rozhraní založené na webových technologiích. Rozhraní lze poměrně dobře konfigurovat, část rozhraní tvoří statické stránky v HTML, které systém generuje pro vložené položky, ale nepočítá se s denním vkládáním velkého počtu záznamů.

Prostřednictvím rozhraní lze vyhledávat v metadatech i v datech, kromě textových souborů systém indexuje i některé jiné běžné formáty (např. PDF),

obsah lze procházet i v logických stromových strukturách, výchozí struktura je stejná jako v *Library of Congress USA*, lze ji však modifikovat. Rozhraní dále umožňuje registrovat nové uživatele, informovat se o novinkách v systému, administrovat jej atd. Celý systém je velice komplexní, vkládání položek je propracované, nicméně relativně složité a časově náročné. Počítá se především s publikujícími vědci, kteří vkládají malé množství jednotek s velkým objemem informací ke každé. Tento rys může komplikovat tvorbu obecných úložišť.

Více informací najdete na adrese <http://www.eprints.org>.

Fedora

Systém Fedora¹⁰⁹ opět pochází z univerzitního prostředí *Cornell University a University of Virginia* (Fedora Commons, 2008). Vše začalo jako výzkumný projekt v letech 1997 a 1998. V roce 2001 se spojily obě vysoké školy a získaly prostředky na další vývoj od Mellonovy nadace se zadáním vyvinout univerzální digitální knihovnu na základě webových služeb a XML. V roce 2007 obě vysoké školy založily organizaci Fedora Commons, která se nyní stará o vývoj jejich společného systému.

Systém Fedora je digitálním úložištěm, které umožňuje ukládat a uchovávat v podstatě libovolné digitální objekty. Nicméně nabízí právě jen služby úložiště ve formě jádra, které se stará o ukládání a správu digitálních objektů, plus programátorské knihovny funkcí a volání služeb, které umožňují vývojářům obsluhovat úložiště. Systém Fedora nemá kromě základního rozhraní GNOME propracované vlastní uživatelské rozhraní, to je nutné vytvořit pro každé nasazení systému zvlášť. Přestože je systém Fedora velice kvalitním úložištěm respektujícím základní teoretické poznatky v oblasti budování digitálních knihoven, tak skutečnost, že je nutno při jeho nasazení dotvářet uživatelské rozhraní, brání jeho použití v projektech, kde je nutné rychlé nasazení, případně kde nejsou vývojové kapacity. Pro tvorbu uživatelského rozhraní jsou ovšem v současné době k dispozici kvalitní nástroje třetích stran, například eSciDoc - sada služeb v architektuře SOA, vývojový nástroj RepoMMan aj.

Základním prvkem architektury systému je podobně jako u jiných systémů digitální objekt, nicméně zde se nejedná pouze o objekt ve smyslu uloženého záznamu, ale systém rozlišuje tři druhy objektů:

- datový objekt (Data Object), který obsahuje metadata a data, případně jednoznačný perzistentní identifikátor ukazující na data v jiném úložišti, systém umožňuje verzování - ukládání různých verzí téhož dokumentu,
- dále objekt popisu chování (Behaviour Definition Object) služeb, které se váží k určitému souboru (např. zobrazení dat určitého typu), tyto služby jsou popsány jazykem WDSL a zajišťují tak relativní platformovou nezávislost pro další typ objektu,

¹⁰⁹ Flexible Extensible Digital Object Repository Architecture

- objekt mechanismu implementace chování (Behaviour Mechanism Object), což je např. software, který umožní data zobrazit, tyto služby mohou být napsány teoreticky v jakémkoliv programovacím jazyce.

Systém Fedora podporuje řadu standardů OAI-PMH, export do formátů METS a vlastního interního formátu FOXML, popisná metadata jsou ukládána ve formátu Dublin Core.

Přestože je jádro systému Fedora velice pokrokové, nejedná se v současné době o plnohodnotný kompletní knihovní systém. Je to kvalitní základ, který je potřeba dotvořit s poměrně velkými náklady a s velkým úsilím. I při provozu je nutno kvůli platformové nezávislosti systému třeba počítat s větší režií, je náročnější na hardwarové zdroje než ostatní systémy.

Více informací najdete na adrese <http://fedora-commons.org/>.

Greenstone

Systém Greenstone¹¹⁰ je exotičtější svým původem než jeho kolegové, nicméně opět se jedná o univerzitní produkt. Je vyvíjen od roku 1995 na *University of Waikato* na Novém Zélandu a vývoj podporuje i organizace UNESCO. Je dostupný pod licencí GNU GPL. Přes svůj exotický původ je systém kromě jiných jazyků částečně lokalizován i do češtiny. Systém podporuje mnoho formátů digitálních objektů.

V oblasti standardů systém podporuje metadatové schéma Dublin Core, protokol OAI-PMH a jako jediný ze zkoumaných systémů podporuje protokol Z39.50 pro obousměrné prohledávání katalogů jiných knihoven, nicméně tato podpora není standardně nastavena a systém je nutné pro její zprovoznění překompilovat.

Základní jednotkou pro uložení v systému je dokument, který se ukládá do předem připravených sbírek, v nichž je definováno, jaké dokumenty budou obsahovat, a tyto sbírky se pak sdružují do knihoven. Pro zpracování jednotlivých typů dokumentů musí systém obsahovat programovou podporu, tzv. pluginy. Tato architektura klade větší nároky na nastavení systému, než je tomu u některých jiných kompletnějších systémů. Při uložení dokumentu do předem definované sbírky v interním formátu XML je vygenerován i jednoznačný identifikátor ve vlastním formátu.

Uživatelské rozhraní je oddělené od vlastního úložiště jako separátní aplikace. Úložiště i uživatelské rozhraní spolu komunikují vlastním interním protokolem, který lze nahradit protokolem CORBA a připojit tak k systému vlastní aplikaci. Uživatelské rozhraní lze nastavovat pro každou sbírku jinak a je založené na webových technologiích. Prostřednictvím něj lze procházet sbírky a vyhledávat v nich. Procházení sbírek je založeno na řazení příbuzných dokumentů

¹¹⁰ Greenstone Digital Library Software

do společných složek, které lze hierarchicky propojovat a vytvářet tak logickou prohledávací strukturu.

Systém obsahuje správu uživatelů a přidělování uživatelských přístupových práv, ale nemá tuto podporu příliš dobře propracovanou. Omezuje se na přidělování práv administrátorům a správcům obsahu, nepočítá se s aktivitou jiných uživatelů.

Systém je zcela jistě životaschopný, jsou k němu dostupné i služby poskytované organizací *DL Consulting* na Novém Zélandě, nicméně jeho nasazení je relativně složité a použití je omezeno na organizace, kde se nepředpokládá vkládání dat uživateli mimo správce sbírek.

Více informací najdete na adrese <http://www.greenstone.org>.

V Tabulce 1 jsou uvedena přibližná číselná skóre jednotlivých systémů podle výběrových vyhodnocovacích kritérií uvedených výše na stupnici od jedné do deseti:

	CDS Invenio	DSpace	Eprints	Fedora	Greenstone
Řízení obsahu	10	8	8	5	9
Získávání obsahu	10	9	10	8	9
Metadata	6	5	5	7	5
Podpora vyhledávání	8	6	5	4	7
Řízení přístupu	8	6	8	4	5
Podpora výstupů	6	6	4	6	10
Udržitelnost, ochrana dat	8	8	7	10	6
Interoperabilita	6	6	6	7	8
Uživatelské rozhraní	8	8	7	5	7
Podpora standardů	9	9	9	9	10
Nástroje pro automatizaci	9	7	5	5	8
Podpora, služby	9	7	8	6	6
Součet	97	85	82	76	90

Tabulka 1: Číselná skóre jednotlivých systémů

Pokud jsou uvedené systémy hodnoceny jen podle jejich vlastností bez ohledu na vhodnost nasazení v podmínkách NUŠL, nejlépe vycházejí systémy Greenstone a CDS Invenio, těsně následovány systémem DSpace. Fedora není hotový systém a jeho dotvoření je pracné a nákladné, systém Eprints je poměrně jednoúčelový pro menší denní přírůstky důkladně popisovaných vědeckých dokumentů.

Pro nasazení v České republice je dostupností služeb a jazykové lokalizace vhodný systém CDS Invenio, který je zároveň velice funkčně bohatý a flexibilní, za což se ovšem platí jeho relativní implementační a provozní složitostí.

Vybranému softwarovému řešení, popisu jeho funkcionalit a stručnému náhledu přímo do správy systému je věnována následující kapitola.

LITERATURA:

VESELÝ, Martin - BARON, Thomas - Le MEUR, Jean-Yves - ŠIMKO, Tibor. *CERN Document Server: Document Management System for Grey Literature in a Networked Environment*. Publishing Research Quarterly, Spring 2004.
Introduction to Fedora, Fedora Commons 2008, Inc. Dostupné z [www:](http://www.fedora.info)
<http://www.fedora.info>

9. Pilotní implementace NUŠL

Jindřich Dolanský a Petra Pejšová

Cílem pilotní implementace NUŠL je vytvoření efektivního automatizovaného sběru dat (metadat i digitálních dokumentů), jejich dlouhodobá archivace a uživatelsky příjemné a snadné zpřístupnění. K zajištění tohoto cíle byly specifikovány požadavky na funkcionalitu softwaru pro digitální repozitář a návazné služby, které jsou podrobně rozebrány v předchozí kapitole a sloužily jako podklad pro výběrové řízení. Požadavky na funkcionalitu softwaru pro NUŠL byly definovány tak, aby zahrnovaly jak požadavky nutné pro pilotní realizaci systému, tak aby na jejich základě byla vybrána moderní a podporovaná technologie s perspektivou dalšího rozvoje. V souvislosti s přípravou výběru softwarového řešení pro NUŠL byla také provedena analýza vybraných Open Source software pro digitální knihovny, které je věnována druhá část předchozí kapitoly. Následně proběhlo veřejné výběrové řízení, kde zadaným požadavkům nejlépe vyhovovalo a bylo vybráno softwarové řešení, jehož základními prvky jsou CDS Invenio pro digitální repozitář a indexační a vyhledávací systém FAST ESP¹¹¹ pro uživatelské rozhraní. Toto řešení získalo jednak nejvyšší ohodnocení dle zadaných specifik, ale také reflektovalo výsledky analýzy, které rozvinulo a nabídlo pokročilé řešení. V tomto případě je nad digitálním repozitářem implementováno uživatelské rozhraní, které umožňuje jak snadné a příjemné vyhledávání a zpřístupnění dat, tak i funkcionalitu integrující platformy samostatných repozitářů. Stejná architektura řešení je úspěšně provozována již několik let ve švýcarském CERN.

V této kapitole je detailně popsán z pohledu správy systém CDS Invenio pro digitální repozitář, protože jeho přednastavená lokální instalace pro sbírky šedé literatury je nabízena spolupracujícím organizacím k instalaci na vlastním hardwaru. Indexační a vyhledávací systém FAST ESP je zde popsán pouze stručně z uživatelského hlediska a ne z pohledu jeho správy, protože tu může provádět pouze certifikovaný administrátor a v rámci projektu je tato služba poskytována externí firmou. V závěru kapitoly jsou popsány další výchozí předpoklady pro pilotní implementaci, které představují perzistentní identifikátory a formáty digitálních dokumentů.

CDS Invenio, jehož popis je v předchozí kapitole, bylo vybráno pro digitální repozitář NUŠL také proto, že patří mezi Open Source software. Je možné ho volně instalovat, používat i upravovat, což umožňuje jeho nastavení pro ukládání šedé literatury a následné šíření mezi spolupracující organizace. Systém CDS Invenio byl v první fázi instalován v základní verzi, tak jak je šířen mimo CERN. Zde byla využita virtualizační aplikace **VirtualBox**, kterou lze

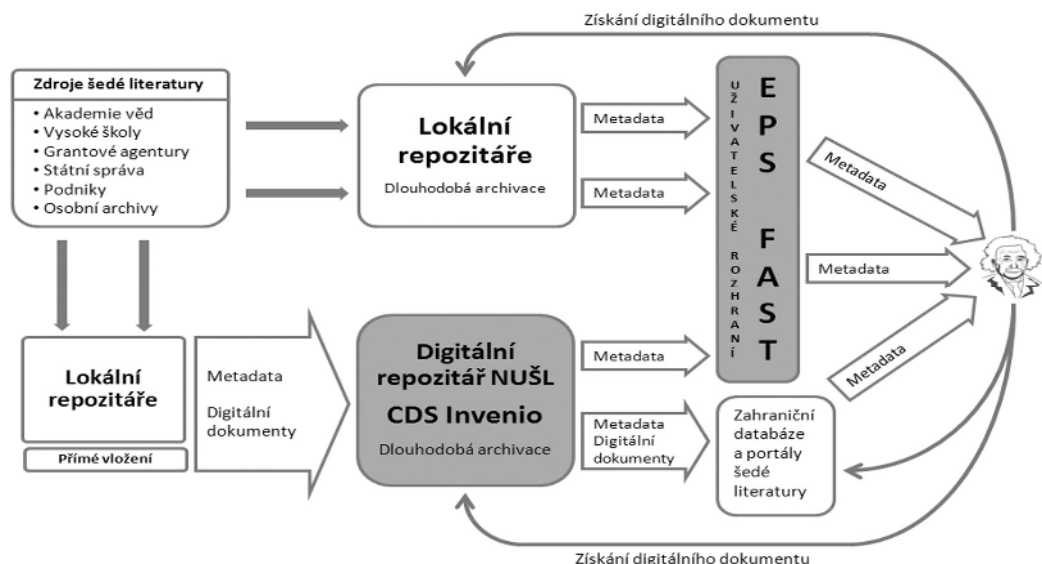
¹¹¹ http://www.incad.cz/jnp/cz/www/produkty/fast/-content-www-produkty-fast-fast_esp.html

nainstalovat na téměř všechny známé platformy.¹¹² V rámci Virtual Boxu bylo instalováno CDS Invenio nad operačním systémem Linux-Debian. Toto řešení umožňuje snadnou distribuci přednastavené verze NUŠL spolupracujícím institucím.

Indexační a vyhledávací systém ESP FAST zabezpečuje uživatelsky příjemné a snadné zpřístupnění dat. Uživatelské rozhraní v systému ESP FAST poskytuje zabezpečené, relevantní a škálovatelné vyhledávání včetně skloňování českých termínů. Vlastnosti řešení ESP FAST umožňují určit kontext a účel dotazu, vyhledat odpovídající termíny jak v metadatech tak v dokumentech a vrátit odpovědi vyskytující se v souvislostech. Uživatelé získají přesné výsledky a kontextovou dynamickou navigaci pro další hledání souvisejících informací. Tyto funkce jsou na příkladu popsány dále v této kapitole.

Výhodou systému ESP FAST je jeho schopnost vytvořit integrující vyhledávací platformu nad repozitáři vědeckých zdrojů obsahujících šedou literaturu. Díky tomu je možné vytvořit již v pilotní fázi projektu komfortní centrální vyhledávací rozhraní, které dokáže zpřístupnit uživatelům data jak z digitálního repozitáře NUŠL, tak z distribuovaných úložišť šedé literatury v jednom interaktivním prostředí. Tento trend představuje posun od budování sbírek k nástrojům, které zajistí vyhledání a přístup dat, zmiňuje v druhé kapitole v části Od sbírky k otevřenému portálu Joachim Schöpfel.

Jednotlivé činnosti a spolupráce digitálního repozitáře v CDS Inveniu a indexačního a vyhledávacího systému FAST ESP jsou znázorněny ve schématu.



Obrázek 6: Schéma software řešení NUŠL

¹¹² Virtuální prostředí VirtualBox podporuje jak operační systém Windows (především XP), tak mnoho Linuxových distribucí (např. Debian, Ubuntu, CentOS, Fedora aj.). Od verze VirtualBox 1.6 jsou podporovány také platformy Mac OS X a Solaris.

CDS Invenio jako digitální repozitář NUŠL

Systém CDS Invenio umožňuje příjem a získávání dat, jejich archivaci a správu a nabízí také široké možnosti správy přístupu. Tyto funkce budou postupně popsány z hlediska systému a jeho funkcionalit.¹¹³ Dále bude popsána problematika formátů v rámci CDS Invenia a předvedeno, že ačkoliv žádná omezení z hlediska softwarového řešení samotného předem kladená nejsou, tak je téměř nezbytné je definovat kvůli budoucí administraci archivu. Nakonec budou popsány komponenty softwarového řešení NUŠL, které umožňují, aby CDS Invenio bylo online.

Základní funkce systému CDS Invenio

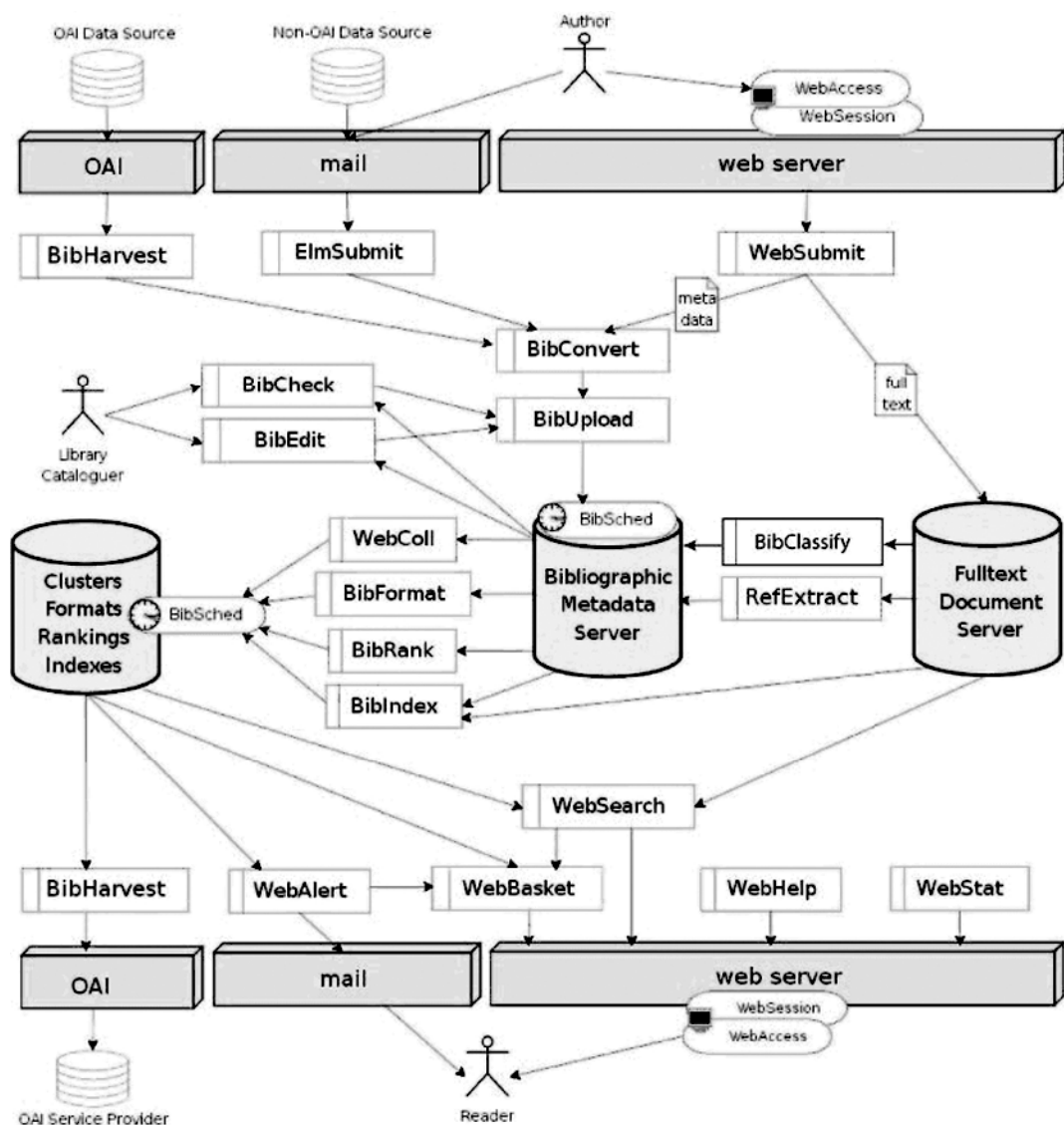
V této části jsou popsány základní funkce systému CDS Invenio s cílem představit variabilitu, které systém dosahuje díky své modulární architektuře. Tato architektura umožňuje spravovat repozitáře o velkém objemu dat a navzdory komplexnosti služeb nad daty si zachovává rychlost.

Přijem/získávání dat

Systém CDS Invenio umožňuje získávat data více způsoby, jak ilustruje obrázek znázorňující workflow diagram CDS Invenia. Díky modulární architektuře CDS Invenia je každý způsob získávání prováděn odpovídajícím specifickým modulem. Jsou možné tři způsoby příjmu dat:

1. přímé vkládání do systému,
2. získávání bibliografických metadat harvestováním jiných OAI repozitářů,
3. získávání metadat formou e-mailu.

¹¹³ V rámci tohoto textu není možné probírat detaily, které jsou přístupné ve vlastní dokumentaci.



Obrázek 7: Workflow diagram (Pepe, 2005, s. 6)

Vložení do systému - modul **WebSubmit**

Modul WebSubmit představuje soubor funkcionalit, s pomocí nichž probíhá vkládání metadat, popř. samotných dokumentů. Konfigurace funkcionalit se provádí ve webovém rozhraní. Toto rozhraní poskytuje intuitivní nástroje pro konfiguraci:

1. Definice WebSubmit elementů

Dopředu se definují elementy modulu WebSubmit, které budou následně používány. Tyto elementy v podstatě představují MARC pole, do kterých se jejich obsah bude později ukládat. Jedná se vlastně o klasické

formulářové prvky jako textový vstup, textové pole či roletové menu. Speciálním případem je pak vstup pro uploadovaný soubor, díky kterému je možné například ukládat digitální dokumenty.

2. Definice akcí

Dále se pouze definují, nikoli implementují akce, které se budou provádět, jako například vložení záznamu, modifikace záznamu, vložení souboru atd.

3. Definice typů dokumentů a kategorií

Definují se typy dokumentů (např. konferenční materiály, zprávy, VŠKP) a jejich další dělení do kategorií (tj. postery, prezentace, výroční zprávy, výzkumné zprávy, diplomové práce)

4. Implementace akcí

Pro každý typ dokumentu se zvolí akce, které se budou provádět. Pro každou akci bude vytvořeno rozhraní složené z definovaných elementů a vznikne tak webový formulář. Ten je po jeho vyplnění a odeslání zpracován sérií funkcí, které zajišťují např. vložení záznamu, vytvoření ID záznamu, přejmenování a přesunutí souborů do příslušného adresáře, extrakce dat z formuláře a jejich transformace do formátu MARCXML, zařazení transformovaného souboru do fronty k uploadu, oznámení o úspěšnosti vložení, zaslání e-mailu vkladateli a zálohování originálních dat.

5. Kontroly zadavatelského vstupu

Data, která zadavatel¹¹⁴ vloží, je možné kontrolovat pomocí funkcí napsaných v jazyce Javascript. Lze tak ošetřit například to, aby zadavatelé vkládali kalendářní data či jména ve správném formátu.

Harvestování¹¹⁵ metadat - modul BibHarvest

Modul BibHarvest je nástrojem pro definici a harvestování jiných OAI¹¹⁶ repozitářů. Konfigurace se provádí ve webovém rozhraní, zatímco vlastní proces se spouští z příkazové řádky. Ve webovém rozhraní se definují parametry pro harvestování jednotlivých repozitářů, kterými jsou jméno repozitáře, začátek harvestování (od začátku, či od data), způsob konverze takto získaných záznamů a nahrávání konvertovaných záznamů do databáze.

¹¹⁴ Termín zadavatel používáme v textu pro pracovníka, který je pověřen zadáváním a úpravami metadat a vkládáním digitálních dokumentů do systému CDS Invenio.

¹¹⁵ Běžně se používá výraz harvestování, který je převzat z anglického jazyka. Do českého jazyka lze tento termín přeložit jako sklizení či shrabování.

¹¹⁶ OAI-PMH Open Access Initiative - Protocol for Metadata Harvesting.

Z příkazové řádky se následně spouští vlastní harvestování OAI repozitáře definovaného ve webovém rozhraní. Pokud příslušný příkaz oaiharvest obsahuje parametr periodicity, je zařazen do fronty periodických úkolů, které spravuje plánovací démon BibSched.

Dalším z parametrů je harvestovaný formát metadat. Obvykle jsou metadata v OAI repozitářích k dispozici ve dvou či více formátech. Z protokolu OAI-PMH plyne, že formát Dublin Core je povinně nabízeným metadatovým formátem. Dále bývá k dispozici alespoň jeden další formát, který buď může vycházet ze standardních formátů, např. MARCXML, anebo bývá „ušit na míru“ dané instituci, např. metadatový formát EVSKP-MS pro eVŠKP.

Harvestování metadat pomocí e-mailu - modul ElmSubmit

Tvůrce systému CERN nyní vyvíjí další způsob, jak získávat data do archivu CDS Invenio. Jedná se o Modul ElmSubmit, který umožní vkládání dokumentů formou e-mailů. Tato funkcionality je již též znázorněna ve workflow diagramu.

Archivace a správa uložených dat

Na Obrázku 2 je vidět, že data jsou uložena podle typu na třech různých serverech, na jejichž obsah se aplikují rozdílné mechanismy:

1. Fulltext dokument server

Fulltext dokument server obsahuje textové (popř. zvukové, vizuální nebo multimediální) soubory vložené spolu s bibliografickými metadaty. Pro každý typ souborů je vhodné zvolit konkrétní formát, což umožní lepší budoucí správu, viz níže.

2. Server pro bibliografická metadata

Server pro bibliografická metadata je propojen s fulltextovým serverem pomocí referenčních čísel, která identifikují dokument příslušný k danému záznamu. Předtím, než jsou metadata získána protokolem OAI-PMH uložena na server, jsou v případě potřeby konvertována pomocí modulu BibConvert na základě transformačních souborů popisujících potřebné úpravy jednotlivých metadatových prvků. Tento modul může být také využit při konvertování metadat ze starého systému.

3. Server formátů, třídících metod a indexů

Provede-li se nad metadaty nebo fulltextem operace formátování, třídění nebo indexování, vznikne obsah třetího serveru.

Správa dat

Na jednotlivé typy dat můžou být aplikovány jim odpovídající moduly. Srovnáním s Obrázkem 2 je vidět, že:

1. Modul BibClassify pro automatickou klasifikaci

Pomocí modulu **BibClassify** se automaticky extrahují klíčová slova vzhledem ke zvoleným řízeným slovníkům a heslářům. Modul BibClassify dokáže pracovat s hesláři ve dvou formátech. Prvním formátem je heslář jako jednoduchý seznam klíčových slov a druhým heslář využívající RDF SKOS taxonomii, který umožňuje extrahovat tzv. koncepty s bohatší a komplexnější strukturou. Stručně řečeno, modul BibClassify počítá výskyt klíčových slov ve fulltextu, která jsou obsažena v hesláři a následně je uspořádá podle jejich počtu. Na fulltext lze také aplikovat funkcionalitu **RefExtract**, pomocí které lze extrahovat seznam referencí obsažených v tomto dokumentu. Tuto funkcionalitu lze použít pouze pro formát PDF.

2. Modul BibConvert pro konverzi dat

Při konvertování do vnitřního formátu užívaného CDS Inveniem prochází metadata základní úpravou prováděnou modulem **BibConvert**. Tento modul je buď volán jinými moduly (např. WebSubmit), anebo je možné jej spustit manuálně, kdy se např. zkonvertuje XML soubor dle předepsané transformace. Pro práci nad metadaty v databázi má zadavatel dva moduly. Pomocí prvního modulu **BibEdit** může přes webové rozhraní přistoupit přímo ke konkrétnímu záznamu a v něm podle potřeby modifikovat, přidávat, či mazat jednotlivá pole. Druhý modul **BibCheck** umožňuje provádět další úpravy a modifikace z příkazové řádky.

3. Modul BibIndex pro tvorbu rejstříků

Na metadata lze aplikovat další nástroje. Základním z nich je modul **BibIndex**, s jehož pomocí lze indexovat vybraná metadatová pole. Modul **BibIndex** umožňuje konfigurovat tzv. word files, které definují metadatová pole pro slovní indexy. Například pro metadatové pole autor je vytvořen rejstřík všech autorů. Tyto slovní indexy jsou pak využity při vyhledávání. Např. je definováno, že index pro pole autor je přiřazen MARCovým polím 100 a 700. Záznamy mohou být navíc zobrazeny v různých metadatových formátech, které lze definovat ve webovém rozhraní modulu **BibFormat**. Dále je možné metadata třídit podle definovaných kritérií a toto roztřídění může být dále použito při vyhledávání např. podobných záznamů podle obsahu podobných slov. Kritéria třídění definujeme pomocí modulu **BibRank**.

Je vhodné, aby některé z dosud zmíněných modulů systému CDS Invenio byly spouštěny periodicky, např. reindexace, zformátování, přetřídění po získání nových záznamů nebo pravidelné harvestování repozitářů či pravidelné

znovuvystavení lokálního repozitáře (či jeho podmnožin, tzv. setů) pro harvestování dalšími institucemi. Periodičnost těchto procesů zajišťuje plánovací démon **BibSched**, pomocí něhož se nejen plánují úkoly pro další moduly, ale také jim přiděluje různou prioritu.

Vytváření a konfigurace **sbírek** je zajištěna modulem **WebSearch**. Pomocí modulu WebSearch se definuje rozhraní pro vyhledávání ve sbírkách a vyhledávání samotné (indexy atd.). V případě NUŠL jsou sbírky strukturovány do dvou dvouúrovňových stromů podle typu dokumentu a podle instituce.



Obrázek 8: Digitální repozitář NUŠL dostupný z <http://invenio.ntkcz.cz/>

Sbírky pro vkládání dokumentů v digitálním repozitáři NUŠL

A) Podle typu dokumentu

Sbírky - skupiny druhů dokumentů, např. materiály z konferencí, VŠKP, zprávy apod.

Podsíbky - jednotlivé druhy dokumentů, např. výroční zprávy, závěrečné zprávy, disertační práce, diplomové práce apod.

B) Podle instituce

Sbírký - skupiny podle druhu instituce, např. ústavy AV ČR, vysoké školy apod.

Podsbírký - jednotlivé instituce, např. Geologický ústav AV ČR, Geofyzikální ústav AV ČR, VŠE apod.

Dalším důležitým modulem je modul **OaiArchive**, díky němuž může archiv CDS Invenio sloužit také jako OAI poskytovatel dat. Data získaná ať přímým vložením, harvestováním či jinak, jsou vystavena v protokolu **OAI-PMH** ve formátu MARCXML (popř. ve formátu Dublin Core). Archiv CDS Invenio hraje tedy i roli OAI poskytovatele dat vůči roli harvestora zprostředkované modulem BibHarvest. Systém CDS Invenio umožňuje dále definovat množiny tzv. sety repozitáře, které je možné specifikovat jednotlivými sbírkami. Dále je možné v rámci těchto jednotlivých sbírek vybírat podmnožiny identifikované podmínkou tzv. phrase, např. na základě hodnoty konkrétního metadatového pole.

Správa dokumentů

Součástí NUŠL je i řešení otázky dlouhodobého uchování a zpřístupnění dokumentů, a to v řádu mnoha desítek let. V současné době nelze předem určit, které formáty bude možné zobrazit na softwarových prostředcích, které se budou používat za padesát či sto let. Pro příklad můžeme uvést textové dokumenty ve formátu Text602, který v první polovině devadesátých let patřil mezi nejčastěji používané na českých osobních počítačích. Přestože uplynulo jen pět let od vydání jeho poslední verze, mnoho uživatelů by bez potřebných konvertorů mělo v současných letech problém tyto dokumenty otevřít a přečíst.

Zpřístupnění dokumentu je možné řešit **emulací**/virtualizací operačního systému a editoru použitého v době vzniku dokumentu nebo migrací dokumentu na vhodnější formát. Při **migraci** dokumentu do novějšího formátu je důležité zajistit maximální možnou míru zajištění autenticity dokumentu, i když někdy se např. není možné vyhnout určité změně ve vzhledu. V případě migrace je proto vhodné zachovávat originální dokument pro případ chyby v migraci či pro opětovnou migraci s vhodnějším konverzním nástrojem. **Formáty dokumentů** použité pro dlouhodobé uložení je možné rozdělit na:

- 1) preferované, které je vhodné dlouhodobě v repozitáři uložit a u kterých je předpoklad dlouhodobé použitelnosti,
- 2) akceptované, kterým je potřeba věnovat v následujících letech vyšší pozornost, ale není potřeba jejich bezprostřední migrace do preferovaného formátu,
- 3) neakceptované, tj. formáty, které je potřeba migrovat na formát preferovaný nebo alespoň akceptovatelný při ukládání do repozitáře.

Problematika dlouhodobého uchování a zpřístupnění šedé literatury v NUŠL je řešena v několika rovinách:

- 1) výběr preferovaných, akceptovaných a neakceptovaných formátů dokumentů v souladu s usnesením vlády České republiky ze dne 3. 11. 2008 č. 1338 o stanovení výstupních datových formátů statických dokumentů textové, obrazové a kombinované povahy v digitální podobě (Usnesení, 2008) a postupem **Národního digitálního archivu** (BERNAS),¹¹⁷
- 2) volbou migrace se zachováním originálního dokumentu jako základní strategie zajištění čitelnosti dokumentů,
- 3) vypracováním metodických pokynů zajišťujících pravidelné testy dostupnosti jednotlivých formátů dokumentů uložených v NUŠL a strategie případné konverze do formátů vhodnějších.

Správa přístupu

Z hlediska přístupu k repozitáři lze rozlišit tři základní skupiny uživatelů systému CDS Invenio:

1) Administrátor

Administrátor se stará o vlastní chod systému, což představuje především konfiguraci jednotlivých modulů, případně vhodné modifikace zdrojového kódu. Přiděluje také přístupová práva k těmto modulům pro zadavatele, zakládá uživatelské účty apod.

2) Zadavatel

Zadavatel dostává přidělená práva především ke vkládání, editaci a schvalování přijetí digitálních dokumentů.

3) Uživatel

Systém CDS Invenio umožňuje nastavit míru otevřenosti systému vůči uživatelům. Uživatel má právo přístupu k metadatům bez omezení a ke všem veřejným digitálním dokumentům. Dle nastavených práv může mít pak dále přístup k některým neveřejným digitálním dokumentům podle sbírek, např. pro danou množinu IP adres. Uživatel může v rámci svého uživatelského konta měnit svůj e-mail a heslo. Dále má možnost recenzovat dokumenty z vybraných sbírek apod.

Konfiguraci přístupu provádí administrátor pomocí modulu **WebAccess** přes webové rozhraní. Tento modul umožňuje definovat přístupová práva k jednotlivým administrátorským či zadavatelským činnostem. Nastavování omezení uživatelského přístupu k některým sbírkám, např. pro množinu IP adres, zahrnuje ovšem i práci s konfiguračními soubory a dalšími skripty.

¹¹⁷ <http://knihovna.nkp.cz/knihovna91/bernas.htm>

Další software

Systém CDS Invenio využívá webový server Apache a databázový server MySQL. Je ovšem vhodné začít vlastní instalací systému CDS Invenio, která předpokládá instalaci operačního systému GNU/Linux (př. Debian, Gentoo) nebo virtuálního prostředí VirtualBox.

VirtualBox

Běžně zájemce o provozování CDS Invenia musí vytvořit balíček z volně dostupného zdrojového kódu, ten pak instalovat a konfigurovat. NTK připravila pro spolupracující organizace přednastavenou defaultní NUŠL verzi, která předpokládá pouze instalaci aplikace VirtualBox.¹¹⁸ Tato aplikace vytvoří na lokálním serveru virtuální prostředí, které může obsahovat jeden či více virtuálních počítačů. Na jednom z těchto virtuálních počítačů bude provozováno CDS Invenio nad operačním systémem Linux Debian.

Po instalaci VirtualBoxu není třeba Debian ani CDS Invenio instalovat, celý virtuální počítač se importuje do VirtualBoxu, a to již s přednastavenou konfigurací CDS Invenia. Po nastavení síťových parametrů (IP, hostname atd.) virtuálního stroje může být spuštěn. Cílem instalace VirtualBoxu namísto přímo CDS Invenia bylo vyhnout se závislosti na operačním systému lokálního serveru, protože VirtualBox lze nainstalovat na všechny víceméně známé platformy.

Navíc je možné přenést virtuální počítač s CDS Inveniem exportovaný do formátu Open Virtualization Format (dále jen OVF) i do jiného virtuálního prostředí než do prostředí VirtualBox. Vyzkoušeno je též další často používané virtuální prostředí VMWare ESXi.

Webový server Apache

Jak bylo řečeno výše, CDS Invenio využívá standardní webový server Apache s otevřeným kódem. Pro účely zabezpečené komunikace mezi uživatelem a serverem systém CDS Invenio vyžaduje certifikovaný protokol HTTPS. Takto je zabezpečené např. přihlašování do uživatelského konta nebo registrace uživatelského účtu. Při instalaci CDS Invenia je možné generovat tzv. self-signed certifikát. Při publikaci CDS Invenia do vnější internetové sítě je ovšem nezbytné zajistit důvěryhodný veřejný certifikát. V případě projektu NUŠL a systému CDS Invenio je využíván serverový certifikát TERENA pro servery organizací začleněných do české sítě národního výzkumu CESNET2

¹¹⁸ <http://www.virtualbox.org/wiki/Downloads>

a poskytovaný firmou Comodo CA Limited, jednou z největších celosvětových certifikačních autorit.

*Databázový server **MySQL***

Součástí instalace systému CDS Invenia je také instalace databázového serveru MySQL, při které jsou vytvořeny všechny nezbytné tabulky a další nástroje této databáze. Databáze MySQL je snadno implementovatelná, volně šiřitelná a také především optimalizovaná na rychlost.

Python

Protože systém CDS Invenio je naprogramován v jazyce Python, je třeba instalovat také interpreter tohoto jazyka. Navíc je třeba přidat modul mod_python do serveru Apache, který integruje interpreter jazyka Python do tohoto webového serveru. Díky tomuto modulu lze efektivněji psát webové aplikace v Pythonu a pracovat s databází.

Centrální vyhledávací rozhraní NUŠL

Od konce roku 2009 je k dispozici centrální rozhraní NUŠL pro snadné vyhledávání šedé literatury v České republice. Do tohoto rozhraní se postupně připojují ústavy AV ČR, vysoké školy, grantové agentury a další. Pomocí tohoto rozhraní jsou zpřístupněny metadata a digitální dokumenty šedé literatury z oblasti vědy, výzkumu a vzdělávání. Centrální vyhledávací rozhraní obsahuje funkce **Web 2.0** a poskytuje vizualizaci dat a kontextovou dynamickou **navigaci**, které se mění se zpřesňováním dotazu. Všechna propojení vznikají dynamicky na základě aktuálně sestaveného dotazu a tak umožňují pracovat se souvislostmi mezi dokumenty.

Navigaci představují **textové filtry** v levé části obrazovky, viz obrázek níže, kterými jsou Typ dokumentu, Osoby, Klíčová slova a Jazyk. Navigace Typ dokumentu je strukturovaná do dvou úrovní, např. konferenční materiály se dále dělí na příspěvky, sborníky a prezentace. Navigace Osoby obsahuje jak autory dokumentů, tak další osoby spojené s tvorbou dokumentu, např. vedoucího práce u VŠKP. Protože zdrojové databáze poskytují termíny buď v českém nebo anglickém jazyce, jsou v navigaci Klíčová slova použity oba jazyky. V navigacích Osoby a Klíčová slova jsou primárně zobrazeny 4 termíny a maximálně 25 termínů. Navigace Jazyk poskytuje výběr všech jazyků digitálních dokumentů dostupných přes toto rozhraní. U jednotlivých položek v navigacích jsou vždy v závorce uvedené odpovídající počty záznamů, např. Sborníky (575). Počty záznamů v závorkách jsou vidět i u jednotlivých připojených zdrojovýchází v horní liště. Z obrázku je patrné, že zdroj NTK obsahuje 8 999 záznamů, zdroj VŠE 16 420 záznamů atd. Na obrazovce je

Vyhledávání v centrálním rozhraní je k dispozici na www.nusl.cz.

Obrázek 9: Centrální vyhledávací rozhraní NUŠL

Výchozí předpoklady pro pilotní implementaci NUŠL

Přijímané formáty digitální dokumentů

Dalším důležitým kritériem digitálního archivu je způsob, jakým pracuje s digitálními dokumenty a možnostmi, které nabízí. Digitální dokumenty mohou být textové, audio či video soubory. Protože také **formáty digitálních dokumentů** podléhají časté změně, je vhodné při budování archivu myslet na udržování digitálních zdrojů v aktuálních formátech, jinak by se brzy staly nečitelnými. Jinými slovy, je třeba vybírat co možná nejrozšířenější formáty, pro něž existují konverzní nástroje, které umožní pravidelnou migraci na aktuálnější verzi. Pro všechny typy digitálních objektů existují takové typické formáty např. PDF a DOC pro textové dokumenty atp. Navíc je třeba přihlídnout k možnostem, které k daným formátům poskytuje samotný archiv.

CDS Invenio v principu neklade žádná omezení na formát digitálního dokumentu. A to bez rozdílu, ať se jedná o textový dokument, grafické či multimediální soubory.

Z důvodu snadnější správy a manipulace s digitálními dokumenty, např. při **migraci** na aktuálnější formát, je vhodné množinu přijímaných formátů omezit. V případě počáteční implementace pro NUŠL byly formáty omezeny u textových souborů na PDF, grafické soubory na JPG a multimediální soubory na MP3 a AVI.

V CDS Invenio se formáty zachovávají a netransformují se na jiné standardní formáty. Není tedy možné bez pomoci jiné aplikace při importu automatizovaně převádět jiné formáty na formát standardní.

Perzistentní identifikátory

Primárními účely digitálních archivů je archivace digitálních informací a jejich zpřístupnění. A právě trvalost (neboli perzistence) přístupu k digitálním dokumentům je základním požadavkem, který musí perzistentní identifikátor splňovat. Problém perzistentních identifikátorů se objevil s digitalizací dokumentů a jejich umístěním na internetu. Tradiční identifikátory, např. LCCN 1898, ISBN 1970, ISSN 1975 - ISMN a ISAN, které se přidělují fyzickým dokumentům, fungují dobře v nedigitálním prostředí. Perzistentní identifikátory by měly splňovat obdobnou úlohu v digitálním světě. V digitálním prostředí je ovšem vztah identifikátor - dokument komplikovanější.

Nejjednodušší způsob identifikace pomocí URL odkazů se ukázal jako nejméně spolehlivý. Webové adresy podléhají časté změně, např. informace, kterou link zpřístupňoval, byla odstraněna, soubor byl přemístěn (např. z důvodu reorganizace webu, změny webové domény, změny vlastníka atd.) a v důsledku toho se objevila potřeba zajistit trvalost odkazu k digitálním dokumentům, což

představuje primární funkci perzistentních identifikátorů. Navíc perzistenci identifikátoru máme na mysli trvalost identifikace bez ohledu na trvalost identifikovaného dokumentu. Je tedy důležité, aby zdroj označený perzistentním identifikátorem nebyl nikdy přesunut nebo odstraněn, aniž by došlo k aktualizaci informací o jeho umístění v registru perzistentních identifikátorů.

Další požadavky na perzistentní identifikátory

Kromě perzistence lze klást na identifikátor další požadavky, které úzce souvisí s virtuálním prostředím, ve kterém se digitální dokumenty vyskytují.

1. Identifikátor musí být jednoznačný. Jednoznačnost identifikátoru, tedy vztah mezi identifikátorem a digitálním dokumentem, je oboustranně jednoznačný v daném systému. Jednoznačnost je zajištěna centrální autoritou přijatého systému. Identifikátory musí být unikátní v rámci digitálních zdrojů organizace, ale mohou být unikátní i celosvětově. Globální jedinečnosti lze dosáhnout, když bude využíván systém, který je používán po celém světě.
2. Identifikátor může mít schopnost pokrýt různé úrovně dokumentů (např. kniha, kniha-kapitola, kniha-obrázek atd.) a tato schopnost se nazývá granularita.
3. Identifikátor může mít schopnost pokrýt různá provedení či verze díla.
4. Syntaxe pro perzistentní identifikátor může reflektovat potřeby dané instituce, např. podřetězce v identifikátoru odkazují na různé podsbírký.
5. Decentralizované přidělování zajišťuje autorita jako správce a další instituce s delegovanými pravomocemi (Bellini, 2007, s. 8).
6. Interoperabilita s dalšími systémy perzistentních identifikátorů.

Příklady perzistentních identifikátorů

Ve světě je rozšířeno několik systémů perzistentních identifikátorů, např. ARK, DOI, NBN a PURL. Tyto systémy byly voleny různými institucemi pro určité účely a splňují tedy různé požadavky. Odpovídající perzistentní identifikátory tedy svoji funkčnost i vlastnosti odvozují z těchto systémů, ve kterých fungují. Například systém Handle s perzistentním identifikátorem DOI používá v České republice např. Knihovna AV ČR. Perzistentní identifikátor URN:NBN používají především národní knihovny Itálie, Německa, Švýcarska, Maďarska, Švédska, Nizozemí atd.

Pro metadatové záznamy v NUŠL byl vybrán perzistentní identifikátor **URN:NBN**, jehož implementaci řeší v České republice NK ČR. Do spuštění URN:NBN v České republice jsou v systému NUŠL přidělovány lokální identifikátory, které půjde lehce rozšířit na plnohodnotné URN:NBN. Syntaxe lokálního identifikátoru

v digitálním repozitáři NUŠL je nusl-číslo, např. nusl-31. Tato hodnota se vkládá do pomocného pole v interním formátu CDS Invenia MARC21. V systému CDS Invenio není potřeba doplňovat nuly před číslo na určitý počet pozic v čísle. Lokální identifikátor je následně používán pro tvorbu trvalého odkazu, ve kterém je jedinou proměnou. Celý trvalý odkaz vypadá následovně:

<http://www.nusl.cz/ntk/nusl-číslo>, např. <http://www.nusl.cz/ntk/nusl-31>, vkládá se do prvku dc:identifier ve formátu NUŠL. Následně byl vytvořen proxy server, který zajišťuje funkčnost trvalého odkazu na identifikátor.

Tato kapitola objasnila softwarové řešení zvolené pro pilotní implementaci NUŠL. Hlavní pozornost byla věnována systému CDS Invenio použitému pro dlouhodobou archivaci a zpřístupnění dat šedé literatury. Byly zmíněny nezbytné standardy pro fungování systému a správu dat - formáty dat a perzistentní identifikátory. Systém ESP FAST umožňuje vyhledávání nejen nad repozitářem NUŠL, ale též nad dalšími repozitáři vědeckých zdrojů obsahujících šedou literaturu v České republice. V době psaní publikace byl připojen katalog NTK, databáze VŠKP VŠE a systém ASEP AV ČR. Průběžně budou do tohoto rozhraní přibývat další zdrojové databáze, tak jak se bude rozšiřovat i partnerská síť spolupracujících institucí. K zajištění správného a dlouhodobého fungování digitálního repozitáře NUŠL byl proveden audit v systému DRAMBORA.

LITERATURA:

PEPE, A. - BARON, T. - GRACCO, M. - Le MEUR, J., Y. - ROBINSON, N. - ŠIMKO, T. - VESELÝ, M. CERN, *Document Server Software: the integrated digital library*. CERN, Geneva, Switzerland, 2005.

BELLINI, E. - CIRINNA, C. - LUNGI, M. - DAMIANI, E. - FUGAZZA, C. *Persistent Identifiers distributed system for Cultural Heritage digital objects*. Dept., University of Milan Via Bramante, Crema, Italy, 2007.

BERNAS, Jiří. *Národní digitální archiv*. Knihovna, 2009, roč. 20, č. 1, s. 22 - 29. Dostupné z [www: http://knihovna.nkp.cz/knihovna91/bernas.htm](http://knihovna.nkp.cz/knihovna91/bernas.htm)

ELEKTRONICKÉ ZDROJE:

Usnesení [online]. Usnesení vlády České republiky ze dne 3. listopadu 2008, č. 1338. 2008. Vláda České republiky [cit. 2010-08-27]. Dostupné z [www: http://racek.vlada.cz/usneseni/usneseni_webtest.nsf/0/22AE7ADE48C55BF8C1257553002951A5/\\$FILE/1338%20uv081103.1338.pdf](http://racek.vlada.cz/usneseni/usneseni_webtest.nsf/0/22AE7ADE48C55BF8C1257553002951A5/$FILE/1338%20uv081103.1338.pdf)

10. Audit NUŠL s použitím nástroje DRAMBORA

Petr Karlach

Audit NUŠL jako **důvěryhodného digitálního úložiště** s použitím nástroje a metodologie *Digital Repository Audit Method Based on Risk Assessment* (**dále jen DRAMBORA**) byl naplánován a proveden v závěru roku 2009 jako nedílná součást vytváření digitálního repozitáře šedé literatury v NTK.

Audit samotný metodologicky navazoval zejména na zkušenosti s auditem digitálního univerzitního repozitáře na Univerzitě Karlově, který používá stejnou metodiku i stejný nástroj, a s auditem digitálního úložiště NK ČR, který byl zpracován s použitím off-line verze nástroje DRAMBORA.

Stejně jako při plánování samotného NUŠL byla i v případě auditu velice důležitá důkladná přípravná fáze. V přípravné fázi se tým připravující audit seznámil s aktuálními trendy a metodami v oblasti plánování, vytváření a certifikace důvěryhodných digitálních úložišť. Dále se tým seznámil se samotným nástrojem DRAMBORA ve verzi on-line¹¹⁹ a s metodikou auditu, který probíhá jako audit vlastními silami bez použití externí specializované auditorské firmy.

Práce s nástrojem také zahrnovala přípravnou fázi a samotný audit, takže příprava na audit se pak přirozeně prolínala s přípravnou prací s nástrojem jako takovým. Tato fáze zahrnovala sběr všech relevantních informací a dokumentů k vlastnímu repozitáři, jako byly jeho status, popis, standardy, procedury, personální obsazení, materiální vybavení, rozpočtové informace apod. (Donnelly et al., 2009). Tyto informace sloužily jako vstupní údaje pro přípravnou fázi auditu a byly vloženy do databáze nástroje DRAMBORA v sekci „Before the Assessment“. Zde bylo popsáno samotné úložiště, definován rozsah auditu a oblasti (Functional Classes - funkční třídy), kterým se audit věnoval, bylo popsáno personální obsazení úložiště včetně popisů rolí a jejich mapování na jednotlivé členy týmu. Definice rolí v rámci personálu úložiště byla zvláště důležitá, protože v další fázi při identifikaci rizik bylo nutné na rizika navázat jejich vlastníky, což se dělo právě prostřednictvím rolí. Už jen přípravná fáze může pro jakékoliv auditované úložiště znamenat veliký přínos. Pomůže celému týmu podívat se na úložiště z jistého komplexního nadhledu, zmapovat a soustředit na jedno místo nejdůležitější popisné údaje o úložišti a případně upozorní na mezery a nedostatky v této oblasti, čímž dá příležitost k jejich nápravě a k doplnění chybějících materiálů.

Audit pak probíhal v části **nástroje DRAMBORA** nazvané „Assessment Centre“ (Donnelly et al., 2009). Zde se definoval **mandát úložiště**, tedy jeho poslání,

¹¹⁹ Více informací zde - <http://www.repositoryaudit.eu>.

důvod existence, zřizovatel apod. Dále byla vložena omezení úložiště, která mají vliv na jeho činnost, a to externí, například legislativní, a interní, například organizační, omezení typu obsahu apod. **Audit** pokračoval definicí **cílů úložiště**, aktivit vykonávaných a prostředků používaných k jejich dosažení. Cíle mají být kvantifikované měřitelné záměry strategického charakteru, například cílový počet záznamů v databázi v určitém období a vztahují se k jednotlivým funkčním třídám. **Aktivity** se mapují na **role** v rámci pracovního týmu úložiště a propojují se s prostředky, které se k nim využívají. Konečnou fází auditu byla **identifikace rizik**, která ohrožují nebo by mohla v budoucnu ohrozit fungování, stabilitu nebo samotnou existenci úložiště.

Pro inspiraci, případně pro přímou pomoc, obsahuje nástroj DRAMBORA řadu odkazů na pomocné texty a celou řadu příkladů jednotlivých vyplňovaných položek z praxe, ať už v přípravné fázi, nebo přímo ve fázi auditu. V oblasti identifikace rizik lze přímo používat předdefinovaná rizika a upravovat je nebo formulovat vlastní. Rizika se mapují na personální role i funkční třídy a lze je propojovat i na **omezení**, **činnosti** a **zdroje**. Po definici rizik se stanovuje zejména pravděpodobnost jejich výskytu a očekávaná míra dopadu na činnost úložiště - nástroj pak stanoví míru jejich závažnosti. Dále se stanovuje strategie, jak s riziky nakládat - předcházet jim, případně řešit jejich následky, a termín kontroly protirizikových opatření.

Po vložení všech potřebných informací lze s využitím funkce Reporting Centre vytvořit **výstupní zprávy** o identifikovaných rizicích repozitáře, jejich závažnosti, provázanosti a jejich řešení. Jsou k dispozici dva druhy výstupní zprávy ve formátu PDF a jeden druh ve formátu HTML. Nelze jednoduše exportovat ostatní vložené popisné informace, nicméně s trochou úsilí je možné je vykopírovat zejména z uloženého snímku aktuálního stavu auditu (Snapshot). Logickým dalším krokem po auditu je vytvoření **akčního plánu** na eliminaci nebo snížení dopadu identifikovaných rizik včetně kontrolních bodů, jejichž termíny audit stanovil. Práce s auditem se tak stane periodicky se opakujícím stálým procesem, který provází život úložiště. Pro prvotně identifikovaná rizika se vypracuje strategie jejich řešení, v kontrolních bodech se rizika znovu vyhodnotí, případně některá zaniknou. Audit je potom třeba opakovat v nově vzniklé situaci a vyhodnotit rizika změněná a nová.

V praxi samotná práce s nástrojem narazila z počátku na jeden nepříjemný problém, kdy se po vyplnění základních údajů zjistilo, že **nástroj DRAMBORA** v on-line verzi nepodporuje jazyky z oblasti kódové stránky ISO-8859-2/Windows-1250, tedy ani češtinu, čímž byly znehodnoceny dosud vložené údaje a dokonce došlo k problémům s přístupovými právy. Překvapivě je celý nástroj na webu napsán s využitím kódové stránky ISO-8859-1, namísto dnes zcela běžného použití UNICODE, např. UTF-8. Autoři nástroje přislíbili nápravu, nicméně to zřejmě nebude rychlý proces, problémem pravděpodobně není jen aplikace samotná, ale i nastavení databáze, do níž se ukládají údaje.

Bylo nutno založit novou databázi pro audit a rozhodnout o použití buď angličtiny, nebo češtiny bez diakritiky. Rozhodli jsme se pro angličtinu z mnoha důvodů, zejména proto, že použití češtiny bez diakritiky je nepřirozené, dále proto, že nástroj i metodika jsou k dispozici pouze v angličtině a že výsledky

auditů se budou prezentovat na mezinárodním poli jak v oblasti projektů šedé literatury, tak patrně i v oblasti auditů DRAMBORA. Existuje i off-line verze nástroje DRAMBORA, která se skládá ze základních jednoduchých formulářů ve formátu Microsoft Office DOC a XLS. Tato verze by umožňovala plnohodnotné použití češtiny, nicméně nenabízí ani zlomek komfortu provázanosti informací a podpory při vyplňování a vyhodnocování údajů jako verze on-line kvůli nimž byl tento nástroj vybrán, proto byla off-line verze vyhodnocena jako v zásadě nepoužitelná.

Po vyřešení těchto počátečních potíží již byla práce se samotným nástrojem v zásadě bez problémů, až na občasně drobnosti, které se však vždy daly obejít. Například při vyplňování rizikových faktorů repozitáře a jejich návazností nebyly vždy k dispozici všechny předem vyplněné údaje, řešením bylo například vrátit se o krok zpět a opakovat akci. Největší problém verze on-line se ukázala nemožnost vytvářet záložní kopie, nejlépe formou exportu, pro případné obnovení poškozené či ztracené databáze. V průběhu práce se vyskytly stavy, kdy vložené informace náhle nebyly k dispozici a databáze vydávala chybová hlášení, nicméně autoři nástroje vždy na vzniklou situaci reagovali relativně rychle.

Nástroj samotný je tedy relativně snadno a intuitivně použitelný, daleko větší problém je kvalitní přípravná fáze. Nestačí sebrat dostupné dokumenty a základní informace, je vhodné připravit si mapu celého auditu a vzájemných vztahů jeho částí. Přestože tato fáze nebyla nijak podceněna, ukázalo se, že přípravy není nikdy dost, občas se stávalo, že v průběhu zadávání informací do databáze bylo nutné se vracet do předchozích kroků a doplňovat chybějící souvislosti. Provázání souvislostí je jeden z hlavních přínosů tohoto nástroje a jak se ukázalo, je zejména v této oblasti velmi kvalitně připraven. Samotná dokumentace NUŠL, popis celého projektu i související dokumenty jsou ovšem na velice dobré úrovni a představovaly velmi dobrý základ pro audit.

Je třeba zdůraznit, že tato přípravná fáze bude často velkým přínosem pro auditované repozitáře, pokud ovšem nebyl repozitář předem zkonstruován a naplánován se striktním využitím metodologie, jako je např. PLATTER a nebyly do něj požadavky auditu přímo předem zakomponovány. Touto cestou se však může vydat málokteré úložiště, zejména proto, že je velice časově i finančně náročná a dlouho se čeká na první použitelné výsledky. Důkladné zmapování úložiště s pomocí **nástroje DRAMBORA** pomůže skutečně získat přehled o tom, co je hotové a co není, jaké významné dokumenty, procedury, prostředky a opatření chybí a kde leží nejvýznamnější rizika, ohrožující jeho existenci a úspěch, jak spolu jednotlivé součásti repozitáře souvisejí, jaká dokumentace je k dispozici apod. K údajům, vkládaným do databáze auditu, lze připojovat externí odkazy formou URL, případně připojovat libovolné dokumenty, takže lze tuto databázi pojmout jako centrum důležitých informací o auditovaném repozitáři.

Nejdůležitějším výstupem kromě zmapovaného repozitáře a jeho relevantního okolí je tedy i podle tvůrců metodiky i nástroje analýza identifikovaných rizik, která ohrožují repozitář, jeho kvalitu, akceschopnost, jeho pověst a pozici u odborné i uživatelské veřejnosti apod. Analýza NUŠL postihuje repozitář ve

fázi pilotního projektu, tedy dostatečně včas, aby se její závěry stihly uplatnit do fáze plného ostrého provozu. Analýza NUŠL byla sice zaměřena na všechny funkční oblasti (Functional Classes), ale v jejich rámci byla analyzována, identifikována a zaznamenána ta nejdůležitější a nejzjevnější rizika podle Paretova pravidla 80/20 - těch 20 procent rizik, která znamenají 80 procent ohrožení. V dalším textu jsou pro ilustraci výstupu z auditu stručně rozebrána prvotně identifikovaná rizika a závěry, které z nich vyplývají. Při čtení rizik je třeba mít na paměti, že se jedná o ilustrativní statický snímek odrážející situaci NUŠL v zásadě již v minulosti.

Rizika

Jak bylo uvedeno výše, audit byl zaměřen na nejdůležitější a nejvýznamnější rizika pro fázi pilotního projektu a jeho nadcházejícího přechodu do ostrého provozu. Tato rizika se týkají zejména stavu popisu činností a procedur úložiště, stavu a rozvoje týmu pracovníků, financování projektu, hardwarových a softwarových zdrojů včetně jejich zálohování a relevantního okolí NUŠL. Bylo identifikováno 16 základních rizik - řazení je nevýznamné, takto rizika očísloval nástroj **DRAMBORA**.

Riziko 1: ztráta členů týmu

Toto riziko se zabývá skutečností, že tým NUŠL je relativně malý, nedávno vytvořený, stěží pokryje všechny související činnosti a už nezbývá prostor na zastupitelnost členů týmu v případě jejich nepřítomnosti nebo odchodu. Takto je činnost NUŠL přímo ohrožena jakýmkoliv pracovním výpadkem. Ideálním řešením by bylo rozšířit tým tak, aby bylo možno sdílet znalosti a zvýšit zastupitelnost, to však nepokrývá rozpočet projektu. Dalším vhodným řešením je doplnit a rozšířit popisy znalostí a činností členů týmu případně i v elektronické podobě, například na základě technologie Wikipedie (elektronická encyklopedie) tak, aby se zastupující nebo nahrazující člen týmu mohl rychleji seznámit se svou novou činností.

Toto riziko souvisí s riziky 6, 12, 15 a volně s rizikem 16.

Eliminace rizika, která proběhla do 30. 6. 2010: Průběžně se rozrůstá projektový tým a tím vzrůstá jeho zastupitelnost. Členové týmu veškerou svoji činnost dokumentují a prezentují na pravidelných schůzkách. Pracovní dokumenty pak ukládají do neveřejné části NUŠL wiki.

Riziko 2: konec pilotního projektu

Jeden z mandátů NUŠL vychází z projektu MK ČR pro léta 2008 až 2011 - Digitální knihovna šedé literatury - funkční model a pilotní realizace. Tento

projekt skončí v roce 2011, což může znamenat oslabení mandátu NUŠL a eventuální finanční potíže. Řešením může být včas se zajímat o účast v nějakém jiném podobném projektu, případně včas hledat jiné finanční zdroje pro pokračování a rozvoj projektu NUŠL.

Toto riziko souvisí s riziky 4 a 6, částečně i 9.

Eliminace rizika, která proběhla do 30. 6. 2010: Funkce dlouhodobé archivace a zpřístupnění dat včetně sběru nových dat jsou a budou i po roce 2011 pokryta z rozpočtu NTK. Na další rozvoj bude nutné zajistit další dotační program.

Riziko 3: plán obnovení úložiště po havárii

Pro NUŠL existuje základní nástin způsobu zálohování databáze, ale není zpracován plán, jak postupovat při obnově po případné havárii databáze nebo po jejím narušení a kdo je za jednotlivé kroky zodpovědný. Toto riziko může znamenat prodloužení času nebo jiné potíže při obnově po havárii. Je nutno takový plán zpracovat včetně určení zodpovědných osob, propojit jej s plánem zálohování, který je nutno také dopracovat a tento plán případně zahrnout do elektronické znalostní encyklopedie NUŠL, zmiňované v riziku 1.

Toto riziko nepřímo souvisí s rizikem 5.

Eliminace rizika, která proběhla do 30. 6. 2010: Současně s metodikou zálohování je řešen i plán obnovení úložiště po havárii. Tento dokument prochází v NTK ověřovacím procesem.

Riziko 4: partnerská síť dobrovolná

Partnerská síť organizací poskytující obsah pro NUŠL je tvořena na základě dobrovolnosti, neexistuje žádný legislativní prostředek pro její vytvoření, vše závisí pouze a jedině na ochotě pracovníků spolupracujících institucí. Toto riziko může znamenat potíže s poskytováním obsahu pro NUŠL. Řešením může být nalezení motivačních faktorů pro spolupracující organizace (např. výhody centrálního úložiště, mezinárodní aspekt projektu apod.), cílená propagace NUŠL zaměřená na spolupracující organizace i na odbornou veřejnost. V neposlední řadě by na politické úrovni pomohlo nalezení motivace srovnatelné s publikováním v odborném tisku, nějaké hodnocení organizací ukládajících šedou literaturu do NUŠL s případným vlivem na jejich financování.

Toto riziko souvisí s riziky 2, 6 a 9.

Eliminace rizika, která proběhla do 30. 6. 2010: Partnerům je nabízena dlouhodobá archivace a lokální instalace CDS Invenia přednastavená pro vytváření sbírky šedé literatury. Průběžně vznikají návody a postupy, jak vytvářet a spravovat sbírky šedé literatury v lokálních úložištích a pravidelně se konají školení pro partnerské organizace. V průběhu roku 2010 probíhá cílená

propagace NUŠL v odborném tisku, na konferencích, zvaných přednáškách a na osobních schůzkách se zástupci producentů šedé literatury.

Riziko 5: ukládání záložních pásek

Záložní pásky, na které se zapisují záložní kopie databáze NUŠL, jsou skladovány ve stejné místnosti s vlastním hardwarovým úložištěm. V případě havárie typu požár, teroristický útok nebo jiné závažné události by byla zničena databáze i její zálohy. V současné době testovacího provozu, kdy databáze obsahuje minimum záznamů, je toto riziko relativně nevýznamné, nicméně ve fázi rutinního provozu je nutné využívat existující alternativní úložiště a tuto skutečnost zahrnout do přepracovaného plánu zálohování.

Toto riziko nepřímo souvisí s rizikem 3.

Eliminace rizika, která proběhla do 30. 6. 2010: viz riziko 3.

Riziko 6: podfinancování

Současný rozpočet a finanční zdroje pokrývají projekt v jeho stávající podobě, nezajišťují však další rozvoj NUŠL v žádné z jeho relevantních oblastí. Žádná instituce, projekt nebo podnik, které se nerozvíjejí a nerostou, nepřetrvají delší dobu. Řešení je stejné jako u rizika 2, včas se zajímat o účast v nějakém jiném podobném projektu, případně včas hledat jiné finanční zdroje pro pokračování a rozvoj projektu NUŠL.

Toto riziko souvisí s velkou většinou ostatních rizik.

Eliminace rizika, která proběhla do 30. 6. 2010: Průběžně je zajišťováno další financování projektu. Například zajištění licence systému ESP FAST a plán obnovy hardwaru.

Riziko 7: rozpočet na služby

Služby podpory pro softwarový systém ESP FAST, podporující uživatelské rozhraní NUŠL a poskytované externím dodavatelem, jsou rozpočtově a smluvně pokryty do konce roku 2010. Je třeba při tvorbě rozpočtu NTK a NUŠL na rok 2011 nezapomenout na tuto položku a včas zahájit i smluvní jednání o podmínkách pokračování spolupráce s externím dodavatelem.

Toto riziko souvisí s riziky 2 a 6.

Eliminace rizika, která proběhla do 30. 6. 2010: Pokračování spolupráce s externím dodavatelem systému ESP FAST je již smluvně pokryto do konce roku 2013.

Riziko 8: ESP FAST ve zkušební verzi

Do konce roku 2011, na základě smlouvy s externím dodavatelem, provozuje NTK softwarový systém ESP FAST podporující uživatelské rozhraní NUŠL ve zkušební verzi. Je třeba, nejlépe již v průběhu roku 2010, nastartovat jednání o podmínkách nákupu, provozu a podpory ostré verze pro rutinní provoz NUŠL a zahrnout nákup licencí a jejich podporu do příslušného rozpočtu.

Toto riziko souvisí s riziky 6 a 7.

Eliminace rizika, která proběhla do 30. 6. 2010: Toto riziko bylo odstraněno zakoupením plné verze.

Riziko 9: slabý mandát

Mandát NUŠL je založen na časově ohraničeném projektu MK ČR a na stručné zmínce ve statutu NTK. Tato skutečnost může negativně ovlivňovat pověst a pozici NUŠL v očích odborné veřejnosti a spolupracujících organizací u nás i ve světě. Tuto skutečnost je vhodné řešit na politické rovině, například snahou o získání formální podpory ze strany NK ČR, AV ČR, propagací na úrovni souvisejících ministerstev apod.

Toto riziko souvisí s riziky 2, 4 a částečně 6.

Eliminace rizika, která proběhla do 30. 6. 2010: Problematika šedé literatury byla zahrnuta do návrhu Koncepce rozvoje knihoven v České republice do roku 2014. Neustále probíhá snaha o řešení tohoto rizika na úrovni jednání vedení NTK s příslušnými institucemi a orgány státní správy.

Riziko 10: není formální procedura pro akvizici

Neexistuje formální popsaná procedura a metodika pro postupy, jak zacházet se získávanými dokumenty, jak je zpracovávat, jakými kontrolami musí projít před uložením do databáze, kdo je za tyto kroky zodpovědný apod. Toto riziko se může projevit nejednotným postupem při ukládání získaných dokumentů, případně jejich nedostatečným zpracováním a kontrolou a tím vést k jistému stupni nekonzistence ukládaných dat. Je nutné takovou metodiku zpracovat včetně určení zodpovědných osob a zahrnout ji případně do elektronické znalostní encyklopedie NUŠL, zmiňované v riziku 1.

Toto riziko volně souvisí s riziky 11, 13 a 14.

Eliminace rizika, která proběhla do 30. 6. 2010: Postupně vznikají metodiky pro získávání a zpracování dokumentů šedé literatury. Současně byla v rámci týmu vytvořena role správce obsahu.

Riziko 11: formáty dokumentů

Přes dobře zpracovanou typologii dokumentů NUŠL nejsou plně popsány formáty dokumentů, které lze ukládat do databáze NUŠL. Nejsou také popsány formalizované procedury a postupy pro případnou migraci formátů na novější verze, získávání nových verzí softwaru potřebných pro zpřístupňování těchto nových verzí apod. Toto riziko by mohlo vést k nekonzistenci uložených dokumentů a k problémům s jejich zpřístupňováním v delším časovém horizontu.

Je třeba vytvořit plán a popsanou metodiku, jaké formáty dokumentů bude NUŠL podporovat a jak bude zajištěna jejich migrace na případně vyšší verze včetně potřebného softwarového vybavení pro jejich zpřístupňování.

Toto riziko volně souvisí s riziky 10, 13 a 14.

Eliminace rizika, která proběhla do 30. 6. 2010: Toto riziko se řeší zadáním externí analýzy této problematiky.

Riziko 12: administrace softwaru

Administrace softwarového systému pro digitální úložiště NUŠL - CDS Invenio - závisí v NTK na jediné osobě, která navíc není oficiálně přímo řízena manažerem NUŠL, je zařazena v oddělení IT. V případě jeho výpadku neexistuje žádná přijatelně rychlá možnost, jak jej nahradit. Tím by byl vážně narušen provoz NUŠL.

Řešením může být přijmout nebo zaškolit dalšího člověka a zvýšit tak jejich vzájemnou zastupitelnost, nebo nahradit tuto činnost službami externí společnosti. Druhé řešení se jeví jako vhodnější, nicméně by v konečném důsledku mohlo být dražší - nezapomenout na tuto případnou možnost při tvorbě rozpočtu.

Toto riziko souvisí s riziky 1 a 6 a volně s rizikem 16.

Eliminace rizika, která proběhla do 30. 6. 2010: Toto riziko bylo výrazně sníženo převedením administrátora přímo pod manažera NUŠL, zaškolením dalšího pracovníka a vytvořením manuálů pro administrátora.

Riziko 13: nedokumentované činnosti a postupy

Přestože všeobecně je dokumentace projektu NUŠL na velice dobré úrovni, audit našel několik konkrétních míst, která nejsou dokumentována dostatečně nebo vůbec neexistuje jejich popis. Audit se zaměřil na nejpodstatnější skutečnosti a rizika pro fázi pilotního projektu. Po skončení auditu je potřeba

se věnovat obecně dokumentaci všech podstatných procesů, identifikovat nepokryté oblasti a vytvořit časový plán dotvoření relevantní dokumentace s případným vytvořením elektronické encyklopedie NUŠL na bázi technologie Wikipedie - viz riziko 1.

Toto riziko souvisí s riziky 1, 10, 11 a 14.

Eliminace rizika, která proběhla do 30. 6. 2010: Toto riziko je eliminováno průběžně tvorbou dokumentace nepokrytých procesů.

Riziko 14: nepopsaná dlouhodobá strategie uchování elektronických dokumentů

Dlouhodobá strategie uchování elektronických dokumentů je jednou z nejdůležitějších položek provozu a rozvoje důvěryhodného digitálního úložiště.

Tato strategie řeší komplexně životaschopnost digitálního úložiště po stránce jak technické (např. migrace formátů dokumentů), tak po stránce fungování týmu pracovníků a celé organizace, po stránce financování provozu a rozvoje úložiště atd. Zárodky této strategie jsou obsaženy v dokumentaci pilotního projektu NUŠL, nicméně je nutno tuto strategii zpracovat jako samostatný dokument.

Toto riziko souvisí s riziky 10, 11 a 13.

Eliminace rizika, která proběhla do 30. 6. 2010: Toto riziko se řeší zadáním externí analýzy této problematiky společně s rizikem 11.

Riziko 15: nedostatečné znalosti a schopnosti členů týmu

S postupujícím rozvojem technologií a metod tvorby a provozu digitálních knihoven může nastat situace, že znalosti a schopnosti týmu neudrží krok s tímto rozvojem. Takové riziko hrozí každému týmu, zejména ovšem týmu malému a plně vytíženému s fakticky neexistující vzájemnou zastupitelností jeho členů, kde nezbyvá mnoho na času na školení a vzdělávání. Management NUŠL by neměl tento problém podceňovat a pokusit se zajistit prostor a prostředky pro pravidelná školení a vzdělávání svých kolegů i sebe sama - se samozřejmými implikacemi pro plánování počtu členů týmu a pro rozpočet NUŠL.

Toto riziko souvisí s riziky 1, 6 a 16.

Eliminace rizika, která proběhla do 30. 6. 2010: Průběžně se rozrůstá projektový tým a tím vzrůstá jeho zastupitelnost. Probíhá průběžné vzdělávání členů jak samostudiem, tak účastí na odborných konferencích a školeních.

Riziko 16: úmyslná sabotáž

Riziko úmyslné sabotáže systému nespokojenými zaměstnanci či zaměstnanci ve výpovědi, narušení databáze, dlouhodobé narušování konzistence dat apod., taková rizika nelze zcela vyloučit nikde a nikdy. Podobně jako u rizika 15 hrozí zejména u malého a plně vytíženého týmu s fakticky neexistující vzájemnou zastupitelností jeho členů a s jedním členem fakticky vyčleněným do jiného oddělení při současné neexistenci přísně procesního řízení. Management NUŠL by měl dbát na pozitivní motivaci členů týmu, na jejich profesní rozvoj a na druhé straně implementovat fungující zálohovací a kontrolní mechanismy, například periodickou kontrolu konzistence dat v databázi NUŠL.

Toto riziko souvisí s riziky 1, 6, 12 a 15.

Eliminace rizika, která proběhla do 30. 6. 2010: Riziko je eliminováno každodenními zálohami dat i systému a vytvořením role v rámci projektového týmu správce obsahu. V rámci posílení týmu dochází k překrývání činností a tím vzrůstá zastupitelnost.

Po dokončení auditu a identifikaci rizik byl vytvořen akční plán, jehož cílem bylo věnovat se jednotlivým rizikům a eliminovat nebo zmírňovat jejich dopad na činnost NUŠL. Jak již bylo uvedeno, výše zmíněný seznam rizik představuje ilustrativní statický snímek z minulosti, protože mnohá rizika již byla v rámci akčního plánu eliminována nebo se s jejich řešením započalo. Tato skutečnost je uvedena vždy u jednotlivých rizik v seznamu. Dalším krokem bude opakování auditu v nových podmínkách, identifikace nových či změněných rizik, vytvoření dalšího akčního plánu atd. Tím se audit stane iterativním procesem opakované identifikace rizik a jejich řešení a přispěje tak k důvěryhodnosti NUŠL.

LITERATURA:

DONNELLY, Martin - INNOCENTI, Perla - McHUGH, Andrew - RUUSALEPP, Ravio. *DRAMBORA Interactive User Guide*. Glasgow, 2009. Dostupné z [www: http://www.repositoryaudit.eu/help/](http://www.repositoryaudit.eu/help/)

Závěr

Se změnou způsobu publikování, které se posouvá od tradičního tisku k elektronickému publikování, vzniká mnoho dokumentů, které spadají do oblasti šedé literatury. Produkce informací a jejich šíření je díky novým technologiím mnohem jednodušší a umožňuje v masové míře publikační aktivity institucím, jejichž základní činností není publikační činnost. Takovou publikační činnost lze pak na všech úrovních podchytit prostřednictvím digitálních repozitářů šedé literatury. To je stěžejní téma, které provází celou tuto knihu.

První část knihy představuje šedou literaturu od obecného vymezení přes její význam v Evropě a České republice až po popis stavu sběru šedé literatury v institucích. Zde se opakuje jedno významné téma, na které je kladen důraz, kterým je spolupráce od lokální, národní až po mezinárodní úroveň. Druhá část knihy je orientovaná prakticky a přináší návod, jak budovat digitální repozitáře obecně s aspekty vlastními šedé literatuře. Snaží se odpovědět na otázky typu co je potřeba pro vybudování repozitáře, jaký software použít, co by mělo být v popisných údajích, nebo jak sbírat, archivovat a šířit šedou literaturu v souladu s českým právem. Poznátky zde uvedené vycházejí ze zkušeností nabytých při řešení projektu Národního úložiště šedé literatury.

Na základě těchto poznatků se bude tým v Národní technické knihovně i nadále zabývat výzkumem v oblasti šedé literatury a digitálních repozitářů, bude podporovat odbornou diskusi a šířit znalosti v oblasti šedé literatury současně s budováním digitálního Národního úložiště šedé literatury. Důraz bude stále kladen na spolupráci jak na úrovni národní, tak na mezinárodní. V rámci projektu se bude tým snažit o přístup k více zdrojům z oblasti Open Access v České republice a podpořit tak sdílení informačních zdrojů v oblasti vědy, výzkumu a vývoje.

Resume

The potential of grey literature has recently attracted increasing attention of specialists working in the areas of education, research, development and information. This publication aims to enlighten the scientific community on grey literature and show a possible method for building a digital repository of grey literature. The scientific publication *Repositories of Grey Literature* is among the first publications of its kind in either the Czech Republic or abroad.

This book introduces the topic of grey literature, its history, types, development and impact on the scientific public. It is also devoted to grey literature producers, who are rather diverse compared to the producers of classical „white” literature. Examples of mandatory policies are stated, as well as strategies for acquiring grey literature in Europe. Outstanding grey literature collections, both printed and electronic, are presented. There is highlighted the importance of cooperation for the acquisition, storage and accessibility of grey literature at all levels, from departmental to national and international collections. It depicts the role both historically and presently of the National Technical Library in collecting grey literature in the Czech Republic.

Cooperation with producers is a key activity for repository construction. Although grey literature is a precious source of information, it is rather difficult to capture using classical means of collection. Contacts, partnership and networking are of immense importance. The legal view of grey literature is of crucial significance. This book is written with a view to Czech law, which has specific characteristics compared to the laws of other countries. There are debated legal recommendations for managers of grey literature databases and options for exploiting free licenses. University final theses are such a distinctive type of grey literature, from a legal standpoint as well, that an entire chapter is devoted to them.

The example of the National Repository of Grey Literature, practically implemented in the National Technical Library, is given to illustrate how such a repository may be built. There are described descriptive elements, so-called metadata; the analysis of the most common Open Source systems, these being CDS Invenio, DSpace, EPrints, Fedora and Greenstone; the software solution for the pilot implementation, which is based on CDS Invenio for the digital repository and the indexing and search system ESP FAST for user interface. For reinsurance credibility of the digital repository is recommended to perform the audit on the issue of its trustworthiness and to repeat this process on a regular basis to establish and achieve quantifiable targets.

Disclosure of grey literature through digital repositories is the most effective form how to provide it to research and scientific community.

Příloha č. 1: Databáze šedé literatury v České republice

Producent	URL	Popis
Akademie věd České republiky	http://www.avcr.cz	Internetový portál zveřejňující současné vědecké informace pro vědeckou komunitu a širokou veřejnost.
Asociace knihoven vysokých škol České republiky	http://www.akvs.cz	Obhazuje a podporuje zájmy univerzit a jejich knihoven při jednání s organizacemi v oblasti vzdělávacích, informačních a komunikačních technologií.
Digitální knihovna Univerzity Pardubice	http://dspace.upce.cz/	Digitální knihovna UPa je vytvořena v systému DSpace a je primárně určena k ukládání a zpřístupňování eVŠKP.
Digitální univerzitní repozitář Univerzity Karlovy v Praze	http://repozitar.cuni.cz/	Digitální repozitář UK budovaný v systému DigiTool.
DSpace MU	http://dspace.muni.cz/	Digitální knihovna MU.
DSpace - repozitář AMU	http://dspace.amu.cz/	Institucionální repozitář AMU, fungující od roku 2008.
DSpace - úložiště VŠKP UJEP	http://edu.ujep.cz/main.php?idwebu=dspace	Digitální repozitář UJEP pro publikování eVŠKP.
DSpace VŠB-TUO	http://dspace.vsb.cz/	Digitální repozitář VŠB-TUO, který buduje a spravuje Ústřední knihovna VŠB-TUO.
Národní Registr VSKP	http://www.theses.cz	Národní registr diplomových a disertačních prací (Česká republika).
Národní úložiště šedé literatury	http://www.nusl.cz	Centrální vyhledávací rozhraní Národního úložiště šedé literatury zpřístupňuje šedou literaturu jak z digitálního repozitáře NUŠL, tak i z dalších zdrojů.
Rejstřík informací o výsledcích	http://www.vyzkum.cz	RIV je jednou z částí IS VaV, kde jsou shromažďovány informace o výsledcích projektů výzkumu a vývoje a výzkumných záměrů podporovaných z veřejných prostředků dle zákona o podpoře výzkumu a vývoje.

Příloha č. 2: Databáze šedé literatury v zahraničí

Databáze	URL	Popis
ANRT	http://www.anrtheses.com.fr/	Ve Francii vyvíjí asociace ANRT aktivitu s cílem digitalizace v rámci své služby Thèses à la carte.
ArXiv.org	http://www.arxiv.org	Preprinty (fyzika, matematika, informatika, nelineární věda).
CEDA Repository	http://cedadocs.badc.rl.ac.uk/	Digitální depozitář je věnován šedé literatuře o pozorování země a atmosférickým vědám. Obsahuje více než 600 digitálních dokumentů.
CNR	http://www.cnr.it/sitocnr/home.html	Institucionální repozitář italské Národní rady pro výzkum (CNR).
DART-Europe	http://www.dart-europe.eu/	Portál eVŠKP zpřístupňuje více než 130 000 plnotextových výzkumných prací z 233 vysokých škol z 16 evropských zemí. DART-Europe je partnerem výzkumných knihoven a knihovnických konsorcií pro evropskou pracovní skupinu NDLTd.
DOAJ (adresář otevřených časopisů)	http://www.doaj.org	Vědecké a odborné časopisy (plné texty).
DOE	http://www.energy.gov/	Jeden z největších amerických digitalizačních projektů, který je zaměřený na sbírky zpráv.
DFG	http://gepris.dfg.de	Německé digitalizační projekty financované DFG zahrnují především primární zdroje kulturního dědictví, ale rovněž odborné publikace (rukopisy, časopisy atd.).
DiVA	http://www.diva-portal.org/	Portál zpřístupňuje 270 000 výzkumných publikací a studentských prací z 27 švédských a norských vysokých škol a institucí vyššího vzdělání, u něhož 44% obsahu tvoří šedá literatura.
DRIVER	http://www.driver-community.eu/	Projekt usiluje o vytvoření evropské sítě digitálních repozitářů nabízejících sofistikované funkční služby jak rešeršérům, tak veřejnosti. Je nadstavbou nad existujícími institucionálními repozitáři a sítěmi.

EKT	http://www.ekt.gr/en/index.html	EKT v Athénách spravuje databázi Hellenic Dissertations odkazující na 13 000 prací z řeckých vysokých škol.
ETH	http://www.ethz.ch/	Knihovnický vysokoškolský portál zprostředkovává přístup ke 2,1 milionům zpráv z dalších knihoven, k databázím a vyhledávacím nástrojům.
EthOS - portal	http://ethos.bl.uk/	Britská knihovna digitalizuje práce z britských vysokých škol pro tento portál, který spojuje možnost volného přístupu k eVŠKP z otevřených repozitářů a dodávku prací digitalizovaných na vyžádání.
Grey Literature LibGuide (Univerzita v Calgary)	http://libguides.ucalgary.ca/graylit	Kompilace dokumentů šedé literatury v oblasti zdravotnictví a v medicínských oborech.
GreyNet	http://www.greynet.org	Archiv GreyText: interní archiv dokumentů šedé literatury.
JISC	http://www.jisc.ac.uk/digitisation	Britský JISC digitalizuje obsah zvláštních sbírek, např. historické zprávy o sčítání lidu či diplomové práce.
LARA	http://lara.inist.fr/	Platforma umožňující přístup k plným textům francouzských vědeckých zpráv z oblasti humanitních i sociálních věd, obsahuje téměř 200 zpráv.
NARCIS	http://www.narcis.info/index	Vědecký portál NARCIS vyvinula nizozemská KNAW a uživatelé najdou informace o výzkumu, programech, projektech, výzkumných pracovnících a institucích spolu s jejich profily.
NDLTD	http://www.ndltd.org/	Mezinárodní organizace věnující se podpoře přijetí, vytváření, používání, šíření a uchovávání eVŠKP, která zajišťuje přístup k téměř 750 000 pracím.
NTIS Technické zprávy	http://www.fedworld.gov	Věda, technologie a obchod.
NUMDAM	http://www.numdam.org/	Matematický archiv s 29 semináři za období 1948 až 2007.
OpenDOAR	http://www.opendoar.org	Adresář akademických otevřených repozitářů.

OpenSigle	http://opensigle.inist.fr	Volný přístup k bibliografickým záznamům šedé literatury v Evropě.
OSTI	http://www.osti.gov/	Vědecký portál pro šedou literaturu typu technických zpráv. Volně přístupné dokumenty přes on-line rozhraní ze zdrojů DTIC, DOE, NASA, EPA atd.
ProQuest (Digitální disertace)	Předplacená služba	Databáze diplomových prací ze severoamerických univerzit.
PUMA	http://puma.isti.cnr.it/	Systém pro zpřístupnění sítě institucionálních repozitářů v Itálii. V současné době zpřístupňuje 40 institucionálních repozitářů prostřednictvím uživatelského rozhraní v italském nebo anglickém jazyce.
ROAR (Registr otevřených repozitářů)	http://roar.eprints.org	Otevřené repozitáře, kurzy e-printu, vzdělávací materiály.
SCIRUS	http://www.scirus.com	Nabídka prohledávání obsahu vědeckých časopisů, vědeckých webových stránek, preprintů, patentů a institucionálních repozitářů.
TEL	http://tel.archives-ouvertes.fr/	Francouzský národní repozitář eVŠKP s více než 10 000 pracemi z let 2005 - 2010 a s téměř 2 000 digitalizovanými pracemi publikovanými do roku 1990.
TDX	http://www.tesisenxarxa.net/	Katalánská síť elektronických diplomových a disertačních prací.
Web 2.0 Ernesta Barrera	http://www.netvibes.com/ernestobarrera#Grey_literature	Osobní stránka se soubory RSS věnovaná sdílení interních znalostí o šedé literatuře z oblasti biomedicíny.

Příloha č. 3: VŠKP jako archiválie

Naskytá se otázka, zda VŠKP jsou archiválií a pokud ano, jaký by měly mít skartační znak. Práce bakalářské, diplomové, rigorózní a disertační jsou školními díly dle **autorského zákona**, vytvořenými pod vedením školy za účelem splnění studijních povinností. Mohou být také považovány za součást protokolu o státní závěrečné zkoušce studenta, který bezesporu archiválií je (Grulich, 2005).

Odpověď na položenou otázku můžeme nalézt ve vzorových **skartačních rádech**, které jednotlivým dokumentům přiřazují odpovídající skartační znaky:

A - dokumenty s trvalou dokumentární nebo informační hodnotou určené do trvalé úschovy v archivu,

V - dokumenty, které jsou po uplynutí skartační lhůty znovu v rámci skartačního řízení posouzeny a zařazeny do kategorie dokumentů typu „A“ nebo „S“,

S - dokumenty, u nichž z hlediska dokumentárního nebo informačního není po splnění jejich provozní a správní funkce nutná další úschova a mohou být v rámci skartačního řízení vyřazeny a skartovány.

Číselné označení uváděné za skartačním znakem udává skartační lhůtu v letech, po jejímž uplynutí se dokument ve skartačním řízení navrhne k předání do příslušného archivu (skartační znak „A“) nebo ke zničení (skartační znak „S“), anebo k posouzení, má-li se písemnost předat do archivu nebo má-li se zničit (skartační znak „V“).

Podle výnosu ministerstva školství č. j. 11834/57 ze dne 16. 7. 1958 a č. j. 10324/60-L ze dne 1. 4. 1960 byly diplomové a jim podobné práce považovány za archiválie se skartačním znakem A10. Aktualizací ve výnosu *Ministerstva školství ČSR* č. j. 19 151/87-491 z roku 1987 byly nově diplomové práce uvedeny se skartačním znakem V20, disertační a habilitační práce se skartačním znakem A20. Podle předpisu z roku 1987 měl být kritériem výběru diplomových prací ve skartačním řízení přínos vědě a výzkumu. Tento výběr však nebyl na školách prakticky prováděn. Seznamy platných předpisů v resortu školství, mládeže a tělovýchovy vydávané po roce 2005 tento předpis již přestaly uvádět.

Nahlédnutím do skartačních řádů jednotlivých škol (viz tabulka Přehled skartačních znaků VŠKP podle skartačních řádů vybraných vysokých škol) zjistíme, že současná praxe v archivování VŠKP se velmi liší. Společným prvkem, který můžeme vysledovat, je význam přikládán pracím habilitačním, které jsou uváděny se skartačním znakem A5 nebo A10, výjimečně i A40 (VŠE). U prací bakalářských a diplomových je často použit skartační znak V20, jak odpovídá výnosu Ministerstva školství ČSR č. j. 19 151/87-491 z roku 1987. Tabulka tak dokládá různou váhu, kterou přikládají jednotlivé školy pracím vznikajícím na nižším stupni studia (bakalářské, diplomové).

Různorodost praxe dokládá např. MU, která zmiňuje pouze Studijní spisy studentů se skartačním znakem A10, nebo *Univerzita Palackého v Olomouci* (dále jen UPOL), která neobvykle zvolila pro práce bakalářské a diplomové ve svém skartačním řádu znak S10, který tyto práce umožňuje vyřadit a skartovat po deseti letech bez provedení skartačního výběru. MENDELU zvolila extrémně nízkou dobu archivace, a to pouhých pět let bez ohledu na typ VŠKP.

VŠKP	VŠE	1. lékařská fakulta UK	UTB	UPOL	MENDELU
bakalářské	V20		V20	S10	V5
diplomové	V20	A10	V20	S10	V5
doktorské	V20	A10	A5	A	A5
habilitační	A40	A10	A5	A	A5

Přehled skartačních znaků VŠKP podle skartačních řádů vybraných vysokých škol

Příloha č. 4: Výsledky průzkumů zveřejňování eVŠKP z let 2006 až 2009

V prosinci 2006 byl Komisí eVŠKP proveden první průzkum zpřístupňování eVŠKP, ze kterého plyne, že elektronické verze textových prací sbíralo již před platností novely zákona o vysokých školách deset škol, z toho 6 povinně a 4 výběrově či volitelně (Horová a Krkošková, 2007). Průzkum tak prokázal vysokou akceptaci metadatového standardu EVSKP-MS, jehož využití v roce 2006 deklarovalo 14 respondentů dotazníku. V opakovaném průzkumu o rok později plánovalo tento jednotný formát využívat 16 respondentů. Zbylí respondenti používali buď vlastní formát, nebo formát MARC/UNIMARC. Ohledně povinnosti sběru v dotazníku za rok 2007 bylo zjištěno, že sběr probíhá již na 21 školách, z toho jednou v testovacím režimu, ve dvou školách výběrově či volitelně. Pouze dvě školy - a šest fakult UK a *Slezská univerzita v Opavě* - uvedly, že sběr se dosud vůbec neprovádí. Jednoznačně převažovalo povinné odevzdávání (Horová, 2007).

Poslední provedený průzkum v roce 2009 zachycuje stav zpřístupňování VŠKP na 16 vysokých školách po spuštění systému theses.cz (Vyčítalová, 2010). Z odpovědí vyplývá pokles vyžadování uzavření **licenčních smluv** (13 respondentů smlouvy neuzavírá, 6 uzavírá a 2 plánovaly). Knihovny ukládají eVŠKP často v knihovních katalozích, proto došlo k nárůstu volby formátu **MARC** (25 % respondentů) v odpovědi na dotaz na používaný formát popisu. Formát EVSKP-MS zvolilo oproti roku 2006 pouhých 25 % respondentů. Lze ale předpokládat, že v případě potřeby může většina škol zajistit export dat i ve formátu EVSKP-MS, neboť mají potřebné údaje pro metadata ve svých systémech.

Pouze TU provádí **retrospektivní digitalizaci** VŠKP s plánovaným záběrem až do roku 1979. Tři univerzity (AMU, JAMU, VUT) zpracovávají i netextové VŠKP (filmy, audiovizuální záznamy, fotografie apod.). Do **Rejstříku informací o výsledcích** (dále jen **RIV**) vykazuje šest škol disertační práce vzniklé v souvislosti s projekty sledovanými Radou vlády, jedna vysoká škola eviduje i práce habilitační.

Velké procento škol nahrává data VŠKP do systému theses.cz ručně (38 %) nebo dávkově (47 %) a pouze 15 % respondentů využívá sklizení dat automatizovaně protokolem **OAI-PMH**.

SEZNAM ZKRATEK

A

AKVŠ - Asociace knihoven vysokých škol

AMU - Akademie múzických umění v Praze

ANRT - L'Association Nationale de la Recherche et de la Technologie; Národní asociace pro výzkum a technologie

ANSI - American National Standards Institute; Americký národní úřad pro normalizaci

aut. - autor, autorský

ASEP - Automatizovaný systém evidence publikací

AV ČR - Akademie věd České republiky

B

BMBF - Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung); Spolkové ministerstvo pro vzdělání a vývoj

BRGM - Bureau de recherches géologiques et minières; francouzská veřejná agentura pro výzkum a odborné znalosti věd o zemi

C

CDS - Cern Document Server

CEDA - The Centre for Environmental Data Archival; Centrum pro archivaci údajů o životním prostředí

CNR - Consiglio Nazionale delle Ricerche; Italská národní rada pro výzkum

cit. - citovaný

Č

č. - číslo

ČDK - Česká digitální knihovna

ČVUT - České vysoké učení technické v Praze

D

DART-Europe - Portál eVŠKP

DCAM - Abstract Model

DCMI - Abstract Model

DFG - Deutsche Forschungsgemeinschaft

DIDL - Digital Item Declaration Language

DiVA - Academic Archive On-line

DOC - přípona souboru textových dokumentů programu Microsoft Word

DOE - Department of Energy

DRAMBORA - Digital Repository Audit Method Based on Risk Assessment

DRIVER - Digital Repository Infrastructure Vision for European Research

DSc. - vědecký titul „doktor věd“

E

EAGLE - European Association for Grey Literature Exploitation

ed. - editor, edičně připravil

EKT - National Documentation Centre; Národní dokumentační centrum v Athénách

et al. - a jiní, a kolektiv

ETD - Electronic theses and dissertations; Elektronické diplomové a disertační práce

ETD db - ETD databáze

ETD-MS - standard pro popis eVŠKP

ETH - Die Eidgenössische Technische Hochschule Zürich; knihovnický portál vysoké školy

EthOS-portal - Electronic Theses Online Service

eVŠKP - elektronické vysokoškolské kvalifikační práce

EVSKP-MS - metadatový soubor pro eVŠKP v České republice

F

FIZ Karlsruhe - Fachinformationszentrum Karlsruhe

FRBR - Functional Requirements for Bibliographic Records; funkční požadavky na bibliografické záznamy

FRVŠ - Fond rozvoje vysokých škol Ministerstva školství, mládeže a tělovýchovy České republiky

G

GA AV - Grantová agentura Akademie věd

GA ČR - Grantová agentura České republiky

GL - International Conference on Grey Literature

GreyNet - mezinárodní organizace pro šedou literaturu Grey Literature Network Service

I

ICC - Katalánský kartografický institut v Barceloně

IGA MZ - Interní grantová agentura Ministerstva zdravotnictví České republiky

INIST (CNRS) - L'Institut de l'Information Scientifique et Technique (Centre National de la Recherche Scientifique)

ISO - International Standard Organization; Mezinárodní organizace pro standardizaci

IS VaV - Informační systém výzkumu a vývoje

J

JAMU - Janáčkova akademie múzických umění v Brně

JISC - Joint Information Systems Committee

K

KIV - Knihovnictví a informační věda

KNAW - Royal Netherlands Academy of Arts and Science (Koninklijke Nederlandse Akademie van Wetenschappen); Královská holandská akademie umění a vědy

Knihovna AV ČR - Knihovna Akademie věd České republiky

kol. - kolektiv, kolektivní

Komise eVŠKP - Odborná komise pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací AKVŠ

KSSL - Kooperační systém šedé literatury

L

LANL - Los Alamos National Laboratory

LARA - Libre accès aux rapports scientifiques et techniques; volný přístup pro vědecké a technické zprávy z různých institucí

LIBER - Ligue des Bibliothèques Européennes de Recherche; nevládní organizace sloužící k podpoře vědeckých knihoven

LIGRIA - Littérature Grise Administrative

M

MENDELU - Mendelova zemědělská a lesnická univerzita v Brně

METS - Metadata Encoding and Transmission Standard

MK ČR - Ministerstvo kultury České republiky

MMVS - Mezinárodní meziknihovní výpůjční služba

MODS - Metadata Object Description Schema

MŠMT - Ministerstvo školství, mládeže a tělovýchovy České republiky

MU - Masarykova univerzita v Brně

N

NARCIS - National Academic Research and Collaborations Information System; vědecký portál, který vyvinula KNAW a kde uživatelé najdou informace o výzkumu, programech, projektech, výzkumných pracovnících a institucích spolu s jejich profily

NATO - North Atlantic Treaty Organization; Severoatlantická aliance

NDK - Národní digitální knihovna

NDLTD - Networked Digital Library of Theses and Dissertations

NIWI - Institut vědeckých informačních služeb

NK ČR - Národní knihovna České republiky

NTIS - National Technical Information Service (USA); Národní technické informační služby

NTK - Národní technická knihovna, do 1. 7. 2009 Státní technická knihovna (STK)

NUMDAM - Numérisation de documents anciens mathématiques; volný přístup k metadatům a novinovým článkům v matematice, Francie

NUŠL - Národní úložiště šedé literatury

O

OA - Open Access; Otevřený přístup

OAI-PMH - Open Archives Initiative Protocol for Metadata Harvesting

OAIS - Open Archival Information System
 OCCL - Online Computer Library Center
 OECD - Organization for Economic Cooperation Development; Organizace pro ekonomickou spolupráci a rozvoj
 OPAC - On-line Public Access Catalog
 OpenDOAR - The Directory of Open Access Repository - projekt na Univerzitě v Nottinghamu (UK)
 OpenSIGLE - System for Information on Grey Literature in Europe; od roku 2005 otevřený přístup k bibliografickým záznamům v SIGLE
 OSTI - Office of Scientific a Technical Information; Vědecký portál pro šedou literaturu typu technických zpráv
 OVF - Open Virtualization Format
 OU - Ostravská univerzita v Ostravě

P

PID - Perzistentní identifikátor
 PDF - Adobe Portable Document Format
 PPSNC DLT - Poznań Supercomputing and Networking Center Digital Libraries Team v Poznani (Polsko)
 PSH - Polytematický strukturovaný heslář
 příl. - příloha
 PUMA - Akademisches Publikationsmanagement

R

RBAC - Role Based Access control
 RDA - Resource Description and Access; katalogizační pravidla
 RDF - Resource Description Framework; obecný rámec pro popis jakéhokoli elektronického zdroje, resp. webové stránky a jejího obsahu, tedy pro vyjádření sémantiky a pro podporu sémantického webu
 roč. - ročník
 RSS - Really Simple Syndication; soubor s příponou .xml, je psán v jazyce XML

S

SGML - Standard Generalized Markup Language
 SIGLE - System for Information on Grey Literature in Europe
 SKOS - Simple Knowledge Organization System
 STAG - Informační systém studijní agendy
 STFC - Scientific a Technology Facilities Council; Vědecká rada technologických zařízení, Velká Británie
 SUDOC - Système Universitaire de Documentation; francouzský národní katalog
 SURF - Southeastern Universities Research Association

T

TDKIV - Česká terminologická databáze knihovnictví a informační vědy
 TDX - Tesis Doctorals en Xarxa; diplomové a disertační práce Online, Španělsko
 TEL - Theses en Ligne; digitální archiv francouzských disertačních prací
 theses.cz - Národní registr VŠKP a systém na odhalování plagiátů

TIB - German National Library of Science and Technology
(Technische Informationsbibliothek Universitätsbibliothek Hannover); Německá
národní knihovna vědy a techniky
TU Delft - Delft University of Technology, Holandsko
TUL - Technická univerzita v Liberci

U

UJEP - Univerzita Jana Evangelisty Purkyně v Ústí nad Labem
UNESCO - United Nations Educational, Scientific and Cultural Organization);
Organizace OSN pro výchovu, vědu a kulturu
UPa - Univerzita Pardubice
ÚOOÚ - Úřad pro ochranu osobních údajů
UTB - Univerzita Tomáše Bati ve Zlíně

V

VaV - Výzkum a vývoj v České republice
VNTIC - Moskevské vědeckotechnické informační centrum
Virginia Tech - Virginia Polytechnic Institute and State University
VISK - Veřejné informační služby knihoven
viz - lze vidět
VŠB-TUO - Vysoká škola báňská - Technická univerzita Ostrava
VŠE - Vysoká škola ekonomická v Praze
VŠKP - vysokoškolské kvalifikační práce
VUT - Vysoké učení technické v Brně
vyd. - vydání, vydáný; vydavatel

W

www - informační systém pro práci s hypertextovými dokumenty na Internetu
(angl. World Wide Web, „celosvětová pavučina“)

X

XML - eXtensible Markup Language; rozšiřitelný značkovací jazyk

Z

ZČU - Západočeská univerzita v Plzni

REJSTŘÍK

A

absence bibliografické kontroly: 15
komerčního zdroje: 14
recenzního řízení: 14
administrátorská práva: 44
Akademie věd České republiky: 14
akční plán: 124
aktivity: 124
AKVŠ: 34
AMU: 49
antičtí vědci: 12
Apache: 117
archiv, osobní: 51
archivace, dlouhodobá: 45
ASEP: 34, 38, 46, 79, 82, 122
Asociace knihoven
vysokých škol České republiky: 14
audit: 123
autonomie vůle: 73
autor: 69
autorská práce: 50
práva: 68
majetková: 69
autorství: 68
AV ČR: 14, 33, 38, 44, 45, 46

B

bakalářské práce: 48
bezpečnost: 12
BibCheck: 113
BibConvert: 113
BibClassify: 113
BibEdit: 113
BibFormat: 113
BibHarvest: 111
BibIndex: 113
bibliografie, studium: 13
BibRank: 113
BibSched: 114
bulletiny: 13

C

CDS Invenio: 46, 82, 98, 107, 109

centrální vyhledávací prostředí: 50
CNRI Handle System: 101
CorpCZ: 59

Č

ČDK: 37
činnosti: 124
vojenské: 12
články: 13

D

DART Europe: 58, 61
databáze: 48, 95
kvalifikačních prací: 72
pořizovatel: 71
šedé literatury: 75
detailnost: 14
digitalizace, retrospektivní: 141
digitální
knihovny: 14
Národní archiv: 116
repozitáře: 14
úložiště: 93, 123
DigiTool: 49, 60
dílo literární: 66
společné: 17
školní: 48, 71
zaměstnanecké: 73
diplomová práce: 13, 48
distribuce, omezená: 14
distribuční systém: 11
dočasnost: 11
dokumenty, politické: 16
dostupnost: 14, 51
DRAMBORA: 123, 124, 125, 126
DRIVER: 39
DSpace: 20, 39, 49, 60, 101
Dublin Core: 59, 79, 80

E

EAGLE: 19, 33
efekty, síťové: 69
elektronický přístup: 13
ElmSubmit: 112

emulace: 115
e-print: 47
EPrints: 102
eScience: 27, 28
ETD db: 60
ETD MS: 58, 60, 81
EVŠKP-MS: 58, 59, 60, 61, 79, 112
eVŠKP: 23, 27, 35, 49, 60
expertiza, právní: 43, 48

F

Facebook: 13
FAST ESP: 108
federace: 14
Fedora: 103
FIZ Karlsruhe: 33
flexibilita: 14
formát, metadatový: 44
 neprofesionální: 15
formáty
 dokumentů: 115, 120
 MARC: 34
 rozličnost: 14
funkční specifikace: 95

G

GA AV: 50
GA ČR: 50
grantové agentury: 47, 50
graue Literatur: 11
Greenstone: 104
grey literature: 11
Grey Literature Network Service: 14
GreyNet: 14, 29, 36, 39

H

habilitační práce: 48
Handle: 83
harvesting: 96, 111
harvestování: 111
herecké umění: 16
humanitní a sociální vědy: 17

CH

charakter, krátkodobý: 16
 mezioborový: 15
charakteristika: 12

I

identifikace rizik: 124
identifikátory, perzistentní: 120
indexování: 96
informace, popisné: 15
informační společnost: 14
INIST: 20, 21, 39
Internet: 14
institute, výzkumné: 14
Institute for Scientific and Technical
Information of the French National
Center for Scientific Research: 20

J

JAMU: 49

K

klíčová slova: 95
knihovna AVČR: 33, 45, 46
knihovny, digitální: 14
 vysokoškolské: 57
komerční zdroj: 14
komise eVŠKP: 48, 59
konferenční materiály: 47, 50
konverze metadat: 95
krátkodobost: 16
kritéria: 93, 97
KSŠL: 33

L

Letteratura grigia: 11
LibGuide: 14
licence, exkluzivní: 70, 75
 školní: 56
 volná: 69
 výhradní: 47
 zákonná: 69
licenční smlouva: 43, 56, 141
literární dílo: 66
literatura
 polopublikovaná: 47
 šedá: 11, 12, 14
 budoucí možnosti: 17
 charakteristika: 12
 databáze: 76
 definice: 11
 historie a vývoj: 12

konference: 14
kořeny: 12
producenti: 14, 44
rysy a význam: 14
typy: 13
v oborech: 16

literature grise: 11
lístky: 13
lokální repozitář: 44

M

mandát úložiště: 123
MARC 21: 81, 141
MARCXML: 112
materiály
 konferenční: 47, 50
 nekonvenční: 11
 pomíjivé: 13
 studijní: 50
MENDELU: 49
metadata: 95
metadatový formát NUŠL: 44
mezioborový charakter: 15
migrace: 115, 120
ministerstva: 51
MIT Dspace: 80
MU: 49
MySQL: 118

N

náklad, omezený: 15
navigace: 118
NARCIS: 26, 27
Národní digitální archiv: 116
národní registr: 60
NASA: 12
NDK: 37
NDLTD: 58, 60
neformálnost: 11
nehmotnost: 11
nekonvenčnost: 11
neprofesionální formát: 15
nízkonákladová vydání: 11
NUŠL: 33, 35, 43, 61
 metadatový formát: 44

O

OaiArchive: 115
OAI-PMH: 20, 39, 44, 59, 61, 93, 99, 115, 141
obchodní tajemství: 73
OCLC: 79
odborové organizace: 14
odevzdej.cz: 62
odpovědnost: 70
omezená distribuce: 14
omezený náklad: 15
omezení: 124
OPAC: 60
Open Access: 14
OpenDOAR: 24
OpenSIGLE: 14, 20, 39, 80
 věcné třídění: 88
Open Source: 93, 98, 101
osobní archiv: 51
 údaje: 73
OU: 49

P

paragraf 47 b): 56
partnerská síť: 43, 50
PDF: 58, 60, 61
PersCZ: 59
perzistentní identifikátory: 120
plagiátorství: 57, 61
plagiátoři: 62
plakáty: 13
podniky: 14, 51
pojetí: 12
politické dokumenty: 16
polopublikovaná literatura: 47
popisné informace: 15
pořizovatel databáze: 71
postery: 47
postprint: 47
práce
 autorské: 50, 68
 bakalářské: 48
 diplomové: 13, 48
 disertační: 13, 46, 48
 habilitační: 48
 rigorózní: 48
 seminární: 62

právní expertiza: 43, 48
 právo: 66
 autorské: 69
 preprint: 47, 50
 prezentace: 47
 prospekty: 13
 předpisy, vysokoškolské: 56, 59
 překlady: 13
 případové studie: 16
 PSH: 84, 88
 Python: 118
 publikace: 13
 rychlé: 14
 publikovat
 částečně: 12
 neformálně: 11

R

radý: 14
 recenzní řízení: 14
 RefExtract: 113
 referáty: 47, 50
 rejstřík informací o výsledcích: 79, 141
 Reports literature: 12
 repozitář, digitální: 14
 lokální: 44
 retrospektivní digitalizace: 142
 RIV: 79, 141
 rizika: 126
 identifikace: 124
 role: 124
 rozhraní, uživatelské: 95
 rozličnost formátů: 14
 rozvrhy: 13
 ruční vyhledávání: 13

Ř

řád, skartační: 57, 139

S

sbírky: 114
 sborníky: 13, 47
 sdělování díla veřejnosti: 57
 sdružení: 14
 Second Life: 13
 seminární práce: 62

Seminář ke zpřístupňování šedé literatury: 45
 SGML: 58
 SIGLE: 19, 33
 síť, partnerská: 43
 síťové efekty: 69
 skartační řád: 57, 140
 softwar, výběr: 93
 softwarové řešení: 97
 systémy: 98
 soubory formulářů: 16
 současné povědomí: 13
 soudní zápisy: 16
 soukromí vydavatelé: 14
 společné dílo: 17
 společnost: 14
 informační: 14

STAG: 48

studie, případové: 16
 studium bibliografie: 13
 systémy, distribuční: 11
 softwarové: 98
 systematicčnost: 12

Š

šedá literatura: 11, 12, 13, 14, 16, 17, 44, 75
 šedá zóna: 44
 školní dílo: 48, 71
 licence: 56
 školy, vysoké: 47

T

textové filtry: 118
 theses.cz: 33, 60, 61
 tištěná podoba: 13
 trusty: 14
 třídění, věcné OpenSIGLE: 88
 TUL: 50
 Twitter: 13

U

údaje, osobní: 72
 UJEP: 49
 uložení, dlouhodobé: 57
 úložiště
 cíle: 124

digitální: 93, 123
mandát: 123
Univerzita Karlova: 49
univerzity: 14
ÚOOÚ: 45
UPa: 49
URN:NBN: 83, 121
utajení: 12
utajované skutečnosti: 73
UTB: 49
užití, volné: 69
uživatelské rozhraní: 95

V

věcné třídění OpenSIGLE: 88
VirtualBox: 107, 117
vládní zprávy: 13
volně dostupný: 14
volné licence: 69
 užití: 69
vojenské činnosti: 12
VŠB-TUO: 49
VŠE: 35
VŠKP: 46, 48, 50, 55, 68
VUT: 49
výběr softwaru: 93, 97
vydání, nízkonákladová: 11
vydavatelé, soukromí: 14
vyhledávání, ruční: 13
výhradní licence: 47
výroční zprávy: 46
vysoké školy: 48
vysokoškolské knihovny: 57
 předpisy: 56, 59
výstupní zprávy: 124
výzkumné instituce: 14
 zprávy: 46

W

Web 2.0: 118
WebAccess: 116
WebSearch: 114
WebSubmit: 110
WorldWideScience.org: 39

X

XMetadiss: 59

Z

zákon č. 216/2006 Sb.: 56
 č. 552/2005 Sb.: 56
 autorský: 139
 o vysokých školách: 55
zákonná licence: 69
zaměstnanec: 74
zaměstnavatel: 74
zápisy, soudní: 16
ZČU: 49
zdroje: 124
 komerční: 14
zkreslení: 12
zprávy: 50
 komisí: 16
 o ukončených projektech: 46
 vládní: 13
 výroční: 46
 výstupní: 124
 výzkumné: 46
 z projektů: 50
 závěrečné: 50

Repozitáře šedé literatury

Petra Pejšová (ed.), Marcus Vaska, Joachim Schöpfel,
Iveta Fürstová, Radim Polčák, Jan Mach,
Bohdana Frantíková, Petr Karlach, Jindřich Dolanský

Vydavatel: Radim Bačuvčík - VerBuM
(Přehradní 292, 763 14 Zlín 12, Česká republika)

Odpovědný redaktor: Marek Adamík
Zlín, 2010

1. vydání. 152 stran. Náklad: 200 ks
Tisk: Kodiak print, s.r.o., Zlín

www.verbum.name
www.verbum.webnode.cz

ISBN 978-80-904273-5-8