

# Projekt NUŠL a další projekty v ČR

## Projekt „Digitální knihovna pro šedou literaturu – funkční model a pilotní realizace“

Petra Pejšová

Státní technická knihovna

[p.pejsova@stk.cz](mailto:p.pejsova@stk.cz)

Idea Národního úložiště šedé literatury (dále NUŠL) vznikla ve [Státní technické knihovně](#)<sup>1</sup> (dále STK) již v roce 2005. Impulsem bylo ukončení evropského kooperačního systému SIGLE (System for Information on Grey Literature in Europe), ve kterém byla STK aktivním přispěvatelem a národním integrátorem. V letošním roce podpořilo projekt pod názvem „Digitální knihovna pro šedou literaturu – funkční model a pilotní realizace“ Ministerstvo kultury České republiky v rámci programových projektů výzkumu a vývoje. Konkrétně se jedná o druhý program zaměřený na zpřístupnění a ochranu kulturních, uměleckých a vědeckých zdrojů.

Projekt „Digitální knihovna pro šedou literaturu – funkční model a pilotní realizace“ (dále jen projekt) je čtyřletý (2008 - 2011). Řešiteli projektu jsou STK a [Vysoká škola ekonomická v Praze](#)<sup>2</sup> (dále VŠE).

Projektový tým tvoří za STK PhDr. Petra Pejšová, PhDr. Martina Pfeiferová a PhDr. Radka Římanová. K 30. 4. 2008 z STK odešel Mgr. Petr Novák, který se vizí projektu zabýval od jeho počátku. Za VŠE se projektu účastní Ing. Jan Mach, který je současně spolureditelem projektu [Národního registru vysokoškolských kvalifikačních prací](#)<sup>3</sup>. Ing. Jan Mach osobní účastí v obou projektech zajišťuje jejich kompatibilitu. Externě na projektu spolupracuje PhDr. Iva Horová z Akademie múzických umění v Praze a externí odborníci na konkrétní úkoly.

Harmonogram projektu je rozdělen do tří etap. První etapa (duben 2008 – červen 2009), která právě probíhá, se zabývá přípravou zadávací dokumentace pro modelovou aplikaci. Tato příprava obsahuje specifikaci metadat, výběr vhodných persistentních identifikátorů, analýzu typologie dokumentů, specifikaci softwaru pro modelovou aplikaci, příprava licenčních smluv pro spolupráci s producenty šedé literatury, dále vytvoření webových stránek projektu. V rámci druhé etapy (červenec 2009 – září 2010) proběhne realizace modelové aplikace a její ověření. Třetí etapa (červenec 2010 – prosinec 2011) se zaměří na testovací provoz a na sestavení standardů, doporučení a metod ověřených na provozu modelové aplikace, které budou zpřístupněny jak na webových stránkách projektu, tak vydány formou odborné publikace. V průběhu všech tří etap bude probíhat doplňování informací na webových stránkách projektu, každoročně bude pořádán

---

<sup>1</sup> <http://www.stk.cz/cs/>

<sup>2</sup> <http://www.vse.cz/>

<sup>3</sup> <http://theses.cz/>

„Seminář ke zpřístupňování šedé literatury“ a seminář „Systémy pro zpřístupňování e-VŠKP“ a prezentace výsledků odborné veřejnosti.

Výstupem projektu bude funkční pilotní aplikace, která bude sloužit pro budování Národního repozitáře šedé literatury. Na základě ověřené technologie a metod definovaných v rámci projektu budou dalšími výstupy projektu odborná publikace a webové stránky projektu. V těchto výstupech budou obsaženy zkušenosti z projektu, doporučení, standardy a metodiky, zejména doporučený metadatový formát, výměnné formáty a šablony; vzory licenčních modelů a řešená právní problematika; metodika ochrany, archivace a zpřístupňování digitálních dat; webové rozhraní pro spolupráci s producenty šedé literatury. Tyto výstupy budou publikovány v české a anglické verzi. Výstupem z projektu bude také každoroční pořádání seminářů k problematice šedé literatury a prezentace výsledků projektu odborné veřejnosti.

Jaký je vztah NUŠL k ostatním projektům v České republice zabývajícím se ukládáním digitálních dat? Národní digitální knihovna (dále NDK) jako jádro České digitální knihovny se orientuje na publikované dokumenty, NUŠL bude naopak podchycovat dokumenty oficiálně nepublikované. V základním funkčním schématu [koncepte NDK](#)<sup>4</sup> se NUŠL nachází v tzv. standardní zóně, která předpokládá zastřešení prostřednictvím portálů a souborných katalogů digitálních dokumentů. NUŠL bude jedním z nich. Samotný NUŠL bude přitom sklízet metadata a plné texty z lokálních úložišť oborových, institucionálních a dalších zdrojů v šedé zóně. Zodpovědnost za tato úložiště včetně finanční ponosou jednotlivé instituce a jejich zřizovatelé jako jsou resortně příslušná ministerstva.

Dalším projektem, který se v České republice zabývá ukládáním digitálních dat, je projekt [WebArchiv](#)<sup>5</sup>. WebArchiv je digitální archiv „českých“ webových zdrojů, které jsou shromážděny za účelem jejich dlouhodobého uchování. Jelikož předpokládáme, že většina primárních dokumentů v NUŠL bude získávána na základě licenčních smluv, NUŠL nebude otevřeným zdrojem pro projekt WebArchiv. Tento postoj se může změnit v případě volně dostupných zdrojů v NUŠL nebo v případě změn v legislativě České republiky.

Od počátku řešení projektu počítáme s Národním registrem vysokoškolských kvalifikačních prací jako zdrojem pro NUŠL. Vysokoškolské kvalifikační práce (dále jako VŠKP) představují významnou část šedé literatury pro NUŠL. Uvědomujeme si, že spolupráce jak s Národním registrem VŠKP, tak s jednotlivými vysokými školami, musíme smluvně ošetřit, proto v rámci řešení projektu připravujeme licenční smlouvy pro spolupráci. Projekt zohledňuje normativní dokumenty [Odborné komise pro elektronické zpřístupňování vysokoškolských kvalifikačních prací](#)<sup>6</sup>. V současné chvíli je metadatový formát e-VŠKP zapracováván do metadatového formátu NUŠL.

V rámci řešení projektu jsme již iniciovali spolupráci se zahraničními projekty. Prvním z nich je systém pro šedou literaturu v Evropě [OpenSigue](#)<sup>7</sup> (System for Information on Grey Literature in Europe) dříve SIGLE. Více o tomto systému a pozadí jeho transformace se můžete dočíst v článku [„Nový online přístup k databázi SIGLE \(projekt OpenSIGLE\)“](#)<sup>8</sup>. Při jednáních o spolupráci na projektu s Knihovnou Akademie věd (dále jen KNAV) bylo dohodnuto, že KNAV

<sup>4</sup> Koncepte trvalého uchování knihovnických sbírek tradičních a elektronických dokumentů v knihovnách ČR.

Etapa 1: do roku 2010 viz <http://www.ndk.cz/koncepce/koncepctext.pdf>

<sup>5</sup> <http://www.webarchiv.cz/ccinfo>

<sup>6</sup> <http://www.evskp.cz/>

<sup>7</sup> <http://opensigle.inist.fr/>

<sup>8</sup> [Bratková, Eva](#). Nový online přístup k databázi SIGLE (projekt OpenSIGLE). *Ikaros* [online]. 2007, roč. 11, č. 10 [cit. 2008-10-23]. Dostupný na World Wide Web: <<http://www.ikaros.cz/node/4346>>. URN-NBN:cz-ik4346. ISSN 1212-5075.

bude dodávat šedou literaturu do OpenSigle přes NUŠL. Dalším naším záměrem je zařadit NUŠL do evropské sítě digitálních repozitářů [DRIVER](#)<sup>9</sup> (Digital Repository Infrastructure Vision for European Research), který nyní představuje vyhledávací portál zahrnující 70 repozitářů typu „open access“. Více o projektu DRIVER se můžete dočíst v příspěvku [Grey Literature ve světě](#).<sup>10</sup>

Při zpracování zadání projektu jsme zjistili potřebu získat informace a data potřebná pro realizaci projektu. Proto jsme provedli dotazníkové šetření, které probíhalo od prosince 2007 do května 2008. Z producentů šedé literatury jsme se v tomto dotazníkovém šetření zaměřili na výzkumné ústavy Akademie věd a veřejné vysoké školy. Celkem jsme oslovili 77 institucí. Hlavním účelem dotazníkového šetření bylo zjistit poptávku po systému sjednocujícím šedou literaturu a zajištění kontaktů na jednotlivých pracovištích. Dále jsme zjišťovali způsob evidence, shromažďování a zpřístupnění šedé literatury v oslovených institucích. Náš dotazník zodpovědělo 47 z oslovených institucí což je 61%. 37 tj. 79% institucí odpovědělo, že považuje za výhodný přístup k šedé literatuře v ČR z jednoho místa. 39 institucí se vyjádřilo, že je ochotno spolupracovat s NUŠL, ale záleží na licenčních podmínkách, které budou pro tuto spolupráci ujednány. Státní technická knihovna tyto instituce osloví v souladu s harmonogramem projektu v druhé polovině roku 2009. [Podrobné výsledky dotazníkového šetření](#) naleznete na webových stránkách.

V rámci řešení projektu jsme vytvořili webové stránky, které naleznete na adrese <http://nusl.stk.cz/>. Webové stránky jsou vytvořeny v aplikaci Media wiki a s použitím grafického stylu [WordPress](#) modifikovaného podle grafického manuálu STK. Aplikace Media wiki podporuje interaktivní rozhraní, které v našem případě představuje sdílená tvorba obsahu v rámci pracovní skupiny a otevřené diskusní skupiny směrem k veřejnosti. Na webových stránkách naleznete informace k projektu, které neustále doplňujeme a otevíráme diskuze k jednotlivým článkům. Dále zde naleznete informace k šedé literatuře z ČR i ze světa, k legislativě spojené s šedou literaturou v ČR, odkazy na portály podporující výzkum a vývoj ve vědě v ČR, on-line sborník ze „Semináře ke zpřístupňování šedé literatury“ a informační zdroje šedé literatury.

Informační zdroje šedé literatury jsou součástí oborové brány [TECH](#), která umožňuje jejich prohledávání. Oborová brána TECH zpřístupňuje zdroje z oblasti vědy, techniky a medicíny. Zdroje tvoří převážně databáze nakupované ve Státní technické knihovně, dále vybrané české i zahraniční zdroje volně dostupné z internetu a katalogy největších českých technicky zaměřených knihoven. Seznam informačních zdrojů šedé literatury prohledávaných přes oborovou bránu TECH je dostupný z <http://tech.jib.cz/informacni-zdroje/profesionalni-zdroje/seda-literatura>. Hlavním záměrem oborové brány [TECH](#) bylo poskytnout uživateli, který pracuje v oblasti techniky a přírodních a aplikovaných věd, jednotné rozhraní pro vyhledávání dokumentů a orientaci v oboru. Oborová brána TECH je zaměřena na tři typy uživatelů: studentům v oboru techniky, dále pak odborným pracovníkům a naposled knihovníkům, kterým může být pomocníkem při tvorbě rešerší. Registrovat se do oborové brány lze jak heslem čtenáře Státní technické knihovny, čímž je zaručen vstup do placených bází, tak i heslem uživatele Jednotné informační brány. V obou případech uživatel získává svůj vlastní prostor k ukládání vyhledaných výsledků, oblíbených zdrojů atd.

V souvislosti s řešením projektu zakoupila STK v letošním roce datové úložiště. Jeho specifikace je následující: server SUNXFIRE 4500, operační systém SOLARIS 10, kapacita úložiště 16 TB = 42 pevných disků o velikosti 500GB a zálohování je řešeno páskovou knihovnou = 24 pásek, 1 páska 1,6 TB komprimovaného formátu.

---

<sup>9</sup> <http://www.driver-community.eu/>

<sup>10</sup> [http://nusl.stk.cz/images/GL\\_ve\\_svete.pdf](http://nusl.stk.cz/images/GL_ve_svete.pdf)

Pro NUŠL definujeme vlastní metadatový formát. Základními požadavky na tento formát jsou jednoduchost, minimum povinných polí, respektování Dublin Core, možnost stahování metadat pomocí OAI-PMH a možnost vytváření základních rejstříků. Výchozími formáty pro metadatový formát NUŠL jsou metadatové formáty hlavních kooperujících registrů, metadatový standart e-VŠKP a systému [Evidence publikační činnosti v Akademii věd České republiky](#)<sup>11</sup> (známý jako ASEP = Automatizovaný systém evidence publikací). Dále metadatový formát [Rejstříku informací o výsledcích](#)<sup>12</sup> (dále jen RIV), který zpracováváme ne z důvodu, že by bylo možné metadata přebírat přímo z RIVu, ale abychom usnadnili práci lokálním přispěvatelům, kteří již vytvoří záznam pro RIV, budou pak moci poskytnout totožný záznam pro NUŠL. V souvislosti s přípravou na předávání dat do systému OpenSigle připravujeme metadatový formát NUŠL tak, aby byl kompatibilní s metadatovým formátem OpenSigle. Při definování metadatového formátu NUŠL zohledňujeme mezinárodní metadatové standardy [Dublin Core](#)<sup>13</sup> a [MODS](#)<sup>14</sup>. V současné chvíli řešíme také hierarchické vztahy v záznamech a jak nakládat s různými identifikátory.

O projektu budeme informovat v příspěvku „Grey Literature in the Czech Republic“ na 10. mezinárodní konferenci o šedé literatuře ([10th International Conference on Grey Literature](#)<sup>15</sup>), 8. prosince 2008 v Amsterdamu.

Veškeré informace naleznete na webových stránkách projektu <http://nusl.stk.cz/>, kde s námi můžete diskutovat nebo nám můžete napsat na e-mail [nusl@stk.cz](mailto:nusl@stk.cz).

---

<sup>11</sup> [http://www.iach.cz/knav/databaze\\_cz.htm](http://www.iach.cz/knav/databaze_cz.htm)

<sup>12</sup> <http://www.vyzkum.cz/FrontClanek.aspx?idsecke=956>

<sup>13</sup> <http://dublincore.org/documents/dcmi-terms/>

<sup>14</sup> <http://www.loc.gov/standards/mods/mods-overview.html>

<sup>15</sup> <http://www.textrelease.com/gl10program.html>