

DIGITALIZACE S GOOGLE A JEJÍ VZTAH K NDK – IOP

MIROSLAVA HEJNOVÁ – ZDENĚK UHLÍŘ

Abstrakt: Národní knihovna České republiky zahájila v roce 2011 spolupráci se společností Google na hromadné digitalizaci historických a vzácných fondů, která potrvá do roku 2016 s možností následného prodloužení vždy o jeden rok a jejímž výsledkem bude zpřístupnění digitálních kopií historických knižních fondů do roku 1800, jakož i některých vzácných fondů mladších s přidáním hodnotou plnotextového vyhledávání. Koncoví uživatelé na celém světě tak získají přímý přístup ke zhruba 150 tisícům svazků historických a vzácných fondů jednak prostřednictvím Google Books, jednak v rámci Manuscriptoria, případně Krameria. To nutí jak Národní knihovnu České republiky a v důsledku toho i další knihovny k poskytování speciálních nadstavbových služeb poskytujících dodatečnou informaci přidáním hodnotu prostřednictvím informačních a komunikačních technologií a metodologií založených na tzv. *culturomics*, resp. *digital humanities*.

Klíčová slova: Hromadná digitalizace, historické knižní fondy, písemné a dokumentární dědictví, analytický popis, materiálová evidence.

V posledních letech se výrazně zvyšuje důraz na přechod do digitálního informačního komunikačního prostředí, zrychluje se digitální konverze tradičních dokumentů nejenom tištěných, ale i rukopisných. Proces digitalizace zasahuje nejen do oblasti živé kultury, ale i kultury mrtvé, tj. písemného a dokumentárního dědictví. A tak se kromě masové digitalizace novodobých knižních fondů a moderních administrativních, účetních a firemních písemností prosazuje i systematická a hromadná digitalizace staršího, středověkého a raně novověkého knihovního a archivního materiálu.

Vedle národních podprogramů *VISK6*, *VISK7* a evropských integrovaných operačních programů se objevují projekty zaměřené na spolupráci s globálními korporacemi (PPP – public – private partnership) při digitalizaci písemného a dokumentárního dědictví, mezi nimiž dominují v menší míře ProQuest v *Early European Books* (1450–1700)¹ a v mnohem větší míře Google v rámci monumentálního počínu *Google Books*.²

Masová digitalizace prováděná Googlem je přijímána ambivalentně. Akademické prostředí zakládající si na antikomerčním a antikapitalistickém étosu, na ni hledí takřka bez výhrady negativně. Prý neodpovídá náročným odborným a vědeckým požadavkům a ke všemu je

1 Viz např. *Early European Books*, dostupné z URL <http://www.proquest.com/en-US/catalogs/databases/detail/eeb.shtml> [cit. 2012-07-08]; KIBBLE, Matt: ProQuest's Early European Books Project: Collaborative Approach to the Digitization of Rare Texts. In: *LIBER Quarterly*. 2011, roč. 20, č. 3–4, s. 372–381, dostupné z URL <http://liber.library.uu.nl/index.php/lq/article/view/8003> [cit. 2012-07-08].

2 Viz např. *Google Books Library Project*, dostupné z URL: <http://www.google.com/googlebooks/library.html> [cit. 2012-07-08]; MAGÁN, J. A. – PALAFOX, M. – TARDÓN, E. – SANZ, A. Mass Digitization at the Complutense University Library: Access to and Preservation its Cultural Heritage. In: *LIBER Quarterly*. 2011, roč. 21, č. 1, s. 48–68, dostupné z URL: <http://liber.library.uu.nl/index.php/lq/article/view/8007> [cit. 2012-07-08]; KAISER, M. Putting 600 000 Books Online: the Large-Scale Digitization Partnership between the Austrian National Library and Google. In: *LIBER Quarterly*. 2012, roč. 21, č. 2, s. 213–225, dostupné z URL: <http://liber.library.uu.nl/index.php/lq/article/view/8020> [cit. 2012-07-08]; HEJNOVÁ, M. Digitalizace starých a vzácných tisků v Národní knihovně společnosti Google. In: *Duha: informace o knihovnách z Moravy*. 2011, roč. 25, č. 3, dostupné z URL: <http://duha.mzk.cz/clanky/digitalizace-starych-vzacnych-tisku-v-narodni-knihovne-spolocnosti-google> [cit. 2012-07-08].

prováděna pro zisk, ačkoli zpřístupnění nejširší globální veřejnosti včetně služeb tzv. *culturo-mics* je bezplatné.

Naopak knihovny, zejména národní a univerzitní, které tvoří páteřní síť vědecké a odborné infrastruktury, na ni hledí mnohem pozitivněji či přinejmenším pragmaticky. Uvědomují si, že svými silami (tj. organizačními a vizionářskými schopnostmi i finančními možnostmi veřejného sektoru) by nebyly schopny dosáhnout ani zlomku výkonu společnosti Google, a to nejenom na tomto poli. Toto tvrzení se týká obzvláště digitalizace staršího písemného dokumentárního dědictví, kde je ve srovnání s moderním materiálem nezbytnou nutností důraz na vyšší kvalitu, a tudíž je jak po technické, tak následně po finanční i personální stránce mnohem náročnější.

Je zřejmé, že bez podstatného podílu společnosti Google by přechod do digitálního síťového a následně virtuálního prostředí byl výrazně pomalejší a mohl by vést k zaostávání a tím i k diskvalifikaci celé kulturní sféry všeobecně a humanitních disciplín zvláště.

Digitalizace staršího písemného a dokumentárního dědictví ve spolupráci se soukromými subjekty se však zakládá na dosti rozdílných předpokladech, neboť vychází z výrazně jiných principů.

Zatímco společnost ProQuest vychází z tradičního knihovního provozu založeného na striktně vymezených institucích (máme na mysli jednotlivé knihovny, ale i státy, v jejichž hranicích se nacházejí) a jež jsou garanty nejenom financování, ale také právního řádu, v tomto případě především autorských práv, resp. copyrightu). Knihy zdigitalizované společností ProQuest tak zhusta nejsou přístupné globálně a všeobecně, nýbrž jen ve vymezených institucionálních hranicích. Tento způsob digitalizace tak v žádném případě nevytváří plnohodnotné síťové prostředí a také není krokem k virtuálnímu badatelskému prostředí, které je podle dnešních představ konečným cílem digitalizace. Nové technologie tak vlastně petrifikují staré informační a komunikační prostředí, takže se nabízí otázka, zda se z dlouhodobého hlediska jedná o perspektivní investice.

Naproti tomu je společnost Google v digitalizačních aktivitách mnohem ambicióznější. Při digitalizaci staršího písemného a dokumentárního dědictví spolupracuje sice také s knihovnami, protože jinak by své cíle ani nemohla uskutečnit, nicméně vztah obou stran je mnohem volnější. Google na jedné straně zpřístupní zdigitalizované knihy v rámci svého zdroje Google Books, jakož i své služby Google Book Search, avšak každá spolupracující knihovna může zároveň nejen uchovat vytvořené komplexní digitální dokumenty pro archivní účely, ale také je podle svého uvážení zpřístupnit ve svých vlastních zdrojích a začlenit je do svých služeb. Přitom není zásadně vyloučena ani agregace dalšími zdroji a integrace do jiných služeb. Pro Google tedy není žádná ze spolupracujících knihoven *hortus conclusus*, naopak jejich spolupráce představuje cestu k vytvoření otevřeného virtuálního prostředí se specializovanými službami pro jednotlivé obory a skupiny uživatelů a nakonec i se službami personalizovanými. Tento otevřený druh spolupráce se oproti striktně institucionálnímu jeví jako výhodnější, protože zvyšuje uživatelský komfort a zároveň rozmnožuje bohatství ideálních hodnot ve společnosti a v návaznosti na tom i hodnot materiálních.

Při novátorských ambiciózních aktivitách se zcela přirozeně vyskytují obtíže. Ty však není nutno chápat jako brzdu a už vůbec ne jako zásadní překážku počínání, ale prostě jako výzvu: odpovědí na ni by mělo být jiné chápání věcí, a v důsledku toho i jiné podoby činností či vůbec činnosti jiné, nové.

Při masivní digitalizaci s cílem nejenom vytvořit síťové virtuální prostředí, ale nahradit jím dosavadní prostředí tištěné, se vynořila i otázka autorského práva, copyrightu. Dosavadní autorské právo totiž vychází z distribuce obvyklé ve fyzickém, analogovém prostředí, v němž

je omezený počet kopií, exemplářů, které nemohou být jednoduše kopírovány a multiplikovány, a tedy není z autorskoprávního hlediska jejich koloběh od jednoho vlastníka k druhému problémem. Vzhledem k relativní náročnosti není v zásadě problémem ani kopírování prostřednictvím fotografie, mikrofilmu, mikrofiše či xeroxu. Situace se však radikálně mění v prostředí digitálním, kde je možno exempláře libovolně multiplikovat a kde v zásadě nelze odlišit originál od kopie. Zdálo by se, že to se nevztahuje na starší písemné a dokumentární dědictví, které je již součástí veřejné domény a není chráněno autorským právem. Zjistě se to týká originálních historických dokumentů, nevztahuje se to však na jejich transformace v podobě vědeckých edic a už vůbec ne na sekundární dokumenty, jež se k originálním historickým dokumentům jakožto dokumentům primárním vztahují a více méně těsně s nimi souvisejí (komentáře, anotace apod.).

Má-li zkoumání staršího písemného a dokumentárního dědictví probíhat *lege artis*, nelze od sekundárních dokumentů tohoto druhu odhlédnout. Nadto je žádoucí využít možností digitálního, virtuálního prostředí a tyto sekundární dokumenty zpřístupnit simultánně v kontextu s oněmi primárními. Otázka autorských práv se okamžitě se vši naléhavostí vynoří. Je ovšem třeba ji řešit adaptací autorského práva na nové digitální prostředí a nikoli trváním na způsobech vzešlých z prostředí tištěného. Nelze to alibisticky řešit důrazem na digitalizaci děl, na něž se sice stále vztahuje copyright, která však jsou *out of commerce*, dohadováním s agenturami kolektivně spravujícími autorská práva lobbingem, nýbrž hlubšími právními změnami, což je pochopitelně kontroverzní. Nemusí to však být nutně konfliktní.

Hromadná digitalizace ve spolupráci se společností Google se ovšem netýká jen těchto na první pohled patrných, v podstatě však pouze vnějších aspektů. Je to aktivita mnohem komplexnější, která se z hlediska celého postupu prací dělí na řadu relativně samostatných, ale úzce provázaných dílčích činností, a jejíž uživatelský kontext je velice rozmanitý, dalekosáhlý. Hlavně je třeba stále znovu si uvědomovat, že je plně závislá na přípravných pracích, které musí provést Národní knihovna na fondu určeném k digitalizaci. Sled těchto přípravných prací tvoří následující workflow:

Elektronický katalog

Předpokladem digitalizace je existence elektronického katalogu vytvořeného podle zásad Google. NK ČR elektronický katalog starých tisků nemá, a proto musela urychleně přistoupit k jejich katalogizaci. Na otázku, zda by situace byla jednodušší, kdyby elektronický katalog měla, lze říci, že zkušenosti knihoven, které již s Google digitalizují, ukazují, že ne.

Katalogizace podle potřeb Google Books se liší od katalogizační praxe knihoven, takže i když knihovny elektronický katalog mají, musí jej předělat. Ukázalo se, že nejrychlejší variantou je udělat záznam znovu s knihou v ruce. Hlavní rozdíl je v katalogizaci konvolutů a děl na pokračování. Důraz je kladen především na fyzickou jednotku a ne na bibliografickou, což je podstatný požadavek síťového digitálního prostředí. Jen tak lze totiž jednoduše a jednoznačně ověřit, že digitalizát a katalogizát, tj. digitální kopie originálního historického dokumentu včetně strukturálních metadat a popisná metadata korelují s fyzickou jednotkou, a že tedy výsledkem procesu digitalizace je virtuální realita, tj. protějšek či odraz fyzické reality v síťovém digitálním prostředí, a nikoli faktoid, tj. libovolná manipulace s dílčími digitálními objekty, které sice jednotlivě odrážejí fyzickou realitu, ale jako celek jejím odrazem nejsou, protože jsou záměrným nebo mimovolným klamem. A tak zatímco fyzické jednotky (zjednodušeně svazky) existují ve fyzické

realitě bezprostředně, bibliografické jednotky (zjednodušeně tituly) jsou pomysly odrážející intelektuální konstrukty, tedy třídy jednotlivostí, jež mají ve fyzické realitě odraz pouze zprostředkovaný a s fyzickými jednotkami nevykazují ani shodu, ani totožnost. Praktickým výrazem této složité myšlenkové struktury je právě zkušenost, že v případě konvolutů a děl na pokračování je rychlejší záznam vytvářet než opravovat či upravovat. Vzhledem k tomu, že děl tohoto typu je poměrně velký počet, neexistence elektronického katalogu v Národní knihovně není tak silnou nevýhodou, jak jsme před zahájením prací předpokládali.

Restaurování

V oblasti přípravy knih restaurátory bylo zjištěno, že konzervování knih zaostává za katalogizací, a to v důsledku několika podstatných faktorů:

1. Knihy jsou více poškozené, než se předpokládalo, tudíž bude třeba, buď zvýšit počet restaurátorů nebo oddálit termín digitalizace.
2. Financování není plynulé, tj. stejná situace jako u katalogizace a navíc finanční dotace na rok 2012 byla oproti požadavku zkrácena o půl milionu.
3. Práce restaurátora rozpočítaná na jeden svazek je časově náročnější, než práce katalogizátora. Aby katalogizace udržela krok s restaurováním, je třeba, aby bylo více restaurátorů než katalogizátorů.

Předávání metadat Google

Google požaduje, aby data byla předávána v MARC XML, což systém Aleph zvládá bez problémů. Je to činnost plně automatická a její nastavení bylo provedeno již dříve v souvislosti s jinými úkoly. Poněvadž proces katalogizace teprve plynule probíhá a protože z důvodu jeho nutné rychlosti je nutno vytvářet jen minimální záznamy s nejmenší ještě přijatelnou informační hloubkou, při transformaci do MRC XML nebylo nutno řešit žádné zvláštní otázky týkající se profilu předávaných dat.

Jak bylo výše zmíněno, při předávání dat je kladen důraz na fyzickou jednotku, což je řešeno pouhými příznaky v záznamech (jinak se seskupí).

Předávání knih Google

Předávání knih je především logistická záležitost. Důležité je, aby jednotlivé dávky byly stejně velké a digitalizace šla plynule a pravidelně.

Společně s knihami se předávají seznamy, a to v té podobě, že fyzické pořadí knih v jedné dávce musí korespondovat s pořadím jednotek na seznamu, pro rychlou a snadnou kontrolu. Tento způsob také usnadňuje kontrolu a evidenci knih po digitalizaci před zpětným zařazením do skladiště.

Přebírání dat od Google

Nastane až po digitalizaci. Data budou předána v podobě Komplexního digitálního dokumentu ve složení:

1. Popisná metadata

Metadata: Dle vzorku z New York Public Library, který na ukázkou předal Google Národní knihovně, poskytuje Google popisná metadata se strukturálními metadaty a daty včetně příslušných odkazů na obrazové a textové reprezentace jednotlivých stran knih. To umožňuje po nezbytných transformacích s digitálními kopiemi s nimi dále nakládat jak při archivaci,

tak při procesu zpřístupnění. Tyto transformace metadat pro LTP, Manuscriptorium³ a Krameria⁴ je třeba připravit – bude řešeno po získání rozsáhlejšího vzorku dat (komplexních digitálních dokumentů) od Google.

Popisná metadata MARC XML, jakož i strukturální metadata budou převedena do TEI P5 ENRICH specification pro následné zpřístupnění v Manuscriptoriu a do příslušného vnitřního formátu také při zpřístupnění jejich menší části v Krameriu.

2. Strukturální metadata

3. Obrazová data (resp. plnotextová)

Obrazová data: Google poskytne obrazové soubory ve formátu JPEG (pro zpřístupnění). NK si bude muset převést data do JPEG 2000 (pro archivaci). Google nemá dostatečnou finanční kapacitu, aby předal NK data v obou formátech, ačkoli o tom proběhla příslušná jednání.

Archivace dat v LTP

Data budou archivována vložení do systému LTP, který je součástí NDK.

Zpřístupnění dat

Po příslušných konverzích budou data vzniklá digitalizací s Google zpřístupněna v Manuscriptoriu a v Krameriu, případně i jinde (TEL, EUROPEANA, The CERL Portal, CENDARI) a samozřejmě budou dostupná v Google.

-
- 3 Viz KNOLL, A. Projekt ENRICH: budujeme Evropskou digitální knihovnu rukopisů. In: *Muzeum a změna III*. Praha: Asociace muzeí a galerií České republiky, 2009, s. 169–172; KNOLL, A. – PSOHLAVEC, S. – PSOHLAVEC, T. – UHLÍŘ, Z. Creation of an International Digital Library of Manuscripts: Seamless Access to Data from Heterogeneous Resources (ENRICH Project). In: *Rethinking Electronic Publishing: Innovation in Communication Paradigms and Technologies*. Mornati, S. – Hedlund, T. Roma: Edizioni Nuova Cultura, 2009, s. 335–347; KNOLL, A. – PSOHLAVEC, S. – PSOHLAVEC, T. – UHLÍŘ, Z. ENRICH: an eContentPlus Project for Creation of a European Digital Library of Manuscripts. In: *Digital Heritage. Limassol: Archaeolingua*. 2008, s. 201–206; KNOLL, A. ENRICH: Un réseau européen des manuscrits anciens numérisés. In: *Culture et Recherche. Numérisation du patrimoine culturel*. Paris: Ministère de la Culture et de la Communication, nos. 118–119, automne-hiver 2008–2009, s. 24; UHLÍŘ, Z. Manuscriptorium nové trendy ve zpřístupnění písemného dokumentárního dědictví. In: *ITLib: Informačné technologie a knižnice*. 2010, roč. 14, č. 2, s. 11–16, dostupné z URL <http://www.cvtisr.sk/itlib/itlib102/uhlir.htm> [cit. 2012-07-08]; UHLÍŘ, Z. Stvaranje ambijenta za virtuelno istraživanje istorijskih izvora. In: *Savremena biblioteka*. 2010, roč. 22, č. 27, s. 24–29; UHLÍŘ, Z. Unitatea polimorfă a culturii Europii Centrale: proiectele ENRICH și REDISCOVER – Multiform Unity of Central European Culture: projects ENRICH and REDISCOVER. In: *Revista Română de biblioteconomie și știința informării*. 2010, roč. 6, č. 2, s. 53–64; UHLÍŘ, Z. Evropský projekt ENRICH a jeho význam pro vybudování virtuálního badatelského prostředí. In: *Knihovna – knihovnická revue*. 2010, roč. 21, č. 1, s. 5–14, dostupné z URL: <http://knihovna.nkp.cz/knihovna101/10105.htm> [cit. 2012-07-08]; UHLÍŘ, Z. Manuscriptorium jako distribuovaná digitální knihovna. In: *Problematika historických a vzácných knižních fondů 2009*. Olomouc: Vědecká knihovna – Brno: Sdružení knihoven České republiky, 2010, s. 71–80, dostupné z URL: [http://www.vkol.cz/data/soubory/import/konf18/po7-Uhl%C3%ad%C5%99-2009.pdf](http://www.vkol.cz/data/soubory/import/konf18/po7-Uhl%C3%AD%C5%99-2009.pdf) [cit. 2012-07-08]; UHLÍŘ, Z. – KNOLL, A. Manuscriptorium Digital Library and ENRICH Project: Means for Dealing with Digital Codicology and Palaeography. In: *Kodikologie und Paläographie im digitalen Zeitalter – Codicology and Palaeography in the Digital Age*. REHBEIN, M. – SAHLE, P. – SCHAASSAN, T. 2009, pp. 65–76.
- 4 Viz FOLTÝN, T. – ŠVÁSTOVÁ, P. Kramerius 4 – řešení pro zpřístupnění digitálních dokumentů. In: *Knihovny současnosti 2011*. Ostrava: SDRUK, 2011, s. 90–97, dostupné z URL: http://www.svkos.cz/data/xinha/sdruk/ks2011/sbornik_2011.pdf [cit. 2012-07-08]; LHOTÁK, M. Uživatelské rozhraní systému Kramerius 4. In: *Čtenář*. 2011, roč. 63, č. 6, s. 227–229.

PARALELNOST DIGITALIZACE PROSTŘEDNICTVÍM GOOGLE A DIGITALIZACE IOP

I když hromadná digitalizace⁵ se společností Google a digitalizace v rámci projektu IOP pro NDK⁶ se jeví jako paralelní nebo dokonce vzájemně zástupná, je to pouhý dojem vyplývající ze zastaralé představy, podle které je digitalizace pořizování digitálních obrazových kopií originálních historických dokumentů. Digitalizace však je mnohem komplexnější aktivitou,⁷ která spočívá jednak v transformaci tradičních tištěných nebo i rukopisných dokumentů do podoby použitelné v síťovém digitálním prostředí,⁸ jednak v tvorbě a využívání různých digitálních nástrojů pro studium písemného dokumentárního dědictví.⁹ A tedy jak digitalizace prostřednictvím Google, tak digitalizace IOP se sice týká národního kulturního dědictví, ale rozdíl spočívá v tom, že digitalizace Google se orientuje na knižní produkci 16.–18. století, zatímco IOP je zaměřen na produkci 19.–20. století. Nejedná se tedy ani o aktivity konkurenční, ani o aktivity zástupné, nýbrž komplementární.

Digitalizace IOP se soustřeďuje na oblast nacházející se na rozhraní mezi živou a mrtvou kulturou s příklonem spíše ke kultuře živé, jež má být konvertována do síťového digitálního prostředí. Digitalizace prostřednictvím Google se naopak obrací výhradně ke kultuře mrtvé, jež má být nejen prezentována, ale také integrována do nově vznikající kultury znalostní společnosti a do utvářející se civilizace digitálního světa.

Již z tohoto rozdílného cíle je zřejmé, že v případě obou těchto významných digitalizačních aktivit není možno kromě nejzákladnějších principů uvažovat o stejném přístupu k přípravě, zpracování, reprezentaci, zpřístupnění a využívání. Z toho také plyne, že nelze zejména v případě popisných metadat, jakož i digitálních nástrojů pro vědeckou a odbornou práci s komplexními digitálními dokumenty používat stejných standardů formátů pro obě tyto oblasti. Všechny standardy a formáty jsou totiž v konečném účelu uživatelsky zaměřené, nelze je tudíž bez nežádoucích následků libovolně přesouvat z oblasti do oblasti. Obě digitalizační aktivity je tudíž nutno vidět jako samostatný příspěvek k přechodu do síťového digitálního informačního komunikačního prostředí a k vytváření stále sofistikovanějšího virtuálního prostředí se specializovanými službami a personalizovanými nástroji pro co nejrozmanitější práci s informacemi různého druhu a zaměření.

Naopak konkrétní prostředí, jakož i zpřístupňující systémy a nástroje jsou a budou v obou případech různé, pokud nemá dojít k defektům v běžném provozu a k degradaci služeb koncovým uživatelům. Zcela nepochybně oběma společnou zůstává a zůstane

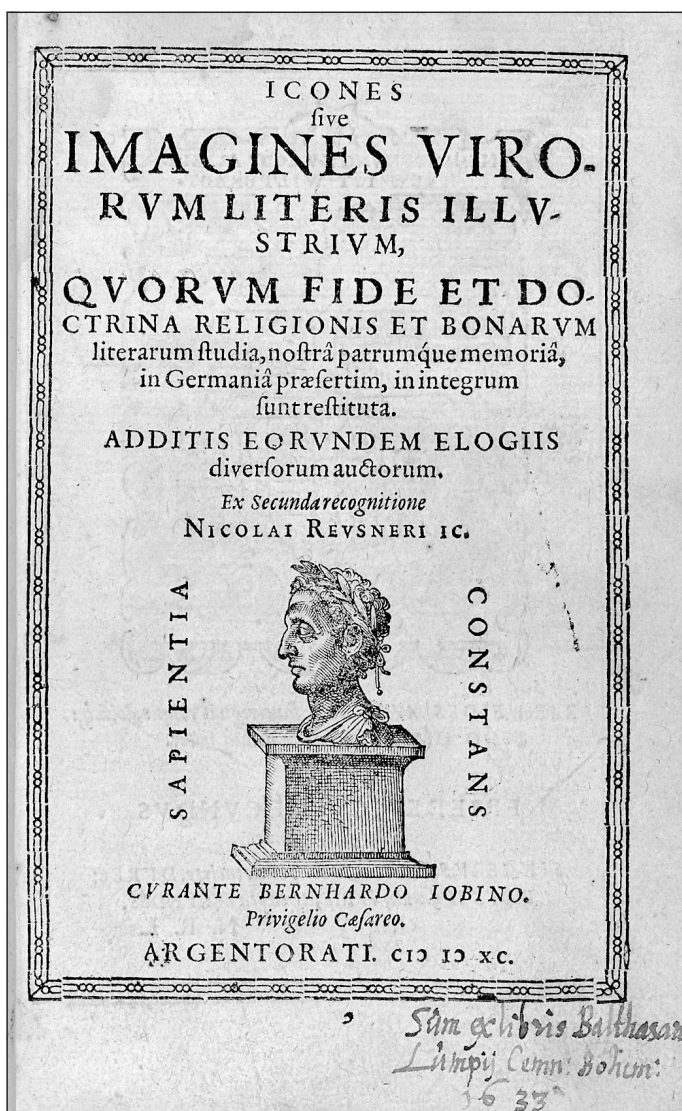
5 V Národní knihovně ČR již dříve zahájena hromadná digitalizace v rámci Norských fondů se týká novodobých monografií, srv. POLIŠENSKÝ, J. Účast NK ČR v projektu Norské fondy – digitalizace bohemikálních monografií ohrožených degradací papíru. In: *Knihovny současnosti 2009*. Brno: SDRUK, 2009, s. 167–172, dostupné z URL: <http://www.sdruk.cz/sec/2009/sbornik/2006-6-167.pdf> [cit. 2012-07-08].

6 Viz STOKLASOVÁ, B. Národní digitální knihovna. In: *Knihovny současnosti 2009*. Brno: SDRUK, 2009, s. 12–18, dostupné z URL: <http://www.sdruk.cz/sec/2009/sbornik/2009-0-012.pdf> [cit. 2012-07-08]; STOKLASOVÁ, B. – HUTAŘ, J. – MELICHAR, M. Národní digitální knihovna. In: *Knihovna: knihovnická revue. Rok 2009, roč. 20, č. 1, s. 6–21*, dostupné z URL: <http://knihovna.nkp.cz/knihovna91/humesto.htm> [cit. 2012-07-08].

7 Viz UHLÍŘ, Z. Digitization is not only Making Images: Manuscript Studies and Digital Processing of Manuscripts. In: *Knygotyra*. 2008, č. 51, s. 148–162. Dostupné z URL: <http://www.leidykla.vu.lt/fileadmin/Knygotyra/51/148-162.pdf> [cit. 2012-07-08].

8 Viz UHLÍŘ, Z. Technické obrazy a transformace kodikologie a bibliologie. In: *Národní knihovna: knihovnická revue. Rok 2003, roč. 14, č. 2, s. 114–127*, dostupné z URL: <http://full.nkp.cz/nkkr/NKKR0302/0302114.html> [cit. 2012-07-08].

9 Viz UHLÍŘ, Z. Digital Codicology and Contextual Editing of Medieval Manuscripts. In: *Revista arhivelor – Archives Review*. 2009, roč. 86, č. 2, s. 36–54; Углирж, Зденек: Цифровые инструменты для изучения средневековых рукописей и текстов (v tisku).



Obrázek č. 1: Jan Carion: Kronika světa. Praha: Daniel Adam z Veleslavína, 1602, Národní knihovna České republiky sign. 65.D.1134.

složka archivace. Národní digitální knihovna ve smyslu centrálního digitálního úložiště je tak vlastně spojnicí digitalizace prostřednictvím Google a digitalizace IOP. V jejím rámci mohou být standardy pro archivní, tj. primární data digitálních obrazů stejná pro obě větve digitalizačních aktivit, poněvadž sekundární data sloužící pro zpřístupnění jsou tak jako tak vždy odvozená, ať už s ohledem na kompresi informace, nebo konverzi formátu. Totéž platí pro některá metadata, jež slouží výhradě organizaci archivu, tj. centrálního digitálního úložiště a nejsou potřebná pro žádný krok při prezentaci koncovému uživateli.

Bezpečným dlouhodobým uložením digitálních dat a z něho vycházejícím různým způsobům prezentace koncovému uživateli, jakož i následnou agregací v dalších zdrojích, zpřístupněním prostřednictvím dalších portálů a metaindexů integrací v různých síťových

službách, se zatím jednoduché digitální prostředí rozvine i v takových oblastech a oborech, kde doposud patřičně nenaplnjuje očekávání a zatím neodpovídá svým možnostem.

Na význam hromadné digitalizace je možné nahlížet ve dvojitým smyslu. Jednak jako na síťové zpřístupnění kulturního dědictví bez ohledu na místo a čas, k němuž lze přistoupit kdykoli a odkudkoli. Tato skutečnost sama je dnes jistě banální a triviální, nicméně málo se uvažuje o následcích, jež přináší a které rozhodně banální a triviální nejsou. Znamenají totiž zásadní změnu paradigmatu jak individuální knihovnické práce, tak organizačního rámce tradičních kamenných institucí knihoven.

Poněvadž podstatná složka hromadné digitalizace je outsourcována soukromým subjektem a větší část zbylé neoutsourcované práce je prováděna externími a nikoli interními pracovníky knihovny, instituce knihovny se tak vlastně stává pouhým rozhraním (ve významu spíše jen formální hranice) mezi koncovým uživatelem a tvůrci digitalizačních a katalogizačních výstupů. Na jedné straně se stává organizační složkou celého procesu, skrze něj procházejí finanční zdroje a část zdrojů materiálních, na druhé straně administrátorem fyzických svazků, které přestanou být koncovými uživateli požadovány, protože uživatelé budou žádat jejich digitální kopie, které však za každou cenu musí zůstat zachovány, aby bylo jisté a ověřitelné, že digitalizáty jsou virtuální realitou a nikoli zmanipulovanými faktoidy a že katalogizáty jsou z kodikologického či bibliologického hlediska spolehlivé a že v sémantické redukci podávají požadovanou informaci.

Vzhledem k tomuto způsobu zpřístupnění je však zřejmé, že řada údajů, které tradiční knihovědná, kodikologická a bibliologická teorie a metodologie považovala pro katalog za nezbytné, neboť v sémantické redukci nahrazují knihu samotnou, je nyní zbytná, neboť s katalogizátem spojený digitalizát nabízí nikoli pouhou sémantickou redukci některých vnějších znaků, nýbrž je představuje v přímém názoru. Digitalizace vůbec a hromadná digitalizace zvláště tedy vede k omezení důrazu kladeného na vnější znaky a naopak vede k umocnění významu vnitřních znaků, jakož i sbírkového, knihovního či archivního kontextu.

Enormně se tedy zvyšuje důležitost analytického popisu oproti obvyklému popisu blokovému a právě tak nabývají na významu podrobné údaje o provenienci, o vazbách atd. atp. Všeobecně lze tedy hovořit o tom, že hromadná digitalizace ve svém důsledku znamená rozhodný důraz na analytickou evidenci se zřetelem k textovému obsahu a na materiálovou evidenci¹⁰ se zřetelem ke kulturnímu kontextu, v němž se pohybovali tehdejší koncoví uživatelé. V konečné důsledku to znamená kvantitativní přístup, důraz na řešení problémů kvantitativní kodikologie a bibliologie a s vědomím, že kvantifikace v případě humanitních disciplín neznamená vyhledávání empiricky zjištělných pravidelností, ale naopak sledování změn,¹¹ protože historie v širokém smyslu je vědou o změně.

Význam hromadné digitalizace lze nahlížet, jak plyne z předchozího, také jako možnost hromadné heuristiky a jako využití digitálních nástrojů pro vědeckou a odbornou práci. Umožňuje to spojení zpřístupnění elektronického publikování a tzv. *Culturomics*.¹²

¹⁰ Srv. databázi, kterou koordinuje Cristina Doni, *Material Evidence in Incunabula*, dostupné z URL: <http://www.cerl.org/web/en/help/incunabula/main> [cit. 2012-07-08], resp. <http://incunabula.cerl.org/cgi-bin/search.pl> [cit. 6.7.2012], jež je zároveň propojena *Incunabula Short Title Catalogue*, dostupné z URL: <http://www.bl.uk/catalogues/istc/index.html> [cit. 2012-07-08], resp. *Index Possessorum Incunabulorum*, dostupné z URL: <http://ipi.cerl.org/cgi-bin/search.pl> [cit. 2012-07-08].

¹¹ Srv. MISES, L. von. *Theory and History*. New Haven, Conn.: Yale University Press, 1957.

¹² Viz BOHANNON, J. The birth of *culturomics*: Google opens books to new cultural studies. In: *Science*. 2010, sv. 330, č. 6011, s. 1600, dostupné z URL: <http://www.sciencemag.org/content/330/6011/1600.summary> [cit. 2012-07-08]; BOHANNON, J. Google Books, Wikipedia, and the Future of *Culturomics*. In: *Science*. 2011, č. 331, s. 135, dostupné z URL: <http://www.terceracultura.net/tc/wp-content/uploads/2011/01/culturomics.pdf> [cit. 2012-07-08].

Google klade důraz na požadavky koncového uživatele a odhlíží od zájmů knihovníků (tj. formálních údajů knihy). Je to vidět na jednoduchosti záznamu a na důrazu na obsahové údaje (obrazy obohacené plným textem). Výstupem je velké množství snadno vyhledatelných a dostupných údajů. Tím nechtěně poukazuje na těžkopádnost knihoven ve zpřístupňování informací. Tento přístup představuje velikou výzvu nejen pro knihovny, ale také pro akademické instituce a všechny zúčastněné individuální osoby vůbec. Důraz na obsahové údaje totiž vyžaduje mnohem výraznější a rozmanitější dělbu práce, než jaká byla obvyklá po stabilizaci institucionální vědecké a odborné infrastruktury v důsledku tzv. první informační krize někdy na přelomu 18. a 19. století. A to tak výraznou, že už od přelomu tisíciletí se nově hovoří o druhé informační krizi, jež spočívá ve stírání hranic mezi tvůrcem, prostředníkem a uživatelem, resp. v překrývání těchto rolí. Dvě zásady, které platily až dosud, tj. jednak organizace založená na členění podle různých funkcí v procesu vytváření, sdělování a recipování, resp. konzumování znalosti a informace, jednak přesvědčení, že jeden deskriptivní metadatový záznam musí být vytvářen právě jedním člověkem, svou platnost ztrácí.

Z toho lze vyvodit, že teoretické konsekvence hromadné digitalizace ve spolupráci se společností Google jsou dalekosáhlé. Neboť se nejspíše neprojeví okamžitě, nýbrž až v delší perspektivě, bude lépe o nich pojednat speciálně. Více méně totiž přesahují tematické zaměření tohoto příspěvku. Bezprostřední praktické důsledky však opomenout nelze, jelikož kladou různé neobvyklé otázky. Tím, že vyvstávají náhle a naléhavě, dožadují se odpovědi a řešení.

Hromadná digitalizace prokazuje, že koncoví uživatelé přestávají mít zájem o instituce typu knihoven a naopak je s lehkým srdcem opomíjejí, mají-li možnost se k požadované informaci dostat rychleji, snáze a pohodlněji. Jestliže digitalizované historické fondy mnoha významných knihoven budou přístupné jednotně prostřednictvím Google Books, resp. Google Book Search, ztrácejí uživatelé důvod obtěžovat se mnohonásobným vyhledáváním v jednotlivých knihovnách nebo dokonce v integrovaných službách typu WorldCat,¹³ TEL¹⁴ či HPB,¹⁵ resp. The CERL Portal.¹⁶ To je velká výhoda pro společnost Google, která jí dovoluje riskovat ohromnou investicí do hromadné digitalizace. Zároveň to dokládá, že za rovných a ekonomicky kalkulovatelných podmínek má komerční business model převahu nad modelem veřejných služeb. Dokladem ostatně je srovnání s výsledky portálu Europeana, masivně podporovaného z veřejných zdrojů.¹⁷

Přesto však knihovny, které spolupracují na hromadné digitalizaci svých historických fondů v rámci Google, mohou při poskytování služeb vycházejících z této digitalizace najít své místo. Musí však změnit svůj přístup. Koncovému uživateli musí nabídnout dodatečnou informaci, kterou ocení jako přidanou hodnotu. Výhodou společnosti Google je dispozice s minimálními popisnými metadaty, s digitálními obrazy a s plnými texty vyrobenými za použití technologie OCR z těchto obrazů, tedy schopnost vytvořit globální síťové prostředí. To je za využití nástrojů postupů culturomics zároveň prostředím virtuálním, musí každá knihovna svou komparativní výhodu hledat na jiném poli. Nalezne ji v poskytování personalizovaných služeb ve speciální oblasti bádání a výzkumu týkajících se písemného a dokumentárního dědictví. Jestliže Národní knihovna ČR bude mít od společnosti Google k dispozici záznamy, obrazy a plné texty, nestačí, aby je pouze zařadila do Manuscriptoria, případně do Krameria, ale tyto informace bude muset nějakým způsobem obohatit, tzn. prohloubit a rozšířit.

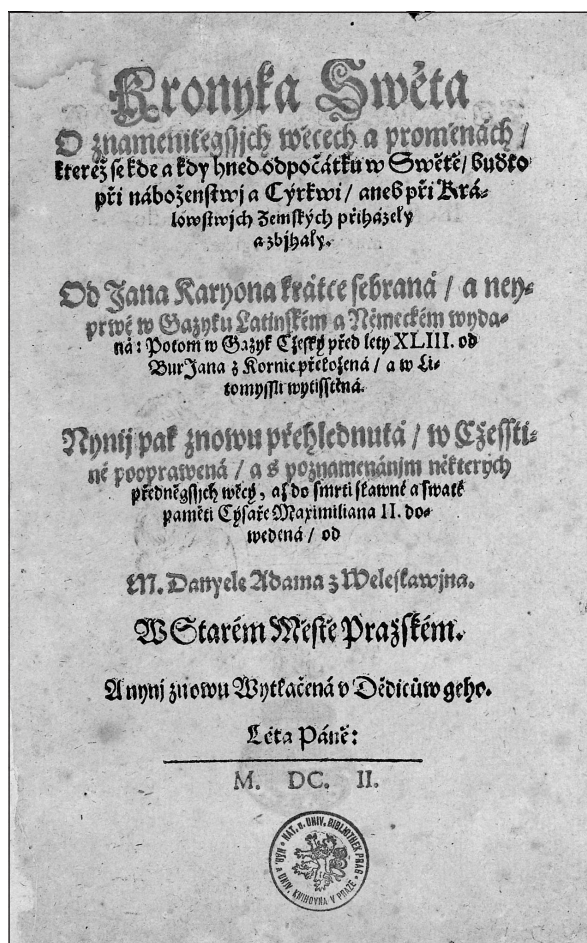
¹³ Viz <http://www.worldcat.org/>.

¹⁴ Viz <http://www.theeuropeanlibrary.org/tel4/>.

¹⁵ Viz <http://www.cerl.org/web/en/resources/hpb/main/>.

¹⁶ Viz http://www.cerl.org/web/en/resources/cerl_portal/.

¹⁷ Viz <http://www.europeana.org/portal/>.



Obrázek č. 2: Nicolaus Reusner: *Icones sive imagines virorum litteris illustrium*, Strasbourg: Bernard Jobin, 1690, Národní knihovna České republiky sign. 1.K.49.

Po přijetí komplexních digitálních dokumentů od Google tedy bude muset kromě konverzí nutných jak pro archivní uložení v centrálním digitálním úložišti Národní digitální knihovny, tak pro prezentaci v Manuscriptoriu, resp. Krameriu postupovat ve dvou liniích.

První z nich představuje obohacování minimálních metadat o analytické popisy spočívající nikoli v tradičních knihovědných, kodikologických a bibliologických popisech vnějších znaků, ale v detailním postžení vnitřních znaků, tj. především v podrobném rozpisu obsahu. Zároveň bude vhodné věnovat se popisu hlediska materiálové evidence a kulturněhistorického kontextu, v tomto případě však bude efektivnější kooperativní postup s jinými knihovnami na vytváření nadnárodních databází proveniencí, knižních vazeb atd., které budou snadno korelovatelné s personalizovaným prostředím Manuscriptoria.

Lze předpokládat, že Národní knihovna ČR nebude schopna vlastními silami určenými ke katalogizaci starých tisků tuto činnost zajistit a že bude nucena ji koordinovat s institucionálním výzkumem a vývojem, případně s externími silami či v rámci domácích i mezinárodních grantových projektů. Ukazuje se, že vývoj v této oblasti spěje k zásadní organizační transformaci v tom smyslu, že profesionální profil historického knihovníka přestává být totožný s oborem knihovědy a že bude nutno klást mnohem větší důraz na obecně historický a obecně filologický základ.

Druhou linií bude další práce s plnými texty, které Google pořídí technologií OCR¹⁸ a v rámci Google Book Search je bude využívat pro vyhledávání, protože příslušné úseky textu budou korelovány k odpovídajícím obrazům. Bude vhodné, ne-li přímo nutné tyto automaticky vzniklé plné texty korigovat tak, aby z nich vznikl surový podklad způsobilý pro použití při tvorbě pragmatických a kontextuálních edic.

Toto je aktivita zcela nová, kterou až dosud nikdo od knihoven a historických knihovníků neočekával, a tedy ji na nich ani nepožadoval. Nyní jsou jí schopni pravděpodobně jen nemnozí pracovníci výzkumu a vývoje Národní knihovny ČR, v zásadě však nikoli standardní historičtí knihovníci, zejména ne ti, kteří pracují se starými tisky. Jelikož ve vztahu k plným textům pouze tato činnost může vytvořit dodatečnou informaci a koncovému uživateli poskytnout přidanou hodnotu, bude nutno se tohoto úkolu zhostit přes jistý odpor pracovníků a přes pravděpodobné pochybnosti instituce; bez něj totiž knihovna nenávratně ztratí komparativní výhodu oproti obecnému globálnímu prostředí vytvořenému společností Google a v důsledku toho mohou vzniknout více méně oprávněné pochybnosti o důvodu její existence. Takže i v této linii se bude muset tato činnost koordinovat s institucionálním výzkumem a vývojem, případně počítat s externími silami a domácími, zahraničními i mezinárodními grantovými projekty. A pro útvar historických a hudebních fondů platí i zde stejné memento jako v předchozím případě.

Tyto dvě linie (obohacování popisných záznamů a zpracovávání plných textů) se spojí na vyšším stupni. Zde bude muset fakticky dojít ke koordinaci globálního síťového prostředí na jedné straně a personalizovaného virtuálního prostředí knihoven (čímž už nyní rozumíme digitální knihovny, nikoli tradiční kamenné instituce) na straně druhé. Přičemž jednomu zůstane komparativní výhoda robustnosti a masivnosti a druhému specializace a personalizace.

Pak bude možné kombinovat a integrovat na jedné straně ontologické a prosté textové vyhledávání týkající se metadat-popisných katalogových záznamů (terciárních dokumentů) a na straně druhé vyhledávání sémantické, plnotextové a bibliografické týkající se dat-plných textů (primárních a sekundárních dokumentů).¹⁹ Zároveň bude možno využívat v návaznosti na nový sofistikovanější druh vyhledávání také výsledky externích nástrojů služeb koordinovaných s Manuscriptoriem (např. automatické srovnávání plných textů a jeho statistické vyhodnocování, jakož i jeho grafická reprezentace, nebo automatické obohacování popisných metadat o informaci o typu obsahu digitálních obrazových kopií, či korelace narativních a diplomatických historických textů s georeferencovaným mapovým zobrazením apod.).

Kromě toho personalizované prostředí Manuscriptoria umožní digitální síťové publikování sekundárních dokumentů týkajících se písemného dokumentárního dědictví vůbec a primárních či originálních historických dokumentů zvláště. K tomu však ještě chybí transnacionální jednoznačná identifikace (persistentní adresování komplexních digitálních dokumentů v příslušných digitálních knihovnách). Je totiž zřejmé, že v důsledku rozsáhlé mezinárodní

¹⁸ V rámci digitalizace novodobých fondů v Národní knihovně ČR se také užívá technologie OCR, srv. FOLTÝN, T. Účast NK ČR v projektu TELplus – vytváření OCR souborů. In: *Knihovny současnosti 2009*. Brno: SDRUK, 2009, s. 159–166, dostupné z URL: <http://www.svkos.cz/data/xinha/sdruk/2009-6-159.pdf> [cit. 2012-07-08]; POLIŠENSKÝ, J. – LEHRL, O. CODEG 4 – nástroj pro podporu OCR a tvorbu rozšířených metadat. In: *Knihovny současnosti 2011*. Ostrava: SDRUK, 2011, s. 80–89, dostupné z URL: http://www.svkos.cz/data/xinha/sdruk/ks2011/2011_KKS_Polisensky.pdf [cit. 2012-07-08].

¹⁹ Viz SCHÄFER, U. The Searchbench – Combining Sentence-Semantic, Full Text and Bibliographic Search in *Digital Libraries*, dostupné z URL: http://www.utlib.ee/liber2012/index.php?id=programme&item=acc_papers [cit. 2012-07-08].

a nadnárodní agregace řešení na národní úrovni nestačí²⁰ a mezinárodní řešení, k němuž se schyluje v případě autorů, resp. přispěvatelů (contributor),²¹ je mimoběžné.

Skutečné aktivity knihoven spravujících a zpřístupňujících historické fondy budou již v brzké budoucnosti probíhat na těchto stupních a nikoli v uzavřených hranicích tradiční instituce chápané jako hortus conclusus. Jinak koncoví uživatelé nebudou mít vzhledem k podstatně změněným možnostem informační znalostní práce důvod brát na takové instituce zřetel. Pravděpodobně to bude markantnější v případě staršího písemného dokumentárního dědictví, tj. v případě historických fondů, nicméně ani oblast novodobých fondů se nevyhne poznání, že digitalizace není jen technická a organizační záležitost, ale že má silný dopad i na obsahovou stránku, zatím však to v této oblasti dostatečně jasné není.²² A tak je jisté, že přes komplementaritu hromadné digitalizace ve spolupráci se společností Google a digitalizace IOP uběhne ještě dlouhá doba, než dojde k vytvoření jednotného prostředí.

DIGITIZATION IN THE NATIONAL LIBRARY OF THE CZECH REPUBLIC IN COOPERATION WITH GOOGLE AND ITS RELATION TO THE NATIONAL DIGITAL LIBRARY – INTEGRATED OPERATIONAL PROGRAM

Summary: National Library of the Czech Republic started in 2011 collaboration with Google concerning mass digitization of historical and rare collections that is agreed to 2016 not excluding further prolongation. Goal of the collaboration is making accessible digital copies of historical holdings to 1800 as well as selected rare collections with added value of full text search. End users will obtain seamless access to of about 150 thousand historical and rare holdings both via Google Books and Manuscriptorium, or Kramerius respectively. Thus both National Library of the Czech Republic and other libraries as well are forced to provide special services with additional information and added value through use of information and communication technologies and methodologies based on so called culturomics and digital humanities.

Keywords: Mass digitization, historical holdings, written and documentary heritage, analytical cataloguing, material evidence.

²⁰ K českému řešení týkajícímu se NDK srv. např. CUBR, L. – HUTAŘ, J. – MELICHAR, M. Kontrolní seznam pro strategii zajištění persistence identifikátorů. In: *Knihovna: knihovnická revue*. 2009, roč. 20, č. 2, s. 54–62, dostupné z URL: <http://knihovna.nkp.cz/knihovna92/0902054.htm> [cit. 2012-07-08]; CUBR, L. Budování důvěryhodného systému trvalé identifikace digitálních dokumentů. In: *Knihovna: knihovnická revue*. 2010, roč. 21, č. 1, s. 23–31, dostupné z URL: <http://knihovna.nkp.cz/knihovna101/10123.htm> [cit. 2012-07-08].

²¹ Srv. např. FENNER, M. ORCID: Unique Identifiers for Authors and Contributors, příspěvek na konferenci „The Value of Unique Scholarly Identifiers to Academics, Institutions and Countries“ konané dne 14. 2. 2012 ve Vilně, dostupné z URL: <http://uniqueids.org/orcid/> [cit. 2012-07-08].

²² Srv. STOKLASOVÁ, B. Mrtvá a živá, aneb Strategie ochrany a zpřístupnění národního kulturního dědictví v Česku a v Norsku. In: *Knihovna: knihovnická revue*. 2009, roč. 20, č. 2, s. 37–48, dostupné z URL: <http://knihovna.nkp.cz/pdf/0902/090237.pdf> [cit. 2012-07-08].