

## Budování důvěryhodného systému trvalé identifikace digitálních dokumentů

PhDr. Ladislav Cubr / Odbor digitální ochrany NK ČR / doktorand UISK FF UK / [ladislav.cubr@nkp.cz](mailto:ladislav.cubr@nkp.cz)

### Resumé:

Cílem článku je přiblížit problematiku trvalé identifikace digitálních dokumentů širší knihovnické obci a upozornit na potřebu budovat národní systém trvalé identifikace digitálních dokumentů v rámci širšího projektu. Systém trvalé identifikace digitálních dokumentů je nutné trvale koordinovat se systémy pro dlouhodobou správu a archivaci digitálních dokumentů a jeho administrace vyžaduje vyšší spoluúčast různorodých institucí než v případě systému trvalé identifikace tištěných dokumentů. V článku podrobněji probíráme rozdíly identifikace tištěných a digitálních dokumentů. Představujeme několik základních tezí vztahujících se k systému trvalé identifikace digitálních dokumentů a upozorňujeme na některá mylná přesvědčení. Nakonec představujeme zahraniční projekty zaměřené na trvalou identifikaci digitálních dokumentů a důvody, proč by podobný projekt měl vzniknout i v České republice.

**Klíčová slova:** trvalá identifikace – digitální dokumenty – dlouhodobá ochrana digitálních dokumentů.

### Summary:

This article aims to approach issues of persistent identification of digital documents for the wider library community and highlight the need to build a national persistent identifiers system for digital documents within a broader project. The persistent identifiers system for digital documents is to be permanently coordinate with the systems for long-term management and archiving of digital documents and its administration requires greater participation of various institutions than in case of the persistent identifiers system for printed documents. We discuss in more detail the differences in the identification of printed and digital documents. We introduce some basic propositions related to the persistent identifiers system for digital documents and point out the mistaken beliefs. Finally we introduce foreign projects for persistent identification of digital documents and the reasons why the project like this should arise in the Czech Republic.

**Keywords:** persistent identification – digital documents – digital preservation.

## Úvodem

Cílem tohoto článku je přiblížit širší knihovnické obci problematiku zavádění a správy digitálních identifikátorů<sup>1</sup> a zdůraznit skutečnost, že úspěšný systém trvalé identifikace digitálních dokumentů závisí na celé řadě dalších souvisejících aktivit, jejichž provázanost s digitálními identifikátory nemusí být na první pohled zřejmá. Problém spolehlivého masového užití digitálních identifikátorů v národním kontextu není možné řešit bez návaznosti na širší kontext ostatních digitálních aktivit knihoven (digitální knihovny, digitalizace, dlouhodobá ochrana digitálních dokumentů). Národní systém trvalé identifikace digitálních dokumentů nelze vyřešit jedním článkem nebo prací izolované skupiny expertů. Musí být vyvíjen v rámci celonárodního projektu zainteresovaných stran, které dospějí ke společnému konsensu o tom, jak by měl takový systém fungovat, jaká by

<sup>1</sup> Trvalé identifikátory digitálních dokumentů budeme v tomto článku označovat jako digitální identifikátory, trvalé identifikátory analogových dokumentů (tištěné knihy, papírové časopisy apod.) jako analogové identifikátory. Digitálním dokumentem zde rozumíme skupinu tvořenou jedním nebo více počítačovými soubory, která je v knihovnickém kontextu pojímána jako jednotka intelektuálního obsahu a využívána pro potřeby správy, ochrany a zpřístupňování, případně jako jedna jednotka identifikována právě digitálním identifikátorem.

měl mít pravidla a jaké by mělo být rozložení odpovědností příslušných institucí při jeho správě.

## Systém trvalé identifikace digitálních dokumentů

Digitální identifikátory mohou být v celonárodním měřítku užívány ke dvěma hlavním účelům. Prvním je jednoznačná identifikace digitálního dokumentu v jasné vymezeném kontextu (minimálně národním, ideálně globálním), která je důležitá zejména pro knihovníky vzhledem k jejich potřebě dlouhodobé správy a ochrany digitálních dokumentů. Druhým účelem je funkce digitálního identifikátoru jako prostředku k přímému získání dokumentu, která je užitečná zejména pro uživatele z hlediska snadného vyhledávání dokumentů nebo citační praxe. Na rozdíl od analogového světa, v němž pomocí analogového identifikátoru (např. signatury) můžeme dokument získat jen zprostředkovaně (přes knihovníka, objednávkou v katalogu apod.), je digitální identifikátor sám tímto prostředníkem, který po zadání do počítače umožní zobrazit požadovaný dokument přímo na monitoru tohoto počítače.

Musíme předem upozornit, že nám v žádném případě nejde pouze o trvalé identifikátory samotné (například jedinečné číselné řady sloužící k identifikaci digitálního dokumentu). Hlavním těžištěm zájmu musí být celý dobře promyšlený systém trvalé identifikace. Tento systém je to, co stojí (ve skrytosti před uživateli) za digitálními identifikátory a co umožňuje jejich bezproblémové využívání v dlouhodobém horizontu. Systém trvalé identifikace lze definovat jako navzájem provázanou kombinaci definic, pravidel, služeb a dokumentů, která je určena pro administraci a využívání trvalých identifikátorů [nesator, 2009, s. 3]. Dodejme, že takto definovaný systém logicky předpokládá odpovídající technické a personální zázemí. Digitální identifikátory jsou jen jedním z článků systému trvalé identifikace, byť jde přirozeně o článek nejviditelnější.

V čem tedy spočívá vysoká komplexita trvalé identifikace digitálních dokumentů? Dobře víme, že expanze digitálního světa měla v posledních přibližně patnácti letech výrazný dopad na práci knihoven. Prověřené postupy pro získávání, zpracování, správu, ochranu a zpřístupňování dokumentů je nutné stále přizpůsobovat novým digitálním výzvám. Starší systémy trvalé identifikace, zavedené původně pro tištěné dokumenty (zejména systémy ISBN a ISSN), fungují již delší dobu relativně spolehlivě. Identifikace digitálních dokumentů je však stále na půli cesty. K dispozici sice již máme několik globálních identifikačních systémů (zejména Handle, DOI, PURL a URN:NBN), liší se však úrovní propracovanosti a jejich integrace do různorodého světa knihovnických procesů je dosud nedokonalá. Širší knihovnická komunita si dosud ne zcela uvědomuje, že vazba mezi správou a ochranou digitálních dokumentů na jedné straně a systémem jejich identifikace na straně druhé je mnohem těsnější než ve světě tištěných dokumentů. Domníváme se, že pochopení rozdílů mezi identifikací tištěných a digitálních dokumentů vytváří základ pro porozumění problémům digitálních identifikátorů. V následující části se proto pokusíme blíže představit tyto rozdíly na srovnání systému trvalé identifikace digitálních dokumentů se systémem ISBN, tedy s dobře zavedeným systémem trvalé identifikace tištěných dokumentů.<sup>2</sup>

## Analogové a digitální identifikátory

Systém ISBN se může opřít o globální infrastrukturu, kterou tvoří soustava národních agentur přidělujících ISBN (často jsou jimi národní knihovny daných zemí), široká komunita zúčastněných stran (vydavatelé knih, knihovny) a uživatelů (čtenáři vyhledávající knihy apod.) a propracovaný systém pravidel a odpovědností (vymezení vztahů mezi vydavateli a národními agenturami, podmínky, za kterých lze číslo ISBN přidělit apod.). Z hlediska vazby mezi správou, ochranou a identifikací platí, že systém ISBN je spojen pouze s ranou fází životního cyklu dokumentů. Identifikátor ISBN je přidělen vydavatelem při přípravě knihy (před jejím vytištěním), zanesen do národního registru,

<sup>2</sup> Identifikátor ISBN sice může být přidělen i některým typům digitálních dokumentů, ale jedná se pouze o velmi malý segment univerza digitálních dokumentů

který zajišťuje jeho jedinečnost a vazbu na konkrétního vydavatele (identifikátor ISBN je transparentní – z jeho struktury lze identifikovat jak zemi, resp. skupinu zemí, tak konkrétního vydavatele), a nakonec vytištěn v knize. Vydaná kniha tak v sobě již obsahuje identifikátor, který knihovny či další subjekty mohou následně využít při akvizici (objednávání knih, zjišťování různých verzí), zpracování (katalogizace) nebo vyhledávání (uživatelské hledisko).

V této fázi končí práce všech subjektů s identifikátory. Výjimkou je národní agentura, která musí zaručit trvalou jedinečnost identifikátoru a jeho soulad s příslušnými pravidly. O identifikátor se však již dále nemusejí starat ani subjekty, které jej získávají (vydavatelé), ani subjekty, které jej využívají při zpracování dokumentů, ani subjekty, které se starají o zpřístupňování nebo dlouhodobou archivaci knih (různá oddělení knihoven). V rané fázi životního cyklu dokumentu končí potřeba koordinace mezi systémem trvalé identifikace a dlouhodobou správou a ochranou identifikovaných dokumentů. Příslušné instituce zajišťující dlouhodobou ochranu tištěných dokumentů nepotřebují ochranná opatření koordinovat se systémem trvalé identifikace, a proto je ochrana tištěných dokumentů v porovnání s digitálním světem mnohem méně náročná. Tištěná kniha je uložena v depozitáři a její fyzický nosič (kniha tvořená listy papíru, deskami apod.) je v porovnání s elektronickými nosiči (pevné disky, magnetické pásky, CD aj.), na nichž jsou uloženy digitální dokumenty, relativně vysoce odolný a má dlouhou životnost. Je také třeba dodat, že s ochranou tištěných dokumentů mají paměťové instituce dlouholeté praktické zkušenosti a specifická ochranná opatření je na knihách potřeba vykonávat obvykle až desítky let po jejich akvizici. Nosiče tištěných dokumentů mohou vydržet až stovky let. Dokonce i nosiče z 19. století tvořené kyselým papírem, u nichž se v současnosti dramaticky zvyšuje degradace, vydrží déle než elektronické nosiče.

Integrita tištěných dokumentů je již do značné míry zajištěna samotnými vlastnostmi fyzického objektu (tištěné knihy). Jednotlivé stránky knihy drží relativně dobře pohromadě ve vazbě, a pokud archivní exemplář nepůjčujeme čtenářům a pečujeme o něj ve vhodných podmínkách, integrita knihy obvykle není vážněji ohrožena. Bezpečnost uložení také zvyšuje archivace v několika identických fyzických kopiích na různých místech (různé exempláře stejného vydání archivované v různých knihovnách) a je prakticky vyloučeno, aby se najednou všechny tyto knihy zničily nebo nežádoucím způsobem změnily. Jak víme z digitalizační praxe, problém s nekvalitní předlohou lze poměrně snadno vyřešit výpůjčkou exempláře z jiné knihovny. Díky této vysoké odolnosti tištěných dokumentů zůstává také číslo ISBN stále na svém místě a nehrozí, že by mohlo náhle nějakým nedopatřením ze všech knih hromadně zmizet (smazat se, přepsat se apod.), a je tedy dlouhodobě dobře využitelné.

*Pro systém trvalé identifikace založený na ISBN je tedy klíčovou pouze raná fáze životního cyklu dokumentu. Jakmile je identifikátor s knihou vytištěn, další fáze nejsou pro zachování systému trvalé identifikace důležité. V analogovém (tištěném) světě jsou systém trvalé identifikace dokumentů a jejich dlouhodobá správa a ochrana navzájem nezávislé a nevyžadují souběžnou koordinaci. Na stabilitu systému trvalé identifikace tištěných dokumentů nemá správa a ochrana těchto dokumentů v knihovnách žádný vliv, a tím je správa tohoto systému relativně dobře zvládnutelná.*

V digitálním světě čelíme zcela jiné situaci. Systém trvalé identifikace digitálních dokumentů nelze oddělit od systému dlouhodobé správy a ochrany. Správa, ochrana a identifikace zde musejí být aktivně vykonávány po celý životní cyklus dokumentů. Trvalá identifikace v digitálním světě zpravidla začíná, stejně jako v případě tištěných dokumentů, před fází produkce dokumentů. Již zde je řada rozdílů. Digitální identifikátor může být přidělován manuálně, obvykle je však přidělován automatizovaně. Druhý případ je typický pro masovou digitalizaci, při které pracujeme s objemem dokumentů mnohonásobně převyšujícím počet dokumentů, kterým je přidělováno číslo ISBN. De facto každá digitalizující knihovna je v pozici vydavatele. Zatímco však klasický vydavatel většinou disponuje dostatečným týmem pracovníků, kteří dohlížejí na bezproblémovost produkce (korektoři, editoři apod.), digitalizující knihovna většinou nemůže nasadit odpovídající množství lidských zdrojů pro masové objemy digitalizace. Vyšší komplexita digitalizačního procesu má tak za následek, že během procesu přidělování digitálního identifikátoru hrozí vyšší chybovost než v popsáném systému ISBN.

Digitální identifikátor se obvykle nekládá do počítačových souborů reprezentujících obsah dokumentu, ale do přidavných *metadatových souborů*<sup>3</sup>. Ať již je však identifikátor zaveden jakýmkoli způsobem, platí, že všechny tyto počítačové soubory je nezbytné po celou dobu jejich životního cyklu aktivně ochraňovat při všech operacích s nimi prováděnými. Životnost a odolnost elektronických nosičů, na kterých jsou digitální dokumenty uloženy, se ani zdaleka neblíží životnosti papíru. Digitální dokumenty je proto nutné v relativně velmi krátkých časových intervalech z původních nosičů kopírovat na nové elektronické nosiče a do nových počítačových systémů.

Podobně není možné uchovávat digitální dokumenty po desítky let v té podobě, v jaké byly původně do digitálních repozitářů knihoven uloženy. Tištěná kniha zůstává v depozitáři v původním stavu (s tím, že postupem let nepatrně degraduje) a je v tomto stavu stále použitelná a čitelná. Digitální dokumenty však nemohou být trvale uloženy v původních počítačových formátech a zůstat v nich uživatelům stále přístupnými. Původní formáty totiž rychle zastarávají, a tak je potřeba dokumenty v relativně krátkých intervalech převádět do formátů nových. Každý takový převod je rizikem pro udržení autenticity původně vloženého digitálního dokumentu.

Integrita digitálního dokumentu je v porovnání s dokumentem tištěným vystavena také mnohem vyššímu stupni ohrožení. Digitální dokument skládající se z desítek nebo stovek jednotlivých počítačových souborů (např. stostránková tištěná kniha může být zdigitalizována do 100 souborů ve formátu TIFF) je potřeba uchovávat jako celek při všech operacích s nimi prováděných, přičemž i při těch nejjednodušších z nich (např. kopírování) hrozí, že některé (ne-li všechny) soubory mohou být smazány, přepsány nebo ztraceny. Toto riziko přirozeně neohrožuje jen archivaci dokumentu, ale také systém trvalé identifikace. Uživatel systému ISBN se pravděpodobně nikdy nedostane do situace, kdy by kniha, kterou hledá, již neexistovala, nebo ji nebylo možné v žádné knihovně najít. U digitálního identifikátoru však snadno může nastat situace, že identifikátor označuje dokument, který buď již neexistuje (byl smazán apod.), nebo byl přesunut jinam a není možné jej zpřístupnit, protože informaci o místě jeho nového uložení odpovědná strana do systému trvalé identifikace nedodala. K tomu všemu musíme připočítat fakt, že digitální dokumenty ani jejich metadata nemohou být přístupné uživateli jinak než prostřednictvím komplexní informační technologie. Z digitálních dat samotných katalogizátor nic nevyčte, což platí i o identifikátoru, který si v digitálním dokumentu nikdo nemůže nalistovat a přečíst jako v tištěné knize označené číslem ISBN. Informační technologie jsou však také rovněž ohrožovány prudkým technologickým vývojem, selháváním zařízení nebo výpadky elektrického proudu, a tak je potřeba se o ně nepřetržitě starat (doslova po dobu 24 hodin denně) a v případě technologického zastarávání přecházet na nové systémy. Jako poslední problém uveďme, že u digitálních dokumentů je pravděpodobnost poškození lidským faktorem, ať již úmyslného (hackerský útok, sabotáž apod.), nebo náhodného (smazání omylem apod.) tak vysoká, že to nemá v dějinách archivace dokumentů obdoby.

Ze všech těchto skutečností vyplývá, že digitální dokumenty jsou vystaveny mnohem vyššímu riziku selhání ochrany a z něho plynoucí ztráty vazby mezi identifikátorem a dokumentem. Na rozdíl od tištěného světa může v digitálním prostředí snadno nastat situace, že identifikátor již nedokáže nic identifikovat nebo zpřístupňovat. To vše výrazně ztěžuje budování důvěryhodnosti systému trvalé identifikace.

Pro dlouhodobou stabilitu systému trvalé identifikace digitálních dokumentů a udržení důvěry v jeho možnosti je tedy nutné tento systém provázat s řízením celého životního cyklu dokumentu. Je nutné nepřetržitě koordinovat trvalou správu a ochranu dat se systémem jejich identifikace. Přidělením digitálního identifikátoru starosti o systém trvalé identifikace nekončí, ale vlastně teprve začínají. Ať již má být takový systém národním,

<sup>3</sup> Nemusí tomu tak být vždy. Někteří producenti mohou vkládat identifikátor přímo do počítačového souboru reprezentujícího obsah dokumentu. Digitální identifikátor zde může být uveden na stejném místě, jako je uvedeno ISBN v tištěných publikacích. To se týká například některých vědeckých publikací, ukládaných do jediného souboru (například ve formátu PDF). Konkrétním příkladem může být identifikátor „urn:nbn:de:0008-2006060703“, uvedený na první stránce známého katalogu kritérií programu nestor pro certifikaci digitálních repozitářů (viz [nestor, 2006]).

mezinárodním nebo jen institucionálním, je nezbytné, aby na něm trvale a soustavně spolupracovaly všechny zúčastněné strany – jak registrační autority, tak vydavatelé nebo oddělení zajišťující dlouhodobou archivaci nebo zpřístupňování. Systém trvalé identifikace je nutné koordinovat jak s digitalizačními aktivitami (vhodný postup při přidělování identifikátoru, výběr formátu vhodného pro archivaci apod.), tak s provozem digitálních knihoven (zpřístupňování dokumentů uživatelům) a digitálních repozitářů (dlouhodobá ochrana digitálních dokumentů). Tyto velmi odlišné činnosti provozuje celá řada různých oddělení a institucí, a nejsou tedy v gesci správce systému, který identifikátor přiděluje. Existuje očekávání, že důvěryhodnost celého systému zajistí samostatně a pro všechny jeho správce. Je zřejmé, že takovéto očekávání je iluzorní. Systém trvalé identifikace digitálních dokumentů může dobře fungovat pouze potud, pokud všechny zúčastněné strany, jejichž dokumentům jsou identifikátory přidělovány, budou v součinnosti s řízením systému trvalé identifikace samy dlouhodobě zajišťovat a zaručovat správu a ochranu těch identifikovaných dokumentů, které jsou v jejich gesci. V digitálním světě tedy platí, že udržet důvěryhodnost systému trvalé identifikace je náročnější úkol než například v systému ISBN. Je zde nutná vysoká míra vzájemné spolupráce zúčastněných stran v dlouhodobém horizontu.

## Některá základní fakta

Srovnání mezi vedením systémů trvalé identifikace tištěných a digitálních dokumentů nám může posloužit jako základ pro plánování důvěryhodného systému pro přidělování a správu digitálních identifikátorů. Pokusíme se shrnout některé základní požadavky na tento systém do dvanácti základních tezí, přičemž také upozorníme na některé mylné představy vztahující se k budování tohoto systému.

**1. Generátor jedinečných znakových řetězců (ať již jsou tvořeny pouze čísly, nebo obsahují i jiné typy znaků) je přirozeně základní složkou systému trvalé identifikace. Zajišťuje jedinečnost digitálních identifikátorů, tedy skutečnost, že již existující identifikátor nebude znovu přidělen jinému dokumentu. Generátor musí být spojen s centrálním registrem identifikátorů, jehož funkcí je trvale udržovat seznam všech znakových řetězců přidělených v daném kontextu. Jedině tak je možné zajistit, že nedojde k chybě, tedy přiřazení téhož identifikátoru odlišným dokumentům. Z toho také plyne, že registr musí udržovat záznamy o všech identifikátorech, tedy například i o těch, které označují dokument, který již neexistuje (např. byl smazán). Samotné generování znakových řetězců je však k ničemu, pokud není součástí komplexního systému trvalé identifikace. Pouhá existence takového generátoru dokonce může vyvolávat falešný dojem, že jeho spuštěním je problém trvalé identifikace vyřešen.**

**2. Přesměrovávací služba, tedy služba, která nás na základě zadání identifikátoru do systému odkáže na aktuální umístění identifikovaného dokumentu v internetové síti, nemůže fungovat bezchybně, pokud není zajištěna její administrativa. Administrativa vyžaduje, aby subjekty, které nechaly svým dokumentům přidělit digitální identifikátory, do systému soustavně dodávaly aktuální informace o internetové adrese, na které se identifikovaný digitální dokument v daný okamžik nalézá. Každá změna internetové adresy v digitální knihovně musí být oznámena centrálnímu systému – ať již automatizovaně, nebo manuálně. To vždy vyžaduje alespoň minimální úroveň technologických a personálních zdrojů na straně zúčastněných partnerů. Pokud instituce takovými zdroji nedisponuje, nemůže se systému účastnit.**

**3. Syntax digitálního identifikátoru je v podstatě druhořadý. Je samozřejmé, že pravidla pro syntax jsou důležitá pro správné fungování systému nebo pro jeho bezproblémovou komunikaci s jinými systémy. Ale domnívat se, že k vytvoření spolehlivého systému trvalé identifikace postačí rozhodnutí jedné instituce o tom, jaký bude syntax identifikátoru, je bláhové. Pravidla pro syntax ostatně již definují některé globální identifikační standardy, a je tedy možné je převzít a zajistit si tak jedinečnost identifikátoru v mezinárodním měřítku. Příkladem je syntax v standardu URN:[Hakala, 2001].**

**4. V procesu digitalizace** je klíčové propojit systémy identifikace tištěných a digitálních dokumentů. Nutně proto již musí existovat nějaký kvalitní systém trvalé identifikace tištěných dokumentů. Systémy ISBN nebo ISSN sice fungují dobře, ale pokrývají jen část univerza tištěných dokumentů. Pro ty tištěné dokumenty, které nedisponují těmito trvalými identifikátory, byly v některých zemích zavedeny speciální typy analogových identifikátorů. Příkladem je identifikátor národní bibliografie (v MARC21 je mu vymezeno pole 015). Zároveň platí, že pokud je systém trvalé identifikace tištěných dokumentů zatížen problémy vzniklými při zpracování dokumentů (nekonzistentní přístup ke katalogizaci, problematické konverze, neexistence identifikátorů u starých děl apod.), nelze očekávat, že tyto chyby vyřeší systém trvalé identifikace digitálních dokumentů. Digitální identifikátory lze při digitalizaci pouze navazovat na stávající analogové identifikátory tištěných zdrojů. Pokud systém trvalé identifikace tištěných dokumentů nefunguje dobře (nebo dokonce vůbec neexistuje), neexistuje způsob, jak by digitální identifikátory mohly tuto situaci vyřešit. Zásadní je proto koordinace obou systémů trvalé identifikace již od rané fáze digitalizace, jejíž součástí může být například rekatologizace problematických záznamů předcházející proces přidělení digitálního identifikátoru. Je zřejmé, že takové činnosti nelze plně automatizovat a vyžadují, zejména při masové digitalizaci, například dostatečný počet katalogizátorů.

**5. Žádné softwarové řešení** není samo o sobě dostačující. Software je bezpochyby nezbytnou technickou páteří celého systému, ale spoléhat se na to, že by při budování systému trvalé identifikace digitálních dokumentů mohl nějaký software nahradit lidské pracovníky nebo neexistující pravidla, není možné. Bylo by to podobné jako očekávat, že software nahradí zaměstnance příslušných národních agentur přidělujících čísla ISBN. Software je v celém snažení o dlouhodobou identifikaci digitálních dokumentů podstatnou pomůckou, ale pouze a jen pomůckou.

**6. Jedna instituce** nemůže udržovat národní systém trvalé identifikace digitálních dokumentů sama. Jak jsme již uvedli, systém trvalé identifikace je nutně v dlouhodobém horizontu soustavně koordinovat s dalšími digitálními aktivitami, které nejsou v gesci jedné instituce. I při sebevíce centralizovaném systému je nezbytné, aby všechny zúčastněné instituce disponovaly vlastními technickými a personálními prostředky a trvale dodržovaly pravidla a podmínky (včetně přidělených odpovědností) jak pro správu trvalých identifikátorů, tak pro správu jimi označovaných dokumentů. Každá instituce, která provozuje digitální knihovnu zpřístupňující dokumenty označované národním digitálním identifikátorem, bude muset vložit nemalé úsilí do údržby národního systému.

**7. Není zdaleka nutné,** aby všechny instituce využívaly jen jeden systém. Stejně tak není nutné, aby se všechny instituce účastnily národního systému. Národní knihovna ČR se rozhodla využít pro systém trvalé identifikace standard URN:NBN. Ten však nemusí být vhodný pro všechny dokumenty. Existují jiné zavedené a stabilní systémy, které jsou schopny zajistit globální jednoznačnost identifikovaných dokumentů a které lze, s určitou mírou přidaného úsilí, implementovat pro české prostředí. Příkladem je systém Handle (<http://handle.net/>), který již také některé české instituce využívají.

**8. Rovněž není nikde řečeno,** že instituce musejí zavádět některé ze stávajících globálních identifikačních standardů. Je samozřejmě v mnoha ohledech výhodné připojit se do již fungující infrastruktury, stejně tak je však možné vytvořit si vlastní identifikační systém, který může zcela uspokojit požadavky instituce v rámci jejího kontextu. Příkladem je Australská národní knihovna (National Library of Australia), která provedla analýzu dostupných řešení a dospěla k závěru, že pro své specifické využití digitálních identifikátorů není vhodné žádné z nich. Vytvořila si proto vlastní systém pro trvalou identifikaci, který propojila se správou, katalogizací a archivací svých digitálních dokumentů [National Library of Australia, 2010].

**9. Zásadní problém** nepředstavuje ani situace, kdy jeden digitální dokument má více identifikátorů (za předpokladu, že každý z nich je užíván v jiném kontextu). Již dnes je

běžné vyhledávat tytéž dokumenty podle několika různých identifikátorů, například podle ISBN a ISSN v případě ročenek.

**10. Identifikátorová metadata** jsou klíčovým prvkem jakéhokoliv systému trvalé identifikace. Tato metadata jsou připojována k přidělenému identifikátoru a pomáhají dlouhodobé administraci systému. Záleží na konkrétních pravidlech systému, jak podrobná musejí tato metadata být. Minimálním požadavkem na metadata jsou údaje o aktuálním umístění digitálních dokumentů v internetové síti, protože bez nich není možné využít digitální identifikátor jako prostředníka pro přímé zpřístupnění dokumentu. Lze říci, že čím podrobnější metadata budou instituce systému trvalé identifikace poskytovat, tím více funkcí bude tento systém nabízet. A patrně také lépe fungovat. Příkladem může být globální identifikační systém DOI, který je nejpropracovanějším a nejlépe fungujícím systémem trvalé identifikace v oblasti digitálního dědictví. Jen uživatelská příručka DOI obsahuje téměř dvě stě stran textu specifickou terminologií popisujícího interní systém pravidel a odpovědností, včetně podrobného datového modelu [Paskin, 2006]. Instituce, která chce identifikátor DOI využívat, musí centrálnímu správci poskytnout právě řadu podrobných metadata vztahujících se k identifikátoru, resp. označovanému dokumentu. Systém DOI je zpoplatněn, příspěvky jednotlivých účastníků se odvíjejí od počtu jejich identifikovaných dokumentů.<sup>4</sup> To dobře poukazuje na skutečnost, že správa komplexního systému trvalé identifikace vyžaduje rezervoár různých zdrojů (personálních, technických apod.) a tyto zdroje také vždy něco stojí. Systém DOI rovněž ilustruje, že čím vyšší je objem identifikovaných dokumentů, tím náročnější, a tedy také dražší je dlouhodobá údržba jejich systému.

**11. Pro konkrétní systém trvalé identifikace je důležité zavést konkrétní datový model. Datový model vymezuje, co přesně se bude v daném systému identifikovat. Bude digitální identifikátor označovat jednotlivý počítačový soubor? Nebo digitální dokument, tak jak je definován v tomto článku? Nebo bude předmět identifikace vymezen jinou definicí?** Datový model si jednotliví aktéři musejí navrhnout dříve, než bude zahájeno přidělování digitálních identifikátorů, jinak bude systém zatížen nekonzistentnostmi a nebude dobře fungovat.

**12. Trvalá identifikace dokumentů, které nebudou dlouhodobě archivovány a trvale zpřístupňovány (alespoň některým) uživatelům, je nesmyslná. Jaký smysl má zavádět trvalý identifikátor pro dokument, u něž není zaručena (nebo se ani neočekává) jeho dlouhodobá existence? Každému rozhodnutí o tom, že instituce přidělí nebo nechá přidělit svému dokumentu trvalý identifikátor, tak musí předcházet rozhodnutí o tom, že instituce tento dokument také zamýšlí dlouhodobě archivovat a zpřístupňovat, a zvážení toho, nakolik je takového závazku schopna dosáhnout.**

## Budování národního systému

V mnoha zemích předcházely vytvoření systému trvalé identifikace digitálních dokumentů několikaleté přípravy v podobě vstupních výzkumných projektů. Projekty se zaměřovaly na základní konceptualizaci systému trvalé identifikace, vytvoření vhodné metodiky pro jeho budování a správu a produkci pilotních verzí podpůrných softwarových nástrojů.

V Austrálii probíhal v rámci národního programu pro výzkumné repozitáře ARROW II v letech 2006-2007 výzkumný projekt PILIN (Persistent Identifier and Linking Infrastructure; <http://www.pilin.net.au/>). Jeho cílem bylo vytvořit základ pro jednotnou národní infrastrukturu trvalé identifikace zaměřenou na vědecký sektor. Jako výchozí systém pro infrastrukturu byl vybrán systém Handle. V rámci projektu vznikla rozsáhlá metodika pro systém trvalé identifikace a také řada specifických softwarových nástrojů. Jeden z výstupů projektu (část obecného doporučení) byl nedávno publikován také v českém překladu [Nicholas, 2009].

<sup>4</sup> Případně podle počtu dalších identifikovaných entit. Identifikátor DOI totiž může označovat nejen dokument, ale například také osoby aj.

V Německu byl v letech 2002-2005 řízen podobný projekt pod vedením Německé národní knihovny (Deutsche Nationalbibliothek). Nesl název EPICUR (Enhancement of Persistent Identifier Services-Comprehensive Method for Unequivocal Resource Identification; <http://www.d-nb.de/wir/projekte/epicur.htm>) a jeho cílem bylo vytvořit systém pro trvalou identifikaci a vyhledávání digitálních verzí závěrečných vysokoškolských prací a dalších online publikací. EPICUR vyšel ze standardu URN:NBN a vytvořil softwarový nástroj pro správu a využívání digitálních identifikátorů s webovým rozhraním (<http://www.persistent-identifier.de/>).

Dobrým příkladem konceptualizace systému trvalé identifikace v rámci širšího kontextu dlouhodobé ochrany digitálních dokumentů je německý program nestor (Network of Expertise in Long-term STORage of Digital Resources / Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen; <http://www.langzeitarchivierung.de/>). Nestor je celonárodní kooperativní iniciativa pro německé knihovny, archivy a muzea a odborníky, kterou od roku 2003 vede Německá národní knihovna. Cílem nestoru je vytvořit expertní síť a trvalý rámec pro všechny otázky dlouhodobé ochrany digitálních dokumentů (tedy včetně trvalé identifikace) v německém národním kontextu. Nejznámějším výstupem nestoru je katalog kritérií pro certifikaci důvěryhodných repozitářů [nestor, 2006]. Digitální repozitáře tvoří nezbytný základ pro vývoj systému trvalé identifikace z hlediska nutnosti úzkého provázání systému trvalé identifikace s dlouhodobou ochranou digitálních dat. Před rokem vydal nestor také doporučení pro důvěryhodný systém trvalé identifikace [nestor, 2009].

V Česku dosud nebyl podobný celonárodní projekt realizován. Za první menší krůčky na cestě k národnímu systému trvalé identifikace můžeme však považovat dvě události. První je zavedení trvalého identifikátoru pro dokumenty české národní bibliografie, který byl nazván čČNB (číslo České národní bibliografie). Identifikátor čČNB slouží zejména k jednoznačné identifikaci starých tištěných dokumentů, které nemají identifikátor ISSN nebo ISBN, a tak je implementace čČNB významným posunem na cestě k provázanosti mezi trvalou identifikací tištěných a digitálních dokumentů. Identifikátor čČNB je v rámci chystané masové digitalizace Národní knihovny ČR (<http://www.ndk.cz/>) jedním z řešení, jak propojit trvalé identifikátory digitalizovaných dokumentů s jejich tištěnými zdroji pocházejícími ze souboru naší národní bibliografie.

Druhým krokem je v současnosti probíhající interní proces Národní knihovny ČR zaměřený na vývoj pilotní verze softwarového systému, který podpoří implementaci prvotní konceptualizace systému trvalé identifikace digitálních dokumentů vyvíjeného v Národní knihovně ČR a založeného na standardu URN:NBN. Tato aktivita je omezena na určitý typ dokumentů a vybrané funkce systému.

Zmíněné dva kroky jsou řešením institucionálního kontextu trvalé identifikace digitálních dokumentů v ČR, v tomto případě v rámci Národní knihovny ČR. Pro celonárodní systém je potřeba podniknout řadu dalších kroků. Je nutné vypracovat komplexní metodiku pro obecné využití a správu digitálních identifikátorů v rámci ČR, která by mohla posloužit jakékoliv paměťové instituci pro rozhodování o budování systému trvalé identifikace digitálních dokumentů. Také je potřeba vytvořit podrobnou konceptualizaci pro zvažovaný národní systém trvalé identifikace. Tento článek je pouze úvodním upozorněním na šíří problému a některé jeho souvislosti. Podrobnou konceptualizaci problému, která poskytne návod na komplexní řešení, musí vytvořit projektově řízený tým složený z různých odborníků v dané oblasti. Tento tým by měl také vytvořit celonárodní rámec pro účast různých zainteresovaných stran a dojít ke konsensu na vymezení jejich dílčích odpovědností.

Z výše uvedené potřeby koordinovat národní systém trvalé identifikace se systémy dlouhodobé správy a ochrany digitálních dokumentů je tak logické, že dalším krokem Národní knihovny ČR je účast v širším národním projektu v rámci Programu aplikovaného výzkumu a vývoje národní a kulturní identity (NAKI), nazvaném „Nástroje pro evidenci, archivaci, agregaci a zpřístupnění národního kulturního dědictví v digitální formě“. Projektu se účastní Národní knihovna ČR, Knihovna Akademie věd ČR a Moravská zemská knihovna a problém trvalé identifikace zde bude řešen jako jedna součást projektu v širším kontextu celonárodních digitálních aktivit.



## Závěr

Budování důvěryhodného systému trvalé identifikace digitálních dokumentů je výrazně náročnější než budování podobného systému pro tištěné dokumenty. Pokud má takový systém fungovat v dlouhodobém horizontu podobně dobře jako systém ISBN, je potřeba, aby byl trvale koordinován s dlouhodobou správou a ochranou digitálních dokumentů, což je požadavek, který neexistuje ve světě dokumentů tištěných. Na správu takového systému se musejí aktivně podílet všechny instituce, které digitální identifikátory přidělené tímto systémem chtějí využívat. Tato správa vyžaduje nemalý podíl personálních, technických nebo finančních zdrojů. Při budování systému musí být dojednáno rozložení odpovědností v rámci všech zúčastněných stran, včetně míry jejich technologické, personální, finanční a jiné účasti. Vybudování systému musí předcházet komplexní analýza problémů (a to nejen problémů uvedených v tomto článku), potřeb a očekávání spjatých s digitálními identifikátory a vytvoření podrobné konceptualizace systému. Budování důvěryhodného národního systému trvalé identifikace digitálních dokumentů by podle našeho názoru vzhledem ke své komplexitě mělo být řešeno v rámci širšího národního projektu, zaměřeného nejen na problém trvalé identifikace, ale obecně na dalších vývoj digitálních aktivit českých knihoven.

## Literatura:

HAKALA, J. 2001. *Using national bibliography numbers as Uniform Resource Names* [online]. 2001 [cit. 2010-05-01]. Internetový standard RFC 3188. Dostupné z WWW: <<http://www.ietf.org/rfc/rfc3188.txt>>.

National Library of Australia. 2010. *National Library of Australia* [online]. 2010 [cit. 2010-06-25]. Activities – Persistent identifiers. Dostupné z WWW: <<http://www.nla.gov.au/initiatives/persistence.html>>.

Nestor Working Group on Trusted Repositories Certification. 2006. *Catalogue of Criteria for Trusted Digital Repositories*. Version 1 (draft for public comment), nestor – materials 8. Frankfurt am Main : nestor – Network of Expertise in long-term STORage – c/o Deutsche Nationalbibliothek, 2006. iii, 39 s. Dostupný také z WWW: <[http://files.d-nb.de/nestor/materialien/nestor\\_mat\\_08-eng.pdf](http://files.d-nb.de/nestor/materialien/nestor_mat_08-eng.pdf)>. URN:NBN:de:0008-2006060710.

Nestor working group long-term preservation standards. 2009. *Catalogue of Criteria for Assessing the Trustworthiness of PI Systems* [online]. Draft for public comment. Göttingen : Niedersächsische Staats und Universitätsbibliothek Göttingen, 2009 [cit. 2010-05-01]. Dostupné z WWW: <[http://files.d-nb.de/nestor/materialien/nestor\\_mat\\_13\\_en.pdf](http://files.d-nb.de/nestor/materialien/nestor_mat_13_en.pdf)>. URN:NBN:de:0008-20080710227.

NICHOLAS, Nick; WARD, Nigel; BLINCO, Kerry. 2009. Kontrolní seznam pro strategii zajištění perzistence identifikátorů. *Knihovna – knihovnická revue*. 2009, roč. 20, č. 2, s. 54-62. ISSN 1801 3252.

PASKIN, Norman. *The DOI® Handbook* [online]. edition 4.4.1. Oxford : International DOI Foundation, October 5, 2006 [cit. 2010-03-01]. Dostupný z WWW: <[http://www.doi.org/handbook\\_2000/DOIHHandbook-v4-4.pdf](http://www.doi.org/handbook_2000/DOIHHandbook-v4-4.pdf)>.