

## Kontrolní seznam pro strategii zajištění perzistence identifikátorů

Bc. Ladislav Cubr, Mgr. Jan Hutař, Mgr. Marek Melichar / NK ČR  
ladislav.cubr@nkp.cz / jan.hutar@nkp.cz / marek.melichar@nkp.cz

NICHOLAS, Nick; WARD, Nigel; BLINCO, Kerry. A Policy Checklist for Enabling Persistence of Identifiers. *D-Lib Magazine*. 2009, vol. 15, no. 1/2. ISSN 1082-9873. DOI: 10.1045/january2009-nicholas. Dostupný z WWW: <<http://www.dlib.org/dlib/janu-ary09/nicholas/01nicholas.html>>.

### Resumé:

Jedním z hlavních úkolů projektu PILIN (Infrastruktura pro odkazování prostřednictvím perzistentních identifikátorů) bylo vytvoření konceptuálního rámce pro správu identifikátorů. Základním zjištěním projektu bylo, že administrativa a strategie jsou v oblasti trvalé identifikace mnohem důležitější než technologie, které systém pro správu identifikátorů používá. Každá strategie trvalé identifikace musí odpovědět na následující otázky: čemu chceme identifikátory přidělovat; jak je budeme používat; jak dlouho chceme zaručit jejich funkčnost.

K rozhodnutí o tom, co chceme identifikovat, musíme mít dobrý informační model. Správce identifikátorů musí mít jasno v tom, co lze identifikovat jako samostatný objekt (nejde jen o digitální soubory, často chceme identifikovat abstraktní entity, jako jsou díla, verze, reprezentace, agregace), a které z těchto objektů musíme udržet dlouhodobě dostupné. Bez toho, že budeme mít zcela jasno v tom, co bychom měli identifikovat, nemá závazek trvalé dostupnosti identifikovaných objektů žádný význam.

Projekt PILIN vytvořil šestibodový kontrolní seznam pro integraci identifikátorů do správy informačních zdrojů, který má zajistit, že každý navrhovaný systém trvalé identifikace naplní naše očekávání.

**Klíčová slova:** trvalé identifikátory – kontrolní seznam – projekt PILIN.

### Summary:

One of the main tasks of the Persistent Identifier Linking Infrastructure (PILIN) project on persistent identifiers was to establish a policy framework for managing identifiers and identifier providers. A major finding from the project was that policy is far more important in guaranteeing persistence of identifiers than technology. Key policy questions for guaranteeing identifier persistence include: what entities should be assigned persistent identifiers, how should those identifiers be exposed to services, and what guarantees does the provider make on how long various facets of the identifiers will persist.

To make an informed decision about what to identify, information modelling of the domain is critical. Identifier managers need to know what can be identified discretely (including not only concrete artefacts like files, but also abstractions such as works, versions, presentations, and aggregations); and for which of those objects it is a priority for users and managers to keep track. Without working out what actually needs to be identified, the commitment to keep identifiers persistent becomes meaningless.

To make sure persistent identifiers meet these requirements, the PILIN project has formulated a six-point checklist for integrating identifiers into information management, which we present here.

**Keywords:** persistent identifiers – checklist – the PILIN project.

## Úvod

Organizace a jednotlivci jsou dnes stále více závislí na zdrojích, které jsou k dispozici on-line. Proto vyvstává potřeba, aby byly tyto zdroje dobře spravovány a byly stále k dispozici po dlouhou dobu. Zásadní význam pro zajištění dlouhodobé dostupnosti zdrojů má udržování trvalých identifikátorů. Každý plán trvalého zpřístupňování musí tedy počítat s trvalými identifikátory.

Projekt PILIN (Infrastruktura pro odkazování prostřednictvím perzistentních identifikátorů) byl financován v letech 2006 až 2008 s cílem podpořit využívání globální infrastruktury perzistentních identifikátorů v Austrálii, zejména v oblasti institucionálních repozitářů. Projekt měl samozřejmě také konkrétní technologický dopad, mezi jeho výstupy patří software a požadavky na služby.<sup>1</sup>

Je však dobře známo, že k zajištění perzistence nestačí technologie. Závislost na konkrétní technologii je překážkou pro zajištění trvalé dostupnosti. Trvalost identifikátorů závisí mnohem více na spolehlivé strategii zajišťující, že někdo udržuje digitální zdroje on-line a aktualizuje odkazy na ně po jasně specifikované době. Uživatele je třeba přesvědčit, že máme dobrou strategii, která zajistí dlouhodobou dostupnost zdrojů, a také ji skutečně mít: skutečným cílem persistence je vybudovat důvěru uživatelů v identifikátory.

Hlavním cílem projektu PILIN bylo vytvořit rámec pro správu identifikátorů a poskytovatele identifikátorů. Strategický rámec obsahuje návody a seznam oblastí, které je třeba brát v úvahu při tvorbě a používání trvalých identifikátorů; příklady strategických dokumentů a dokumentaci požadavků na identifikátory v některých komunitách. Pokud chceme, aby byl takový rámec nezávislý na konkrétní technologii (a dokázal přežít změny v technologiích), je nutné vytvořit na technologii nezávislý model identifikátorů a identifikačních služeb. Model PILIN obsahuje ontologii, slovníček pojmů a model služeb pro uživatele. Nezávislost na technologii znamená, že PILIN dogmaticky nelpí na žádném konkrétním systému identifikace. PILIN nakonec pro implementaci identifikátorů využil řadu různých technologií, zejména Handle System® jako model identifikačních služeb. Projekt PILIN také vytvořil dokumentaci pro trvalé využívání HTTP URI.<sup>2</sup>

PILIN byl mezi-komunitní projekt. Byl zaměřen na potřeby výzkumných institucí, knihoven a vzdělávacích institucí. V průběhu projektu byly shromážděny požadavky na perzistentní identifikátory, i na to, jak by měly být spravovány a využívány.

Když projekt shromáždil požadavky různých stakeholderů (dočasných držitelů) na správu a využití identifikátorů, vyvstalo několik základních otázek týkajících se problému jak vytvořit a udržet perzistenci identifikátorů. Uvědomili jsme si, že je potřeba vytvořit kontrolní seznam (*checklist*), pomocí kterého bude možné porovnat různá řešení systému identifikace.

Znovu a znovu se objevovaly tyto otázky:

- Q1. Co je myšleno slovem „trvalý“?**
- Q2. Jak dlouho by měly identifikátory existovat?**
- Q3. Co bychom měli trvale identifikovat?**
- Q4. Jaké kroky je třeba podniknout k zajištění perzistence identifikátorů?**
- Q5. Jaké technologie bychom měli použít k zajištění perzistence identifikátorů?**

První z těchto otázek nás přiměla k detailnímu modelování identifikátorů a jejich vlastností pro účely správného vymezení jejich perzistence. Toto modelování bylo velmi náročné. V dalším textu se pokusíme shrnout naše úvahy o perzistenci. Na poslední otázku neexistuje jediná správná odpověď. Záleží totiž na tom, jak odpovíme na otázky předcházející a jak budou identifikátory komunikovat s jinými počítačovými systémy. Prostřední tři otázky směřují přímo k jádru plánování trvalosti. Pro jejich řešení projekt vytvořil „šestibodový program modelování informačních toků a služeb“, který zde popíšeme.<sup>3</sup>

Šestibodový program je definován na abstraktní úrovni a musí být doplněn konkrétními detaily specifickými pro danou oblast. Můžeme ho považovat za kontrolní seznam pro úvahy nad otázkami Q2, Q3 a Q4 v rámci daného kontextu. Q3 závisí na zapojení/názoru poskytovatelů zdrojů a Q4 na zapojení správce zdrojů, Q2 se zabývá potřebami koncových uživatelů, tedy vyjasněním toho, co systém poskytuje těmto uživatelům. Začneme se zabývat otázkami Q3 a Q4. Odpověď na otázku Q2 závisí na odpovědi na Q1. Všechny odpovědi pak můžeme nakonec spojit dohromady.

## Co bychom měli trvale identifikovat?

Problém trvalosti identifikátorů, jímž se zde zabýváme, je třeba oddělit od problému trvalosti zdrojů, jež identifikátory označují. Následující debata o perzistentních identifikátorech předpokládá, že identifikované zdroje mají trvalý význam a trvalé identifikátory mají zajistit, aby také zůstaly trvale dostupné. Plány, které mají zajistit, aby zdroje byly stále k dispozici (např. plán správy dat uvedený níže), je třeba při úvahách o perzistentních identifikátorech vzít v potaz.

Samy od sebe však perzistentní identifikátory nevzniknou. Zajistit persistenci znamená zaručit uživatelům, že identifikátory jsou stále udržovány v aktuálním stavu, a to vyžaduje trvalý závazek zpřístupitelnosti zdrojů. Aby takový závazek měl nějaký význam, není možné pokoušet se identifikovat vše, co v dané oblasti vzniká. Je potřeba rozhodnout, které zdroje v dané oblasti musí mít perzistentní identifikátory.

Doporučujeme upřednostňovat zdroje, které budou používány, citovány, popisovány, pořádány nebo jinak využívány komunitou uživatelů po dobu, kterou označíme v definici trvalosti. O tom, co má dlouhodobý význam a dostane trvalý identifikátor, nemůže rozhodovat správce identifikátorů. Je potřeba zapojit uživatele (často jde o představitele poskytovatelů zdrojů), kteří mají dobrou představu o využívání zdrojů a odkazování na ně. Také je nutné zapojit správce zdrojů, který je bude udržovat dostupné a v dobrém stavu, který určí, jak dlouho je realistické zaručit jejich dostupnost a jaká mohou být technická omezení jejich zpřístupňování (alespoň dokud jsou on-line).

Rozhodnutí, co přesně chceme identifikovat, souvisí s informačními modely používanými v dané oblasti. Perzistentní identifikátory mohou být spojeny nejen s konkrétními objekty, jako jsou digitální soubory, ale i s abstraktními entitami, které mohou být využívány pro vyhledávání a správu zdrojů. Abstraktními entitami mohou být například kategorie zdrojů (např. díla, vyjádření nebo provedení v terminologii FRBR)<sup>4</sup>, skupiny, podskupiny nebo transformace zdrojů. Definice a správa těchto abstraktních entit vyžaduje modelování. Správci a poskytovatelé zdrojů musí vědět, do jaké úrovně abstrakce lze jít a jak by měly být převedeny do konkrétních reprezentací.

Otázka „Co bychom měli trvale identifikovat?“ tedy odpovídá dvěma strategiím, které vyžadují zapojení poskytovatelů zdrojů:

1. Zjistěte, jaké všechny objekty ve vaší oblasti mohou být identifikovány pomocí identifikátorů (co ve vaší oblasti existuje v abstraktní a konkrétní rovině) a jak spolu navzájem souvisí.
2. Rozhodněte, jaké objekty ve vaší oblasti budou označeny trvalými identifikátory (tedy co je v dané oblasti důležité).

Tyto strategie odpovídají prvním dvěma bodům našeho kontrolního seznamu:

- I. Vytvořte informační model platný pro vaši oblast.
- II. Do tohoto informačního modelu začleňte perzistentní identifikátory.

Explicitní informační model umožňuje rozlišovat mezi konkrétními a abstraktními entitami, umožňuje znázornit a popsat vztahy mezi nimi, např. vyjádření verzí, kopírování, transformace nebo míru granularity.

## Jaké kroky je třeba podniknout k zajištění trvalosti identifikátorů?

Komunita uživatelů očekává, že perzistentní identifikátory budou poskytovat trvalou možnost vyhledávání a zpřístupňování zdrojů. Kromě toho je také třeba zohlednit činnost správců dat: s označenými zdroji by mělo být nakládáno tak, aby se minimalizoval dopad na trvalost identifikátorů. Například každou změnu lokace digitálního objektu by měla doprovázet odpovídající aktualizace identifikátoru, aby byla zaručena trvalost identifikátoru. Lepší by však bylo, kdyby se změny lokace digitálních objektů prováděly jen zřídka. Také aktualizace digitálního objektu by neměla vždy vyžadovat aktualizaci identifikátoru.

Správa identifikátorů by měla být součástí pracovních postupů pro správu digitálních zdrojů. Pokud například používáme pro správu zdrojů jeden identifikátor („lokální“), ale pro správu identifikátorů používáme jiný identifikátor, který je trvalý, pak jsou správa zdrojů a správa identifikátorů odděleny. Tento přístup, který je podobný systému DNS, je běžně rozšířen. URL nebo klíč zdroje se obvykle používá k přímé manipulaci se zdroji, a to jako identifikátor lokace. Zde jde přesněji řečeno o přístupové klíče (lokátory) pro získání zdrojů místními systémy. Mohou být sice použity i jako identifikátory, protože spojují jméno (přístupový klíč) se zdrojem, ale jejich trvalost je zaručena pouze v případě, že je zaručen trvalý způsob jejich uložení. Perzistentní identifikátory (jako HANDLE, PURL, a „cool“ URI)<sup>5</sup> se naproti používají pro nepřímé označení zdrojů a jejich lokace de facto identifikují nepřímo.

Z těchto důvodů musíme udržet trvalou vazbu mezi správou identifikátorů a správou zdrojů. V případě, že se identifikátor lokace změní, pak musí být aktualizována i metadata pro perzistentní identifikátor. Provádění těchto aktualizací je základní podmínkou zajištění trvalé identifikace, jde o věc nezávislou na technologii. Pokud nedokážeme provádět takovéto aktualizace, je narušena perzistentní identifikace i postup nepřímé identifikace. Jak již bylo řečeno, záruka aktualizace je do značné míry věcí strategie. Pracovní postupy pro správu informačních zdrojů, které vycházejí z této strategie, by měly zachovávat volnou vazbu mezi identifikátory lokace a perzistentními identifikátory: operace vedoucí ke změně identifikátoru lokace musí okamžitě vyvolat aktualizaci metadat perzistentních identifikátorů. Aktualizace by měla pokud možno probíhat zcela automaticky.

Pracovní postupy pro management identifikátorů musejí být načasovány tak, aby identifikátory vznikaly ve správný okamžik. Pokud má být objekt přidělen perzistentní identifikátor (například pokud má být do webové stránky vloženo její URI), pak musí identifikátor vzniknout dříve než objekt samotný. Ovšem identifikátor může vracet digitální objekt až tehdy, kdy označovaný objekt již skutečně existuje a je identifikátorem označen! Typickým řešením je vytvořit identifikátor předem s tím, že vrací prázdný objekt, nebo nevrací vůbec nic, je neaktivní. Jakmile je označovaný objekt aktualizovaný a dostupný on-line, aktualizuje také identifikátor. To vše se však musí stát předtím, než bude objekt nebo identifikátor zveřejněn. („Zveřejněním“ zde myslíme, že zdroj je k dispozici lidem mimo okruh správců obsahu, tj. mimo skupinu lidí, kteří pracují na vytváření a přípravě dokumentů.)<sup>6</sup>

Správa digitálních dokumentů by také měla být maximálně perzistentní a neměla by být vázána na konkrétní technologie nebo interní identifikátory. Správa digitálních zdrojů založená na perzistentních identifikátorech minimalizuje riziko, které mohou přinést změny v identifikátorech lokace. Také snižuje riziko, že identifikátor lokace se bude používat častěji než trvalý identifikátor. Takovéto události by mohly ohrozit důvěru v systém perzistentních identifikátorů.

Abychom uvedli příklad: pokud pro identifikaci různých verzí jednoho objektu používáme identifikátory lokace a nikoliv perzistentní identifikátory, může dojít k tomu, že identifikátory lokace budou používat i koncoví uživatelé a budou pomocí nich přistupovat k jednotlivým verzím digitálních dokumentů. Ovšem ve chvíli, kdy se identifikátor změní (například proto, že vznikla nová verze objektu s jiným identifikátorem), stává se odkaz, který má uživatel k dispozici, nefunkčním. V případě, že od samého začátku používáme interně pro správu verzí výlučně trvalý identifikátor (například s využitím parametru pro verze), je mnohem menší riziko, že budou zveřejněny dva protichůdné identifikátory.

Otázce „Jaké kroky je třeba podniknout k zajištění persistence identifikátorů?“ tedy odpovídají tři strategie, které vyžadují zapojení poskytovatelů zdrojů:

3. Je třeba zajistit, aby identifikátory přetrvaly změny, které vyžaduje správa digitálních dokumentů.
4. Musí existovat mechanismy, které zajistí, že jsou identifikátory aktuální.
5. Pro správu a přístup ke zdrojům je třeba používat identifikátory důsledně navzdory změnám, které management zdrojů vyžaduje.

Tyto strategie odpovídají třem bodům kontrolního seznamu:

**III.** Oddělit správu identifikátorů od správy digitálních zdrojů.

**IV.** Definovat pracovní postupy pro správu identifikátorů a integrovat je do správy digitálních zdrojů. (Sem patří automatické procesy pro aktualizaci identifikátorů při změně zdroje, stejně jako rozhodnutí o tom, ve které fázi životního cyklu dat identifikátory přidělovat.)

**V.** Navrhnout systém služeb pro správu digitálních dokumentů, který využívá identifikátory (jak pro vnitřní správu dokumentů, tak pro zpřístupňování navenek).

Volnější vazba mezi zdroji a identifikátory umožňuje flexibilitu nepřímé identifikace a zachovává perzistenci i přes změny v identifikátorech lokace. Dobře integrovaný systém pro správu identifikátorů poskytuje volnější vazbu mezi zdroji a identifikátory, takže dereference není narušena ani v případě, že se identifikátor lokace změní. Využívání trvalých identifikátorů při správě dokumentů pak minimalizuje závislost na identifikátorech, které nejsou perzistentní. Perzistentní identifikátory se stanou další informační vrstvou, jejímž prostřednictvím jsou zdroje zpřístupňovány a citovány spolehlivě a metodicky.

## **Jaké technologie bychom měli použít k zajištění trvalosti identifikátorů?**

Nelze očekávat, že persistenci identifikátorů zajistí nějaká konkrétní technologie: identifikátory by měly přetrvat změny technologií obsluhujících identifikátory, stejně jako všechny ostatní technologické změny. Jak bylo už mnohokrát řečeno, trvalost zajistí dobrá konceptualizace a administrativa, nikoliv konkrétní technologie. Jediný požadavek na technologii pro identifikátory je, aby napomáhala naplnění této koncepce a provádění administrativy.

Technologie musí být především kompatibilní s již používanými systémy v dané oblasti. Pokud nebude kompatibilní, budou perzistentní identifikátory oddělené od každodenního managementu zdrojů, čímž v ničem nepomohou při správě informačních zdrojů. Bez této integrace se vrátíme tam, kde jsme byli na začátku: identifikátory nebudou součástí správy informačních objektů a jejich trvalost nebude možné zajistit. Perzistentní identifikátory se musí stát přidanou informační vrstvou. Tato vrstva však bude k ničemu, pokud nebude účinně komunikovat s existujícími systémy a procesy správy informačních zdrojů.

Dostáváme se tak k nové strategii:

6. Perzistentní identifikátory musí být integrovány do existujících služeb a systémů.

Tato strategie je ovšem již součástí bodu IV. našeho kontrolního seznamu:

**IV. Definovat pracovní postupy pro správu identifikátorů a integrovat je do správy digitálních zdrojů.**

Pracovní postupy pro správu informací by neměly vést k tomu, že identifikátory budou integrovány někdy v budoucnosti. Nutné je, aby byly integrovány do stávajících procesů již nyní. To znamená, že musíme využít stávající služby pro zpřístupňování zdrojů, o kterých lze říci, že budou velmi pravděpodobně používány ještě po nějakou dobu.

To také znamená, že pokud mají perzistentní identifikátory tvořit samostatnou informační vrstvu, pak musí identifikační služby naplňovat všechny požadavky správy informačních zdrojů. Klíčové služby správy informačních zdrojů, které by měly používat perzistentní identifikátory, jsou:

- Načtení (*retrieval*): zpřístupnění zdroje, který je identifikátorem označen.
- Řešitelnost (*resolution*): zjištění, k čemu identifikátor odkazuje (což nemusí být nutně shodné s načtením).
- Vytvoření zdroje, aktualizace a smazání.

HTTP-REST je dobrým příkladem přístupu k poskytování těchto služeb založeného na URI: vytvoření zdroje, jeho aktualizace a smazání, stejně jako načtení/zpřístupnění zdrojů, jsou realizovány prostřednictvím operací HTTP na URI zdrojů.<sup>7</sup>

## Jak dlouho by měly identifikátory existovat?

Jakmile jsou perzistentní identifikátory zveřejněny pro koncové uživatele, je třeba zaručit, aby skutečně existovaly po dlouhou dobu. „Ale jak dlouho to je?“, to byla otázka, kterou jsme nejčastěji slyšeli od všech zainteresovaných stran. Zpočátku jsme odpovídali bez rozmyšlení („asi 25 let“), ovšem časem nám začalo být jasné, že každý projekt implementace perzistentních identifikátorů bude mít vlastní časový plán závisící na konkrétních podmínkách v dané oblasti a na řadě externích vlivů. Takže tvrdit, že existuje jeden všude platný časový odhad „trvalosti“, řekněme například 25 let, nemá smysl.

O trvalosti bychom měli uvažovat jako o době, po jakou má smysl systém identifikátorů v dané oblasti udržovat. Definujeme trvalost jako časové období, po které bude identifikace zajištěna a udržována, ať je to třeba 25 let nebo 25 dnů. Koncový uživatel by se měl dozvědět, jak dlouho lze očekávat, že bude systém identifikace fungovat, a jak dlouho se může na systém spolehnout. Doba, po kterou má být „trvalost“ zachována, nemusí mít archivní rozměr, ale nesmí být přerušena. Její skutečná délka zaleží na řadě okolností.

Protože zajištění trvalé dostupnosti a funkčnosti identifikátorů předpokládá závazek poskytovatele konečným uživatelům, měla by být doba, po kterou se poskytovatel zavazuje udržovat systém identifikace funkční, explicitně zveřejněna. Uživatelé perzistentních identifikátorů se musí připravit na dobu, kdy systém spravující identifikátory již nebude fungovat. Poskytovatelé identifikátorů mohou mít v úmyslu systém spravující identifikátory provozovat věčně, ale pokud na tomto systému mají záviset služby jiných systémů, je lepší zveřejnit přesné termíny. Z hlediska koncového uživatele znamená trvalost očekávání důvěryhodnosti. Poskytovatelé identifikátorů musí tuto důvěryhodnost ustanovit a definovat. Těmito problémy se zabývá se své práci John Kunze, zejména operátorem „??“ systému trvalých identifikátorů ARK.<sup>8</sup>

## Co je myšleno slovem „trvalý“?

Důvěřovat v persistenci identifikátoru znamená věřit, že něco nastane, nebo nenastane. Pouhé prohlášení, že identifikátor bude trvale existovat, nám neříká nic o tom, co se

bude ve skutečnosti dít. S identifikátorem se může v průběhu času přihodit několik věcí, které je třeba posuzovat odděleně. Pro lepší pochopení trvalosti navrhuje následující přístup:

Identifikátor je v nezákladnějším slova smyslu spojením nějakého znaku s nějakým objektem. Identifikátor nemusí být nutně on-line, nemusí to být hypertextový odkaz nebo URL.

Trvalost je vždy příznakem nějaké vlastnosti identifikátoru.

Vlastnosti identifikátoru, které mohou být perzistentní, zahrnují:

- **Znak**, tvořící identifikátor. Pokud se řetězec obsažený v URL nemění v čase, pak to znamená, že je trvalý, i když odkaz například nefunguje a objekt nevrací.
- **Vazbu** mezi identifikátorem a objektem, který identifikuje. Pokud existuje záznam o vazbě znaku tvořícího identifikátor k nějakému objektu, tato vazba přetrvává, a to i tehdy, kdy již nelze tento identifikátor použít k načtení objektu.

- Příklad 1: `.cs` je například doménové jméno pro Československo, i přestože je již vymazáno z registru IANA. To znamená, že vztah mezi `.cs` a doménovým jménem pro Československo přetrvává.

- Příklad 2: Esej Leo Felkina z roku 1999 dostupný na WWW: `<http://www.progress.org/leon01.htm/>` odkazuje na text, který „byl kdysi on-line“ na WWW: `<http://www.libertymatters.org/flyers.htm/>`. URL `<http://www.libertymatters.org/flyers.htm/>` stále označuje text *Liberty matters* z roku 1998, protože Felkinův esej tuto vazbu udržuje. Vazba je zaznamenána, a to i přesto, že samotný objekt označeným daným URL již není on-line.

- **Služba** spojená s identifikátorem. URI `<http://www.libertymatters.org/flyers.htm/>` již nefunguje, takže služba načtení objektu označeného tímto URI netrvá, ačkoli vazba mezi URI a objektem trvá.

- Fungování služby můžeme chápat jako řadu různých **případů**. Například služba načtení spojená s URI `<http://www.libertymatters.org/flyers.htm/>` (HTTP GET na vlastní URI) se nezdařila. Ale (na rozdíl od roku 1999) „wayback machine“ Internetového archivu nyní umožňuje objekt získat prostřednictvím vlastní služby fungující (nepřímo) na URI: `<http://web.archive.org/web/19981206012626/http://www.libertymatters.org/flyers.htm/>`. Identifikátor URL tedy lze stále načíst, ovšem nikoli pomocí výchozí instance služby, lokátoru přes HTTP.

- Fungování služby můžeme také rozdělit na řadu **typů**. Identifikátor může být řešitelný (*resolvable*), tj. umožňuje získat informace, které jednoznačně popisují, co je identifikátorem označováno, aniž by zároveň umožňoval přímo načtení zdroje (*retrieval*). Například článek identifikovaný pomocí „`http://arxiv.org/abs/gr-qc/0609101`“ byl stažen z `arXiv.org`, protože byl nepůvodní, ale název, autoři, abstrakt a informace o zveřejnění jsou stále k dispozici. To, co dostaneme, když klikneme na odkaz, již není načtení zdroje (*retrieval*), ale je řešení (*resolution*).

- **Odpovědnost** za identifikátor. Správci digitálního archivu často chtějí vědět, kdo je odpovědný za aktualizaci identifikátoru, a kdo ho vytvořil.

Trvalost je obvykle chápána jako zachování služby načtení zdroje (*retrieval*). To vyplývá z dlouholetých problémů s nefunkčními URI (které již neumožňují načtení zdroje) od okamžiku, kdy byl zdroj, na který URI odkazovalo, přesunut nebo smazán. Identifikátory ale mohou být udržovány i poté, co zdroje, které označovaly, již nejsou on-line. Pro

účely archivace je to nezbytné. V takovém případě bychom očekávali, že řešení (*resolution*) identifikátoru vrátí záznam metadat, a tak zachová vazbu identifikátoru se zdrojem (výše popsaný případ staženého článku na *arXiv* je příkladem této možnosti). Pokud jde o dlouhodobé uchování identifikátorů, je udržení řešitelnosti (*resolution*) důležitější než zachování možnosti načtení zdroje (*retrieval*). Údržba metadat směřuje k celkově spolehlivějšímu managementu elektronických zdrojů.

Záruky trvalosti musí být jasně vymezeny nejen s ohledem na to, jak dlouho má být trvalost garantována. Také musí být jasně řečeno, jaký aspekt identifikátoru je garantován. Z dlouhodobého hlediska je zejména třeba oddělit řešení (*resolution*) identifikátoru a načtení (*retrieval*) zdroje, má-li mít garance zachování identifikátoru nějaký smysl i po zániku identifikovaného objektu.

Otázce „**Jak dlouho by měly identifikátory existovat?**“ a s ní spojené otázky „**Co je myšleno slovem ,trvalý‘?**“ odpovídá následující strategie:

7. Stanovit, jak je chápána trvalost různých aspektů identifikátorů a přesně definovat záruky provozovatele systému koncovým uživatelům.

To odpovídá poslednímu bodu v našem kontrolním seznamu:

**VI.** Rozhodněte, na jak dlouho se zavazujete k identifikaci zdrojů a jak dlouho budete zajišťovat přístup k službám resolveru (co od trvalosti očekávají uživatelé a co může provozovatel systému spravujícího identifikátory poskytnout). Tato rozhodnutí je třeba jasně zveřejnit (tj. zveřejnit strategii trvalosti).

Nedávný rozvoj plánů pro správu dat, zejména v oblasti výzkumu, poskytuje formální rámec pro zapojení správců dat a uživatelů. Předpokládáme, že zajištění trvalosti identifikátorů bude součástí tohoto procesu.<sup>9</sup>

## Shrnutí a Kontrolní seznam

Vyvinuli jsme šestibodový kontrolní seznam pro integraci perzistentních identifikátorů do procesů správy informací. Shrňme ho tedy takto:

S poskytovateli zdrojů je třeba:

- I. Vytvořit informační model platný ve vaší oblasti.
- II. Začlenit perzistentní identifikátory do tohoto informačního modelu.

Se správci digitálních zdrojů je třeba se dohodnout na:

- III. Oddělení správy identifikátorů od správy digitálních zdrojů.
- IV. Definici pracovních postupů pro správy identifikátorů a jejich integraci do správy digitálních zdrojů. (Sem patří automatické procesy pro aktualizaci identifikátorů při změně zdroje i rozhodnutí o tom, ve které fázi životního cyklu dat identifikátory přidělovat.)
- V. Navržení systému služeb pro správu digitálních dokumentů, který využívá identifikátory (jak pro vnitřní správu dokumentů, tak pro zpřístupňování navenek).

Uživatelům zdrojů je třeba jasně sdělit:

**VI.** Jak dlouho se zavazujete k identifikaci zdrojů a jak dlouho budete zajišťovat přístup k službám resolveru (co od trvalosti očekávají uživatelé a co může provozovatel systému spravujícího identifikátory poskytnout). Tato rozhodnutí je třeba jasně zveřejnit (tj. zveřejnit strategii trvalosti).



Jednotlivé body jsou založeny na strategiích, které vycházejí z klíčových otázek týkajících se identifikátorů.

## Poděkování

Tento článek vychází z práce vykonané v rámci projektů PILIN a PILIN ANDS Transition Project. PILIN byl financován *Commonwealth Department of Education, Science and Training* (DEST) v rámci iniciativy systémové infrastruktury (SII) jako součást projektu Australské vlády, jejího Akčního plánu inovací pro budoucnost (BAA). PILIN ANDS Transition Project byl financován Australskou vládou jako součást Národní strategie infrastruktury výzkumu (NCRIS) v rámci přechodu na Australian National Data Service (ANDS).

Autoři děkují za podporu a zpětnou vazbu celému týmu projektu PILIN.

Copyright © 2009 University of Southern Queensland

### Poznámky a odkazy:

<sup>1</sup> *PILIN Project* [online]. Dostupný z WWW: <<http://www.linkaffiliates.net.au/pilin2/>>. Software dostupný z WWW: <[http://www.linkaffiliates.net.au/pilin2/outputs/outputs\\_software.html](http://www.linkaffiliates.net.au/pilin2/outputs/outputs_software.html)>.

<sup>2</sup> *The Handle System*® [online]. Dostupný z WWW: <<http://www.handle.net/>>. *Using URLs as Persistent Identifiers* [online]. Dostupný z WWW: <<http://resolver.net.au/hdl/102.100.272/DMGVQKN-QH>>. Všechny výstupy projektu PILIN jsou dostupné z WWW: <[http://www.linkaffiliates.net.au/pilin2/outputs/outputs\\_reports.html](http://www.linkaffiliates.net.au/pilin2/outputs/outputs_reports.html)>.

<sup>3</sup> *Persistence of Identifiers Guidelines* [online]. Dostupný z WWW: <<http://resolver.net.au/hdl/102.100.272/V89DC0DQH>>.

<sup>4</sup> IFLA Study Group. *Functional Requirements for Bibliographic Records, Final Report* [online]. 1998. Dostupný z WWW: <<http://www.ifla.org/VII/s13/frbr/frbr.pdf>>.

<sup>5</sup> *PURL (Persistent Uniform Resource Locator)* [online]. Dostupný z WWW: <<http://purl.org/>>. BERNERS-LEE, Tim. *Cool URIs don't change* [online]. 1998. Dostupný z WWW: <<http://www.w3.org/Provider/Style/URI>>.

<sup>6</sup> TRELOAR, Andrew; GROENEWEGEN, David; HARBOE-LEE, Cathrine. The Data Curation Continuum : Managing Data Objects in Institutional Repositories. *D-Lib Magazine*. 2007, vol. 13, no. 9/10. ISSN 1082-9873. DOI 10.1045/september2007-treloar.

<sup>7</sup> Původní definice REST viz kapitola 5. In FIELDING, Roy T. *Architectural Styles and the Design of Network-based Software Architecture*. 2000. Disertační práce. University of Kalifornie. Dostupný z WWW: <<http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>>.

<sup>8</sup> KUNZE, John A.; ROGERS, Richard. *The ARK Identifier Scheme*. 2008. Dostupný z WWW: <<http://tools.ietf.org/html/draft-kunze-ark-15>>. KUNZE, John A. *Towards Electronic Persistence Using ARK Identifiers*. 2003. Dostupný z WWW: <<http://www.cdlib.org/inside/diglib/ark/arkcdl.pdf>>.

<sup>9</sup> FITZGERALD, Anne M.; PAPPALARDO, Kylie M.; AUSTIN, Antony C. Legal Framework for e-Research Project & Open Access to Knowledge Law Project. In *Practical Data Management: A legal and policy guide*. 2008, Queensland University of Technology. Dostupný z WWW: <[http://eprints.qut.edu.au/14923/1/Microsoft\\_Word\\_-\\_Practical\\_Data\\_Management\\_-\\_A\\_Legal\\_and\\_Policy\\_Guide\\_doc.pdf](http://eprints.qut.edu.au/14923/1/Microsoft_Word_-_Practical_Data_Management_-_A_Legal_and_Policy_Guide_doc.pdf)>.