

## Trvalé identifikátory pro kulturní dědictví

Internetové zdroje mají omezenou životnost. Jejich identifikace a trvalé umístění je komplexním problémem, který může ovlivnit řadu technologických a organizačních problémů jako jsou citace, vyhledávání a ochrana kulturních/vědeckých zdrojů. Nejde zdaleka jen o technický problém: trvalá identifikace digitálních objektů, včetně textů, hudby, videa, obrazových a vědeckých dokumentů atp. je stále velkým problémem a ztěžuje používání internetu jako důvěryhodné platformy pro výzkum a šíření vědeckého a kulturního obsahu.

### Proč potřebujeme « trvalé identifikátory »?

Jednou z klíčových činností kulturních institucí, jako jsou univerzity, archivy, muzea a knihovny, je dnes dlouhodobá ochrana, zpřístupňování a šíření kulturních digitálních objektů. Používání URL nelze považovat za přístup pro spolehlivé řešení tohoto problému z důvodů strukturní nestability odkazů (např. již neexistujících domén) a s nimi souvisejících zdrojů (přemístění nebo aktualizace). V současnosti převažující používání URL zvyšuje riziko ztráty kulturních dokumentů a nedostatečného využívání existujících sbírek kulturních digitálních objektů. V oblasti kulturního dědictví nejde jen o identifikaci nějakého objektu, ale také o garanci jeho trvalé dostupnosti.

Důvěryhodným řešením je spojit digitální objekt s trvalým identifikátorem (TI), který zůstane stejný bez ohledu na to, kde se daný digitální objekt právě nachází.

K zavedení systému trvalé identifikace jsou nezbytné následující kroky:

- 1) výběr objektů, které potřebují TI
- 2) přidělování TI a vytvoření jejich registru
- 3) zprostředkování vyhledávání URL na základě zadání TI
- 4) udržování registru, který spojuje TI <-> URL, a garance trvalé dostupnosti k digitálním zdrojům.

První krok je výhradně v pravomoci kulturní instituce, následující kroky mohou být delegovány jiným institucím takovým způsobem, aby bylo dosaženo co nejlepší ekonomické a funkční stability této služby.

### Požadavky na trvalé identifikátory

Instituce kulturního dědictví by měly vybírat systém trvalé identifikace podle následujících kritérií:

- globální jednoznačnost
- trvalost
- zpřístupnitelnost
- spolehlivost
- spravující autorita
- pružnost
- interoperabilita
- nákladnost

### Globální jednoznačnost

Identifikátorem rozumíme označení, které je v určitém kontextu spojeno s nějakým objektem. Kontextem se zde rozumí jak standardy syntaxe názvu identifikátoru (např. URN:NBN:IT:xxx-xxxx), tak identifikace autority (podřízený jmenný prostor), která tento název přiděluje.

## Glosář:

**Objekt:** jakákoli entita, která je předmětem intelektuální transakce. Je definována metadaty a terminologií z datového slovníku (v rámci rejstříků atp.) k zajištění vzájemného porozumění (interoperabilita). Objekty mohou být fyzické, digitální, nebo abstraktní, např. lidé, organizace, smlouvy atd.

**Zpřístupňovací služba (dereference):** proces, při němž je identifikátor vstupem (ve formě požadavku) do síťových služeb a na jehož základě se vrací specifický výstup (zdroj, digitální objekt, metadata atd.).

**Pojmenovávací autorita:** Nezávislá autorita, která přiděluje názvy identifikátorů a zajišťuje jejich jedinečnost a trvalost. Pojmenovávací službu by měla provozovat každá autorita. Měla by zajišťovat zpřístupňování objektů na základě zadání jejich TI. Distribuovaný systém trvalé identifikace předpokládá, že odpovědnost za generování a zpřístupňování může být delegována jiným institucím, tzv. periferním autoritám, které mají na starost určitou část jmenného prostoru.

**Jmenný prostor:** abstraktní kontejner, který uvádí do kontextu jednotky, které obsahuje, a umožňuje jednoznačnou identifikaci jednotek se stejným názvem (vyskytujících se v jiných jmenných prostorech).

**Registr:** Tabulka spojující URN (nebo jiný jiný TI) a jedno nebo více URL.

**Repozitář:** Místo, kde jsou uloženy digitální objekty (DSpace, Fedora, Codex atd.), a kde je nebo není zaveden nějaký systém organizace informací (systém souborů).

**URI (jednotný identifikátor zdrojů)** - je řetězec znaků s definovanou strukturou, který slouží k přesné identifikaci objektů.

**URL (jednotný lokátor místa)** - je URI, který spolu s identifikací zdroje, který slouží k přesné specifikaci umístění zdrojů informací. Umožňuje zpracovávat nebo získávat kopie zdrojového objektu popisem jeho primárního přístupového mechanismu nebo síťové lokace.

## Trvalost

Trvalost se vztahuje na dobu existence identifikátoru. Trvalý identifikátor nelze zrušit nebo přidělit jinému digitálnímu objektu. To znamená, že TI je globálně jedinečný jednou provždy a funguje jako odkaz na daný objekt i po té, co již neexistuje autorita, která ho přidělila. Trvalost je ve službách a strategiích institucí spravujících kulturní dědictví zjevně velmi specifická. Jedinou zárukou použitelnosti a trvalosti systému identifikace je závazek organizace, která přiděluje a spravuje trvalé identifikátory a poskytuje k nim vyhledávací služby.

## Zpřístupnitelnost

Zpřístupnitelnost znamená možnost získat digitální objekt v případě, že je publikovaný. Je třeba rozlišovat mezi zpřístupněním objektu a jeho pouhou identifikací. Výběr identifikačního názvu nutně neurčuje podobu odpovídající vyhledávací architektury.

## Spolehlivost

K zajištění spolehlivosti systému trvalé identifikace je nutné, aby infrastruktura, na které systém trvalé identifikace funguje, byla stále v provozu (zálohování služeb, zálohování úložiště atd.) a aby byl registr TI <-> URL nestále aktuální (automatická aktualizace).

## Spravující autorita

Jedinou zárukou použitelnosti a trvalosti systému identifikace je závazek organizace, která přiděluje identifikátory, provozuje jejich registr a zprostředkovává vyhledávací služby. V oblasti institucí spravujících kulturní dědictví je obvyklé spoléhat se na služby veřejných institucí, jako jsou národní knihovny, státní archivy apod. Požadavky na pověření a důvěryhodnost systému trvalé identifikace bychom měli velmi pečlivě zvážit, dříve než se pro nějaký rozhodneme.

## Pružnost

Efektivní systém trvalé identifikace musí být schopen dostát nárokům pro různé typy dokumentů a sbírek. Systém trvalé identifikace by měl být například schopen zvládat různé úrovně granularit digitálních objektů, protože identifikátor musí v různých uživatelských oblastech umět odkazovat k různým, velmi odlišným typům zdrojů.

## Interoperabilita

Tato vlastnost je zásadní pro zajištění možnosti šíření a zpřístupňování kulturních digitálních objektů.

K dispozici je celá řada technologických řešení a mnohá z nich jsou upravena pro určitou konkrétní oblast. Interoperabilita mezi jednotlivými systémy musí být zajištěna přinejmenším na úrovni služeb pro koncové uživatele. Systémy musí umožňovat použití ve společném a jednoduchém uživatelském rozhraní. Interoperabilitu na systémové úrovni může zajišťovat přijetí otevřených standardů.

## Náklady

Systémy trvalé identifikace používané v institucích spravujících kulturní dědictví by měly být dostupné zdarma nebo za rozumný poplatek, protože tyto instituce mají za úkol zajišťovat volný přístup k datům a měly by předcházet vytváření digitální propasti.

**PURL** (persistent URL/Trvalé URL): Z hlediska funkce je PURL v podstatě URL. Nesměřuje ovšem přímo na internetové umístění objektu, ale na zprostředkovatelskou zpřístupňovací službu, a s využitím standardních serverových přesměrování odkáže přeměřovat dotaz trvalého identifikátoru do aktuální lokace zdroje nebo dokumentu.

[www.purl.org](http://www.purl.org)

**URN** (Uniform Resource Name/Jednotný název zdroje): URN je URI využívající schéma URN a nutně nepředpokládá dostupnost identifikovaného zdroje. Cílem URN je sloužit jako trvalý identifikátor, který je nezávislý na umístění objektu. URN je navrženo tak, že může včleňovat jiné jmenné prostory (které mají stejné vlastnosti jako URN) do svého schématu (např. URN-NBN atp.). Syntaxe URN poskytuje nástroje ke kódování znaků v takové formě, která může být odeslána v existujících protokolech, přepsána na většině klávesnic atd.

[www.ietf.org/rfc/rfc1737.txt](http://www.ietf.org/rfc/rfc1737.txt)

**Systém Handle:** Handle je globální systém identifikace, který umožňuje zabezpečené zpřístupňování a administraci na sítích jako je internet. Systém Handle spravuje „handly“, což jsou jedinečná označení digitálních objektů a jiných internetových zdrojů. Pojmenovávající autorita přiděluje a spravuje jména „handlů“, identifikátor musí být v rámci autority jedinečný, ale jeho syntaxe jinak není ničím omezena.

[www.handle.net](http://www.handle.net)

**XRI** (OASIS Extensible Resource Identifier/Rozšířený identifikátor zdrojů): XRI definuje schéma URI a odpovídající jmenný prostor URN pro distribuované služby, které mají zajistit identifikaci zdrojů (včetně lidí a organizací) a sdílí data mezi různými doménami, aplikacemi a oblastmi podnikání.

[www.oasis-open.org](http://www.oasis-open.org)

**ARK** (Archival Resource Key /Klíč pro archivní zdroj) (IETF Internet draft): Schéma vyvinuté pro účely usnadnění trvalého pojmenování a identifikace informačních objektů. Základním předpokladem ARKu je, že trvalost je pouze vlastností služeb/organizace a není vlastností objektu ani jakéhokoli systému identifikace. Identifikátor ARK poskytuje tři výsledky zároveň: digitální zdroj, metadata, vyjádření závazku trvale objekt uchovávat.

[www.cdlib.org/inside/diglib/ark](http://www.cdlib.org/inside/diglib/ark)

**N2T** (Name to thing/Pojmenování věcí): N2T je konsorcium paměťových institucí a malého, obvyklého webového serveru, který je kvůli spolehlivosti zrcadlen na několika místech po světě. Cílem tohoto projektu je chránit URL 200 organizací před nestabilitou, pomocí 200 pravidel přepisu jednoduchým pravidlem pro HTTP přesměrování pro každou organizaci.

[www.n2t.info](http://www.n2t.info)

## Další důležité body

### Granularita

Granularita vyjadřuje míru podrobnosti, s jakou je digitální objekt identifikátorem označován. Požadovaná granularita bude mít zásadní dopad na to, jaký systém trvalé identifikace si instituce zvolí.

Někdy bude potřeba citovat webovou stránku, která slouží jako vstupní stránka pro celou sbírku webových souborů, jindy odborný časopis, knihovni jednotku nebo kapitolu monografie. Z právního hlediska je někdy nutné být v identifikaci ještě přesnější. Každá instituce musí zvážit, zda ten který systém trvalé identifikace odpovídá požadavkům na granularitu typu jejich sbírek.

### Nettransparentní nebo transparentní trvalé identifikátory

Trvalý identifikátor nemusí sám o sobě obsahovat žádnou informaci o objektu, který identifikuje (nettransparentní TI). Může být složen z náhodných znaků, které dohromady netvoří žádný význam. Nettransparentní systém trvalé identifikace funguje pouze s fungující vyhledávací službou. Je však možné, aby v sobě měl identifikátor zabudován nějaký význam nesoucí prvek (transparentní TI).

Obecně je samozřejmě snazší si zapamatovat mnemotechnicky vytvořené identifikátory než identifikátory složené z náhodných řad znaků; na strojové zpracování ovšem nemá transparentnost identifikátoru vliv.

### Nové verze objektů

Každá nová verze objektu musí mít vlastní trvalý identifikátor. Novou verzi lze považovat za jiný digitální objekt, protože jeho obsah nebo formát může být jiný. Správa různých verzí jednoho objektu může využívat nějakých pravidel pro přidělování názvů identifikátorů nebo metadat.

## Jak nám mohou pomoci technologie?

Trvalá identifikace si vyžaduje udržovat databázi, která obsahuje informace o aktuálním umístění digitálních objektů. Nazývá se „databáze resolveru.“ Databáze resolveru udržuje informace o aktuálním umístění objektů a přesměrovává na něj uživatele. Databáze resolveru a zpřístupňovací služba může fungovat několika rozdílnými způsoby: centralizovaně, nebo distribuovaně, může využívat DNS, nebo nemusí.

Centralizovaná architektura je založena na centrálním uzlu, který generuje názvy objektů a trvale zajišťuje jejich vyhledávatelnost a spolehlivost. Toto řešení předpokládá centralizaci odpovědnosti a řízení nákladů. Centralizovaný systém má jeden centrální bod, který může selhat.

Distribuovaná architektura vyžaduje, aby každá periferní část systému, měla distribuované registry a vlastní zpřístupňovací službu. Každý periferní uzel si přiděluje poslední část názvu TI a má odpovědnost za jejich správu. Centrální registrační autorita zařizuje přesměrovávání na zpřístupňovací služby příslušných periferních uzlů.

## Zdroje

ERPANET workshop Persistent Identifiers  
[www.erpanet.org/events/2004/cork/index.php](http://www.erpanet.org/events/2004/cork/index.php)

DCC Workshop on Persistent Identifiers  
<http://www.dcc.ac.uk/events/pi-2005/>

URN:NBN  
<http://www.ietf.org/rfc/rfc3188.txt>

URN:NBN:DE  
<http://www.persistent-identifier.de>

URN:NBN:IT  
<http://www.rinascimento-digitale.it>

DOI  
<http://www.doi.org>

ARK  
<http://www.cdlib.org/inside/diglib/ark/>

PADI  
<http://www.nla.gov.au/padi/topics/36.html>

PILIN  
<https://www.pilin.net.au/>

OpenURL  
[http://www.niso.org/committees/committee\\_ax.html](http://www.niso.org/committees/committee_ax.html)

Architektura založená na DNS využívá http protokol k aktivaci citovaného odkazu na internetu. Systém založený na DNS nemá žádné zvláštní nároky na uživatele, nejsou třeba žádné doplňky do obvyklého internetového prohlížeče.

Další implementace pomohly vytvořit speciální protokol spravující pojmenovávání a zprostředkování vyhledávání TI (např. DOI). Architektura nezaložená na DNS vyžaduje speciální software (nebo doplněk do prohlížeče) k vyhledávání lokalizaci speciálního identifikátoru.

## Možnosti dalšího výzkumu

S růstem významu informačních technologií je stabilitě URL a dostupnosti digitálních zdrojů na internetu věnována stále větší pozornost. Trvalé identifikátory jsou relativně novou reakcí na tento problém. Velmi dynamický kontext, ve kterém tyto systémy operují, odhaluje nové oblasti pro další výzkum. Zde jsou některé zajímavé a dosud nevyřešené aspekty trvalé identifikace, kterým by stálo za to se dále věnovat:

Převládá tendence využívat systémy, které se vztahují k určité oblasti (např. NBN v knihovnách). Digitální objekty však mohou být součástí více oblastí a mohou být identifikovány pomocí různých systémů. Je proto nutné zajistit interoperabilitu mezi různými identifikačními systémy a jejich implementacemi na základě stejného jmenného prostoru.

Trvalé identifikátory zprostředkují přístup nejen k digitálním objektům, ale i k jejich metadatům, která jsou zásadní proto, aby uživatel mohl posoudit obsah dokumentů. Je proto velmi důležité pracovat na dalším vývoji komplexní správy metadat a uživatelských služeb, například takových, které zprostředkují obsah několika digitálních repozitářů najednou.

Sémantické vztahy mezi multimediálními objekty je možné vzít v úvahu při definici ontologií a snaže lépe porozumět internetovým zdrojům.