

Stav implementace perzistentních identifikátorů v NK ČR a výhled do budoucna



Jan Hutař
Marek Melichar
Ladislav Cubr

Osnova



1. Perzistentní identifikátory (PID) obecně
2. PID v digitálním světě
3. Současná situace v ČR
 - *Handle (Kramerijs)*
 - *URN:NBN (obecně, zahraničí, role NK)*
4. Resolver NBN v NK
5. Co nás ještě čeká, možnost spolupráce

Co je PID?



- řetězec znaků, který reprezentuje určitý objekt
- význam má pouze v určitém kontextu (systému)
- tento systém je udržován zodpovědnou institucí (např. knihovnou)
- nutno znát nejen PID, ale i správce (instituci)
- tradice PID v knihovnách
 - 1898 – Kongresová knihovna – systém LCCN pro katalogizační lístky
 - 1970 – systém ISBN, další systémy: ISSN, ISMN, ISAN aj.
- tradiční PID fungují dobře, v digitálním prostředí je věc komplikovanější
 - větší okruh problémů: digitální objekty lze na základě PID nejen identifikovat (metadata), ale také okamžitě získat (data) ap.

PID v digitálním prostředí



Postup zavádění PID pro digitální objekty:

- **Výběr objektů pro identifikaci**
- **Rozhodnutí o granularitě** (označovat jen celé knihy / kapitoly / stránky?)
- **Pravidla pro pojmenování** (volba znaků a struktury PID, možnost hierarchizace)
- **Registr PID a informací o umístění objektů** (vazby PID+URL)
- **Údržba** registru (aktuálnost a bezchybnost PID)
- **Služba** pro vyhledávání a získávání objektů na základě PID (resolver)
- **Dlouhodobá ochrana – perzistence** (trvalost) PID „na věky“



Hlavní funkce PID

- **jednoznačnost PID**
 - 1 PID odkazuje pouze na jeden objekt, a NE jiný
 - mělo by platit i naopak - 1 objekt má jen jeden PID
 - to vše však platí pouze v daném kontextu (systému)
- **perzistence PID**
 - trvalý a neměnný vztah mezi řetězcem (PID) a digitálním objektem, který označuje - PID bude mít význam pouze tak dlouho, jak dlouho vydrží tento vztah
 - není vlastností technologie, ale spravující instituce
- **perzistence objektu**
 - problematika perzistence objektu samotného – dlouhodobá ochrana digitálního dokumentu

Systemy PID v digitálním světě



Různé architektury systémů, nejdůležitější je však efektivní fungování správce systému (instituce)

- DOI (správcem je DOI Foundation)
- ARK (California Digital Library)
- PURL (OCLC)
- Handle (CNRI)
- URN:NBN (IANA + jednotlivé národní knihovny)

PID v ČR 2007



- Dosud neimplementován plnohodnotný systém PID pro digitální objekty
- Vytvořena pracovní skupina pro PID
- Úkol: zjistit požadavky zainteresovaných institucí v této oblasti
- Vytvořen kooperativní webový prostor (pid.ndk.cz)
- Výstup: je nereálné vybrat jediný systém pro všechny - každá instituce má specifické požadavky, které nesplňuje jediný systém
- V zahraničí jsou běžné případy, kdy jedna instituce využívá více systémů PID

PID v ČR 2008



- Pokrok oproti „iniciačnímu“ roku 2007
- Několik institucí začíná plnohodnotně využívat systémy PID
 - **KNAV** – *DOI*
 - **Kramerius v NK ČR** – *Handle* od března 2008
 - **ÚVT UK** – aktivní nasazení *Handlu* v DigiToolu do konce r. 2008
- Koncepce NDK > 14.5.2008 – vláda schvaluje záměr NDK v rámci Strukturálních fondů
- NK ČR pracuje na zavedení systému URN:NBN a národního resolveru

Handle a Kramerius



- Kramerius je napojen na server systému handle.net (verze 3.1.0)
- pro každou instanci Krameria nutnost registrace u CNRI
- získání unikátního prefixu – zpoplatněno
- identifikátory označují monografie i periodika
- identifikace jde až na nejnižší úroveň popisu (titul> ročník> výtisk> strana> obrázek)
- <http://kramerius.nkp.cz/kramerius/handle/ABA001/11008209>

KRAMERIUS - PERIODIKA (498/3608815) - F (33/127491) - **Filmová práce** (1/160)

Typy dokumentů

Abeceda

Výběr titulu

Titul

Ročník

Výtisk

Stránka

HLEDAT

[Pokročilé vyhledávání](#)

Filmová práce

* **ISSN:**

1801-948X

* **Hlavní název:**

Filmová práce

* **Místo uložení:**

Národní filmový archiv - knihovna

* **Signatura:**

III-31

* **Zobrazit titul v Souborném katalogu ČR**

* **URI:**

<http://kramerius.nkp.cz/kramerius/handle/ABA001/11008209>

METS dokument

Datum vydání ročníku: 1945 Číslo ročníku: 1 (1/160)

Proč URN:NBN?



- URN je identifikační systém s celosvětovou působností (IANA / Internet Society)
- NBN je jeho podsystém pro národní knihovny
- snaha o národní resolver (vč. dalších institucí)
- mezinárodní rozšíření NBN
- NBN je velmi živé v okolních zemích
- open source SW k dispozici (Itálie)



Cíl:

Zajistit „trvalou dostupnost“ digitálního objektu

Předpoklad:

Tento digitální objekt má „trvalou hodnotu“

Řešení:

Použití nějakého systému „trvalé identifikace“
(PID)

Co k tomu potřebujeme ?



1. Resolver (nástroj, který to zajistí)
2. Důvěryhodná spravující instituce (registrační autorita)
 - udržuje vztahy PID-URL
 - garantuje autenticitu, úplnost a dostupnost objektu

URN:NBN



- zaregistrován u URN Finskou národní knihovnou
- RFC 3188
- PID pro identifikaci publikací, které nemají jiný identifikátor (např. ISBN)
- smějí jej implementovat pouze národní knihovny

Resolver



- obtížně přeložitelný termín („lokalizátor“, „vyhledávač“ ap. váže různé konotace) – nepřekládáme
- přiděluje nové jedinečné PID
- udržuje registr vztahů PID-URL-digitální objekt
- na základě zadání PID pomáhá vyhledat digitální objekt
- spravuje systém PID – sběr a kontrola PID
- doplňkově funguje i jako záložní archiv objektů samotných

Existující implementace



- Upsala University > portál Diva (Skandinávie)
 - zpřístupňování VŠ prací
- Projekt Epicur (Německo)
- SURF Foundation v Nizozemí
- Severské země (snahy o spolupráci)
- Itálie, Maďarsko

Persistent Identifier

...eindeutige Bezeichner für digitale Inhalte...

[Home](#) [Assignment](#) [Resolving](#) [Implementation](#) [FAQ](#) [Service](#) [Press](#) [References](#) [Imprint](#)

[Home](#)

[+ Assignment \(URN\)](#)

[+ Resolving \(URN\)](#)

[+ Implementation \(URN\)](#)

[FAQ](#)

[+ Service](#)

[+ Press](#)

[+ References](#)

[Imprint](#)

Worth knowing

- What are Persistent Identifiers (PI)?
- Demonstration of the URN resolver of the Deutsche Nationalbibliothek
- URN (NBN) Resolving Plugin for Internet Explorer, Mozilla, Netscape, Firefox
- Examples

Current

- URN Generator (Swiss National Library, 21.04.2005)
- EPICUR-Resolver (21.02.2005)
- URN (NBN) - Generator (18.02.2005)
- URN (NBN) Resolving Plugin for Internet Explorer (25.01.2005)



02.01.2008

GEFÖRDERT VOM





URN:NBN service in the NSZL

Identification of electronic documents

Home

Register

Modify

Extend

Delete

Resolve

Reverse Resolve

URL-List

Descriptions

Hungarian

Welcome!

On these pages you can register, manipulate and resolve URN:NBN identifiers (URN -- Universal Resource Number, NBN -- National Bibliography Number) which supports the long-term identification of your documents on the web.

For safety reasons each operation, except resolving, consists of two steps. You can find detailed description on the [Descriptions](#) page, but the most important information for a given operation is always indicated on the appropriate page as well.

You can send your questions, remarks and suggestions to the [owner](#) of the page.



nuove
tecnologie
per i beni
culturali

Fondazione
RINASCIMENTO
digitale

National Bibliography Number



Versione Italiana

► Menu principale

Home

Project

Resolver

Download

Partner

Resources

Contacts

► Login Form

Username

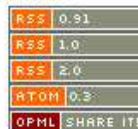
Password

Ricordami

Entra

[Password dimenticata?](#)

► Syndication



cerca nel sito...

Home

National Bibliography Number project

giovedì 29 maggio 2008

One of the most important problems of cultural heritage is the constant identification of internet resources, particularly in the framework of Digital Library and Digital Preservation, that aim at using internet as platform for contents research and dissemination. National Libraries, having mandate of storage and preservation of the national cultural heritage and having a leading role also about internet resources, can fulfil these requirements. The principal project' s objectives are:

- Test a new technology, based on the Uniform Resource Name (URN), using National Bibliographic Number (NBN) as namespace of reference,
- Develop a prototype for the implementation of a identification system of the digital cultural objects on internet,
- Make easier resources access from citizens,
- Awaken/Bind cultural institutions to apply/implement digital preservation programs for their own digital resources.

Ultimo aggiornamento (giovedì 29 maggio 2008)

Italský projekt URN:NBN

spolupracující instituce:

- Fondazione Rinascimento Digitale (FRD)
- National Library in Florence (BNCF)
- University consortium (CILEA)

Výsledek italského projektu:

➤ **Prototyp aplikace pro správu NBN**

Software



-open source (základem je DSpace)

využívá:

Java - jre1.5.0

Databáze – postgresql

Server - Tomcat

První fáze



- vývoj resolveru
- otestování nové technologie přístupu k URN, založení na distribuované architektuře
 - vzájemně sdílené (peer-to-peer) sítě
 - decentralizovaný systém

Druhá fáze



- Rozšíření užívání po Itálii
- Podpořit fungování resolveru dle PID ve vzájemné síti
- Vyřešit propojení URN:NBN s DOI, ARK

Decentralizovaný přístup



= využití výhod vzájemného sdílení
(peer-to-peer)

- zvýšení spolehlivosti (funkčnost zachována při výpadku lokálních uzlů i uzlu centrálního!)
- přenesení části odpovědnosti na registrátory druhé úrovně – snížení nákladů
- přirozená otevřenost k jiným systémům

Složky systému



- Centrální uzel – NK ČR
 - registrátor první úrovně - rozhoduje o všech jmenných prostorech za URN:NBN:CZ
 - Např. MZK → URN:NBN:CZ:MZK
- Lokální uzly – registrátoři druhé úrovně
 - rozhodují o poslední části řetězce
 - např. ABC123
 - výsledný PID: URN:NBN:CZ:MZK-ABC123

Registrátor 1.úrovně



Centrální uzel

- 1. automatický sběr nových / aktualizovaných PID ze sítě
- 2. kontrola → řešení duplicit (stejný PID / různé digitální objekty, stejné MD5 / různé PID)
- 3. zařazení bezchybných dat do centrální databáze
- 4. průběžná distribuce všem členům sítě

Registrátoři 2.úrovně



Lokální (podřízené) uzly:

- 1. provoz SW na připojení do sítě
- 2. vytváření nových PID v rámci dané kompetence
- 3. správa lokální databáze
- 4. podíl na lokalizaci PID v síti
- 5. provoz resolveru (volitelně)

NBN creator

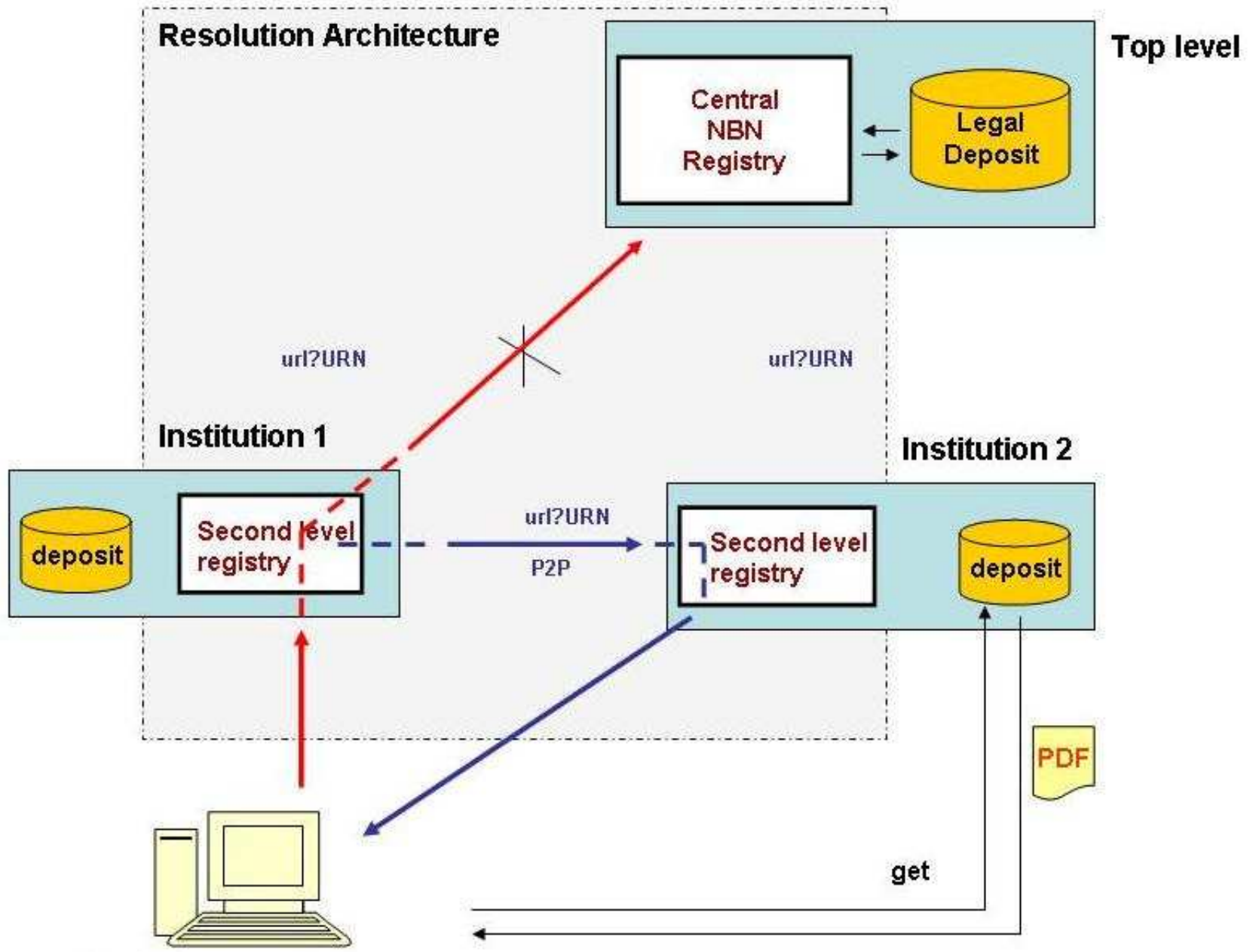
<p>Home</p> <p>Browse</p> <p>Communities & Collections</p> <p>NBN</p> <p>Sign on to:</p> <p>Receive email updates</p> <p>My NBN-creator authorized users</p> <p>Edit Profile</p>	<p>NBN-creator ></p> <p>NBN-creator is Live</p> <p>Communities in NBN-creator</p> <p>Choose a community to browse its collections.</p>	<p>This is a default installation of NBN-creator</p>
--	---	---



Lokalizace PID

Lokalizace po zadání URN:NBN:

- Přímo z lokální databáze
- Lokální resolver nezná, odkáže na centrální → ten zprostředkuje dotaz příslušnému lokálnímu uzlu
- Lokální resolver nezná, odkáže na centrální → ten zprostředkuje dotaz příslušnému lokálnímu uzlu → ten nefunguje → centrální uzel použije svoji databázi
- Lokální resolver nezná → centrální nefunguje → nefunguje ani příslušný lokální resolver → lokalizace z ostatních lokálních uzlů



Resolution request for
URN:NBN:xxx-yyyy
e risposta

Shrnutí - co máme?



- syrový testovací provoz na virtuálním stroji
- doménu resolver.nkp.cz
- snahu ;-)

Problémy, aneb co dále



- prototyp – ještě jedna vývojová fáze bude následovat
- open source – potřeba lokalizace, otázka podpory ?
- vůle partnerů v ČR (podpora pracovní skupiny?)
- pomoc při testování vzájemně sdílené sítě
- nasazení na skutečný finální server
- doladit syntax PID
- **dořešit financování celého systému**



Děkujeme za pozornost

jan.hutar@nkp.cz

marek.melichar@nkp.cz

ladislav.cubr@nkp.cz