



Národní knihovna České republiky

Odbor digitální ochrany

**Národní knihovna a předpoklady
pro českou infrastrukturu trvalé
identifikace digitálních
dokumentů**

Ladislav Cubr
Jan Hutař
Marek Melichar



Cíle/obsah prezentace

- Úvaha o problematice identifikace v digitálním světě
- Souhrn podmínek a předpokladů funkčnosti identifikátoru URN:NBN
- Vyvolat institucionální debatu o možném využití identifikátoru (kdo? proč? nač?)
- Opakující se nejasnosti
- Vyvolat zájem o zapojení dalších institucí do projektu pro vytvoření společné infrastruktury pro identifikaci založenou na systém URN:NBN



Identifikace v digitálním světě

- identifikací (vygenerováním čísla/ID) věc nekončí, ale začíná
- je potřeba stále udržovat identifikátory i data jimi označovaná a vazby mezi obojím

identifikace =

- přidělení identifikátoru
- správa identifikovaných objektů a identifikátorů
- administrace infrastruktury

Tedy - v digitálním světě NELZE oddělovat otázku přidělení a identifikace dat od trvalé správy dat.



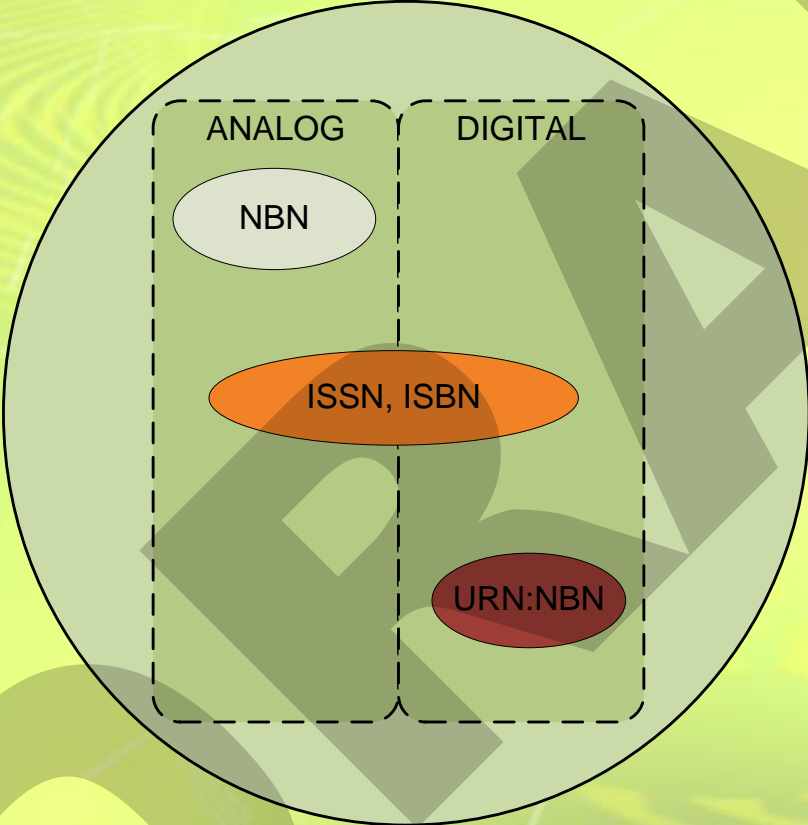
- odpovědnost (nejen) národních knihoven
- pro to, co bude stanoveno jako součást českého národního dědictví
- možno i pro další (vydavatelé aj.) – dle dohody
- možnost vazby s povinným elektronickým výtiskem v budoucnu
- není jediným řešením (více ID systémů v 1 knihovně – dle projektů např., dle granularity aj.)
 - 1 IE více ID > není třeba čekat na NK (vlastní analýza, vlastní systém)
- pro řadu jiných věcí se nehodí
 - např. identifikace stránek



Primární cíle URN:NBN

- **Identifikace obsahu digitálních dokumentů a jejich aktuálních lokací**
- **Možnost získat přímo dokument – proto dokumenty musí být přístupné online**
- **Dokumenty mají vztah k národní bibliografii a národnímu kulturnímu dědictví**

Identifikátory a typy zdrojů, které mohou označovat - standard



National Bibliography Number (NBN) is a generic name referring to a group of identifier systems utilised by the national libraries and only by them for **identification of deposited publications which lack an identifier, or to descriptive metadata (cataloging) that describes the resources.**

The materials identified by a national bibliography number may exist only in printed or other physical form, not electronically.

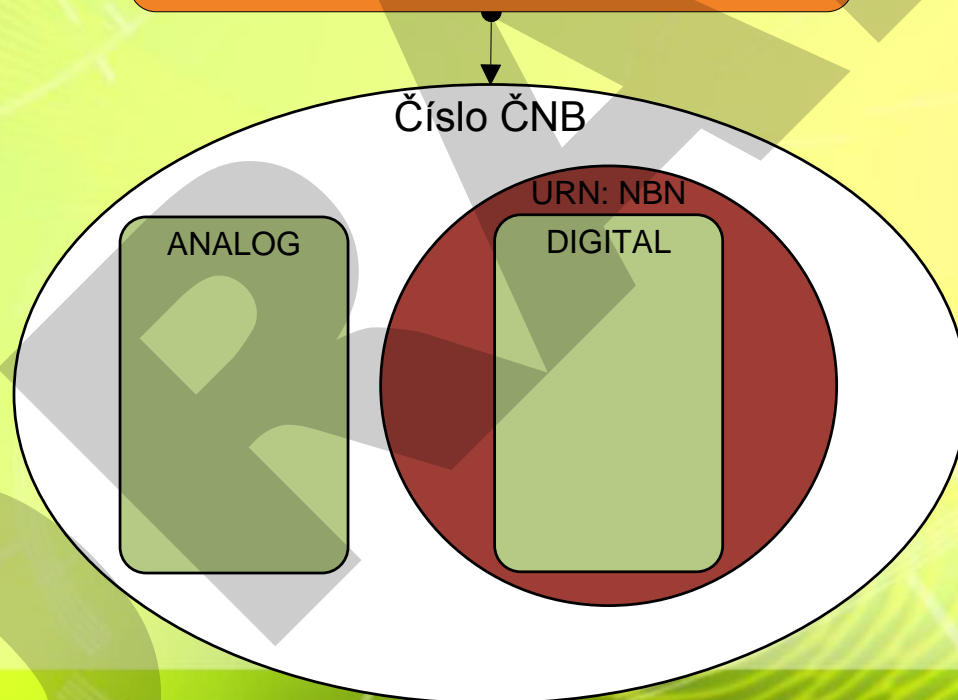
RFC3188



Pro náš kontext:

Identifikátory a typy zdrojů, které mohou označovat – aplikace pro CZ

Základní intelektuální entita (PREMIS) odpovídá 1 záznamu v katalogu ČNB





- Konflikty při realizaci systému URN:NBN:

Uživatelské hledisko x hledisko správy dokumentů

- **Různé potřeby, jiný důraz na granularitu, problém verzí a formátů má v těchto dvou kontextech zcela jiný význam**

Jedinečnost a trvalost identifikace DD x přímé zpřístupnění objektů (lokace)



Pro funkční systém URN:NBN zdaleka nestačí

- *generátor* jedinečných znakových řetězců
- *směrovací služba* (na aktuální URL);
resolution vs. retrieval
- *softwarová řešení* (resolver apod.) – nejsou
bezúdržbová, potřeba lidského vkladu
(koncepce, dohled)
- *stanovení syntaxe* identifikátoru
- *technická infrastruktura* – servery, databáze

**TO VŠE: jsou nutné podpůrné
součásti/pomůcky, nikoliv ŘEŠENÍ**



Předpoklady pro to, aby to celé mohlo fungovat ...



Předpoklad 1

ANALÝZA STAVU

- zmapování informačních toků, při nichž se užívají stávající identifikátory v systémech instituce
- stanovení, pro co je identifikátor URN:NBN vhodný a pro co není (v každé instituci a pro procesy mezi nimi)
- datové a procesní modely zainteresovaných institucí
- konsensus na tom, co je české digitální dědictví



Předpoklad 2

KONCEPTUALIZACE

- **vytvoření informačního modelu pro identifikátory**
 - míra abstraktnosti identifikátoru (jednotka vs. dílo)
 - granularita apod.
- **vazby na jiné identifikátory**
 - mj. vazba na původní zdroje (digitalizace)
- **definice terminologie**
- **stanovení pravidel, vč. signifikantních vlastností**
 - identifikace intelektuálního obsahu
 - identifikace digitálního objektu
- **definice trvalosti!! (ID samotné, vazba ID<>DO, systém ID)**
- **představa o možnostech nahrazení stávajících ID novým (tj. použití!)**



Předpoklad 3

ADMINISTRATIVA

- rozdělení odpovědnosti za administrativu a financování infrastruktury
- dodržování pravidel a modelu
- udržování systému ze strany jednotlivých aktérů
- zajištění všech přidružených činností, s kterými identifikace DD souvisí
- rozdíl oproti např. ISBN
 - (vydavatel > NK > ostatní knihovny pasivní využití)
 - URN:NBN – knihovna=vydavatel, aktivní údržba NK i ostatní knihovny (svou část DD)



Předpoklad 4

PŘIDRUŽENÉ ČINNOSTI

- **trvalé zpřístupňování DD**
 - správa digitálních knihovny
- **trvalá archivace DD**
 - zajištění uložení v TDR (ideálně certifikovaný TDR, síť TDR)
 - identifikovaný DD musí být stále „přístupný“
- **trvalá autentifikace DD označovaných pomocí URN:NBN (udržování linků aj.)**
- **správa systému přidružených metadat**



Předpoklad 5

AUTENTIFIKACE

- **zajistit, aby soubor všech počítačových souborů označovaných jedním URN:NBN**
 - byl udržován jako celek (nesmí se žádná část ztratit) – údržba metadat apod.
 - soubory nebyly porušeny
 - různé uživatelské verze odpovídaly signifikantním vlastnostem archivních verzí
 - aktualizace umístění URL

= úkol pro konkrétní knihovnu, ne pro NK



Předpoklad 6

METADATA

- **identifikátor bez systému přidružených metadat je k ničemu**
- **system metadat – stanovení základních metadat vázaných k identifikátoru**
 - zdaleka nestačí jen aktuální internetová adresa
 - autor, název, lokace v jednotlivých DK, další ID, různé verze DD aj.
- **údržba metadat v registru / databázi**
- **aktualizace metadat**



Předpoklad 7

KATALOGIZACE / DIGITALIZACE

- návaznost katalogizace a systému identifikace – silná závislost (přenos chyb apod.)
- systém identifikace zdrojových dokumentů v případě digitalizace
 - k URN:NBN musí být vázán systém identifikace tištěných publikací (např. na ISBN, čČNB)
- kvalitní katalogizace, případně rekatalogizace
 - nedigitalizovat nezkatalogizované nebo špatně publikace, vracet je
 - PRIMO v Singapuru (nutnost rekatalogizace, odložení projektu o 2 roky!)



Předpoklad 8

STRATEGIE VYUŽITÍ

- je nutné mít rozmyšleno na co se bude po zavedení ID používat, jaké ID nahradí (interní, externí komunikace knihovny) a jaké výhody identifikace přinese
 - slouží čtenářům nebo správě dokumentů?
 - slouží při digitalizaci nebo při archivaci?
- dle toho pak pro jednotlivé instituce (projekty) vybrat nejvhodnější systém identifikace (nezávisle na NK) > 1 systém tohle nevyřeší
- bez masového užití je nový systém v podstatě zbytečný!
 - tj. navázání stávajících procesů, systému apod. na nový identifikátor (opět otázka typu systému, jeho možnostech granularity apod.)



Infrastruktura

- všechny tyto předpoklady lze splnit jen v rámci AKTIVNÍ spolupráce všech zainteresovaných institucí
- zainteresovaná strana – každá strana, která chce přidělit URN:NBN
- centralizace – NK přiděluje ID (třetí strany volají generátor) a spravuje resolver a registr
- ostatní – vykonávání všech přídatných činností



Společný projekt

- **Jediné řešení (jinak se stále bavíme o nejasných zájmech jednotlivých institucí, které čekají na řešení NK, aniž by věděly co jim přesně přinese)**
- **Příležitost : NAKI (nástupce VaV)**
- **Vytvoření konceptualizace, pravidel, modelu, základů infrastruktury a dohod o jejím provozu**
- **V zahraničí předcházet všem aktivitám celonárodní projekt – vyjasnil zájmy, cíle, zájemce aj.**



Finance: Bundesministerium für Bildung und Forschung
Vedoucí: Národní knihovna + další knihovny a školy

Projekty: EPICUR, nestor

2002 – 2005 EPICUR

- infrastruktura na URN:NBN, původně pouze pro vysokoškolské práce (ALE např. GDZ využívá PURL, Handle)
- celonárodní program pro digitální archivaci; spolupráce všech institucí; identifikace je jednou ze sekcí

- 2009 – doporučení pro důvěryhodné identifikační systémy (tedy nedávno)
- chápou, že celonárodní identifikace digitálních dokumentů musí být spojena se správou a ochranou dat a stanovením národních standardů a doporučení a spoluprací všech zúčastněných stran



Finance: Nadace FRD (Fondazione Rinascimento Digitale) + Národní knihovna + univerzity

Projekt: JNBN 2008 – dosud

- distribuovaná správa identifikátorů v celonárodním měřítku
- resolver – finální plná verze – podzim 2009 (testy verzí v NK od jara 2009)
- využití resolveru pro mezinárodní spojení resolverů – odlišná architektura – každá instituce si drží vlastní registr



Austrálie

Finance: ARROW II (australský národní program pro výzkumné repozitáře)

Řešitelé: výzkumné instituce

Projekt: PILIN 2006 – 2009

- základ pro národní infrastrukturu identifikace pro výzkumné instituce:
- výstupy: konceptualizace, pravidla, výborná doporučení na co vše se zaměřit při plánování
- A Policy Checklist for Enabling Persistence of Identifiers“ [DLIB, 2009] + *překlad Knihovna 2010*



Skandinávie

Švédsko

Univerzita v Uppsale

Projekt: DIVA Portal 2000-dosud

- identifikace odborných publikací ve skandinávských univerzitách
- výstupy – **infrastruktura, resolver**



Projekt v NAKI

- širší záběr – nejen identifikace, ale i pravidla a standardy pro digitální knihovny, repozitáře, digitalizaci a katalogizaci – co se týče potřeb identifikace
- každá instituce bude mít určitý úkol/odpovědnost
- společná práce na modelu a pravidlech pro ČR (v oblasti URN:NBN)



Možné výstupy projektu

- model, pravidla, konceptualizace pro identifikaci s přesahem pro digitalizaci, katalogizaci, digitální knihovny a repozitáře
- zmapování ostatních identifikátorů (výhody/nevýhody jejich užití) <> lze použít různé systémy ID! současně
- vybudování infrastruktury – instituce (správa, aktualizace) – síť – centrální správa a přidělování
- vytvoření specializovaných pracovišť/znalostí (nejen v NK)
- SW nástroje – resolver + registr aj.
- další?? k debatě ...



Čeho se dosáhlo

- **Vazby na identifikaci provedení (manifestation) v ČNB – číslo národní bibliografie (čČNB)**
 - Marcovské pole 015
- **Stabilní řešitelský tým v NK**
- **Základní konceptualizace v ODO (zatím pro NK)**
 - Pilotní fáze resolveru a infrastruktury pro potřeby NK – zatím pouze pro digitalizovaná data (výzkumný záměr NK 2010), vývoj nástrojů a testování



Jak dál?

- **Zkusit NAKI v roce 2010 pro národní implementaci URN:NBN**
- **Vyhlášení květen/červen 2010 ??**
- **Úvodní informační schůzka březen 2010?**
- **Oslovení v první fázi – NTK, KNAV, MZK, UISK, UVT UK**



Bez společného projektu?

- Úhel pohledu NK
- Vyhovět potřebám NK a jejích systémů
- Omezená funkcionality
- Jen pro některé typy dokumentů
- Pomalejší vývoj



Bez společného projektu ?

Pokud nevyjde NAKI, lze čekat následující scénář

- **V rámci NK ověřit pilotní implementaci pro Kramerius > její rozvinutí pro NK a případně instituce, kt. Kramerius používají (K4)**
- **Snaha o implementaci URN:NBN do workflow digitalizace a dlouhodobého archivu NK & MZK (IOP projekt)**
- **Pokusit se rozumně využít URN:NBN v katalozích NK – tak aby to mělo smysl pro NK i uživatele**
- **Časem zkusit identifikaci dat z Webarchivu, Manuscriptoria, případně dalších projektů v NK**
- **Ostatní instituce si budou hledat vlastní řešení (handle, DOI, ARK, PURL aj.) nebo aktivní spolupráce s NK**



Děkujeme za pozornost

Debata?

Otázky?

Návrhy?

Za odbor digitální ochrany NK ČR

jan.hutar@nkp.cz

