

Persistent identifiers in the national bibliography context

Ladislav Cubr
Jan Hutař
Marek Melichar



National Library of the Czech Republic (NLCR)



Content

- Historical background: PIDs in CR
- Why do we need PIDs?
- National bibliography number (CNB number)
- URN:NBN:CZ – Issues
- URN:NBN:CZ – Next steps
- Questions?



Historical background: PIDs in CR

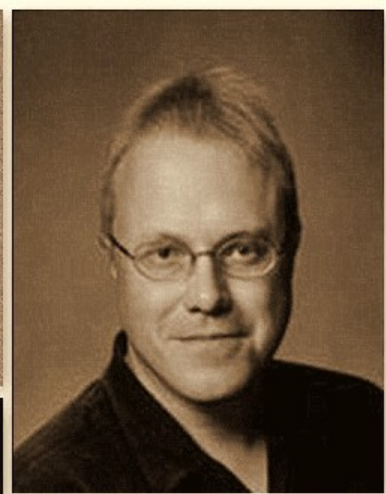
URN:NBN:CZ: are we really the first generation touching this issue?

CASLIN workshop 2001 – Juha Hakala
Czech Translation 2003

HAKALA, Juha. 2003. Popis dokumentů a přístup k nim - nové výzvy. In KLOUČKOVÁ, Z. a MACHALOVÁ, L. (eds.). *Moderní informační a komunikační technologie v knihovnictví 2003 : sborník příspěvků*. Praha : Státní technická knihovna, 2003, s. 22-30. ISBN 80-86504-09-3.



CASLIN 2001





Historical background: PIDs in CR

Hakala 2001 at CASLIN:

- clear ideas
- too optimistic expectations about the URN future

...we have two requirements for dynamic - and efficient - **resolution service** for electronic resources. **It must be able to adapt to the location (URL) changes**, and **it has to be able to take into account the user's priorities**, based on his/her credentials and / or location.... Last, but not least, **the system should be well integrated into the Internet infrastructure**.

In the present Internet, URN resolution will be based on Domain Name Service. The idea is that a user can type URN into the Location-window of his/her Web browser or other Internet client, just like URLs can be used now. Once the user has hit the Enter key, DNS service will find the resource using the metadata stored in the DNS system.

The URN:NBN has never been integrated into Internet infrastructure



Historical background: PIDs in CR

2007

- Working group for PID (10 partners, libraries + commercial companies)

Goals:

- reviewing current systems
- specify requirements of participating institutions

Results:

- use cases, analysis
- no solution for everyone ☹️
- go for URN:NBN



Currently used PIDs in CR

- 2007/08: implementation of Handles in NLCR (Kramerius digital library) for user copies access
- Other libraries started to use other PIDs (DOI, Handle in Dspace etc.)
- Repeatedly **expressed need for national bibliography** identifier for **digital world**



Why do we need PIDs?

- to connect the library catalogues with the digital collections (user copies)
- to connect the digitized material (user and archival copies) with the catalogues: to identify the right catalogue record of the „manifestation“ (FRBR) from which the „digital representation“ (PREMIS) was made
- Identify born digital collections (webarchiving)
- **Preserve access to the mass digitizing results**
 - mass digitization project likely to start 2010



Present Issues (NLCR)

- Catalogue duplications
 - no FRBR
 - various approaches to cataloguing
 - **what is the „intellectual entity“**(PREMIS) we need to identify?
- Problems of Union catalogue cannot be solved by any identifiers (de-duplication)
- Separation of archival copies, user copies, catalogue records.....



URN:NBN on the National level

What is needed on the **national level**?

The answer is **PROJECT** of few or more
institutions ...

...to design and implement URN:NBN:CZ

- **system**
- **infrastructure**



URN:NBN on the National level

Consortium of institutions which will

- decide the basic conceptual questions (what do we need to identify? computer files X documents?)
- formulate national policy for URN:NBN
- apply for the project funding (VISK3 for 2010?)
- establish organization and standards (who will be responsible for long-term storage and accessibility of the identified objects? admin rights? etc.)
- prepare functional requirements of the software
- develop and implement the software
- sort out financial issues – who, how?



Basic facts/needs to keep in mind

- 99% of the success of any identification system seeking to be persistent and unequivocal is about **the conceptualization and administration!**
- Identifiers lacking **descriptive and administrative metadata, rights, URLs** have no use
- Identifiers lacking **links to existing catalogues** have no use
- Identified objects must be **permanently accessible** and have lasting value and significance (stored in certified repository with elaborated preservation policy....)



Issues with URN:NBN

Which URN:NBN system do we need?

Different interpretations in different countries!

- **Germany** (identification based on computer files identification)
- **Finland** (identification based on intellectual content, i.e. extension of ISBN model for digital world)
- **Netherlands** (digital documents archived in e-depot only – articles)



What do we propose?

1) CNB number

- implement unique and persistent identifier of **Czech National Bibliography** elements/records (i.e identify the „manifestation“ of work in this database)
- the record in CNB database then may contain all identifiers of „digital representations“(PREMIS) of given „manifestations“(FRBR) (URN:NBNs or Handles to all digital copies)
- the record in CNB will include all other identifiers (ISSN, ISBN)
- use the CNB record as the base for identifying what is the basic „intellectual entity“(PREMIS)



What do we propose?

2) URN:NBN system

- Conceptual issues should be decided in the project by all the partners
- We have some ideas about how the system should work, but if this is to be a national identifier also others need to have their say (draft of the conceptualization to be available in end of October 2009 for public comment)



URN:NBN system

The URN:NBN:CZ infrastructure should include:

- **Register** of identifiers and all related metadata (CNB, URL, ISBN/ISSN, descriptive, administrative and rights metadata)
- **Resolver** (searching by metadata, URN:NBN or CNB, other identifiers,... – return full record with all links to computer files stored in different locations)
- **administration module** (harvesting metadata, generating URN:NBN numbers, checking URLs, managing user accounts, etc.)



URN:NBN system: Issues for the project

- Computer files or **intellectual content**?
- Centralization or decentralization of **responsibility**?
- Use **RD.CZ** database (base for digitized documents) as the data source and add additional databases for digital born documents?
- How to include **documents outside CNB**? Who will decide that the object has lasting value and significance?
- Do we need **identifiers for permanent access**? In such case there has to be a system of „**trusted repositories**“ which will store back-up copies
- Administration, organization, funding
- Programming: start from scratch or use Italian or other existing software?



URN:NBN:CZ

Next steps 2009

- **October 2009** – basic concept of URN:NBN implementation (NKP) distributed to relevant institutions
- **November 2009** – meeting of interested institutions (comments on the concept and expressing the will to participate at the project)
- **November/December 2009** – prepare project application to VISK3



URN:NBN:CZ

Next steps 2010

- **January 2010** – the group starts work on:
 - URN:NBN policy specification
 - Basic specification of the URN:NBN concept, use cases, review of global resolver project,
 - Functional requirements for the software
 - June 2010 – presentation of the results of the first phase to the community, review of feedbacks
 - October 2010 – tender for programming, web interface,
 - February 2011 – application of URN:NBN in testing,
 - June 2011 – application is running



Questions, comments?

jan.hutar@nkp.cz