Names, Things, and Open Identifier Infrastructure: N2T and ARKs

John Kunze
California Digital Library

Open identifier infrastructure?

- URL infrastructure is mostly open
- But "persistent identifier" infrastructure is mostly closed

Non-traditional id persistence

Traditional silos: doi.org, handle.net, purl.org, urn.arpa

2001: ARK (Archival Resource Key) scheme published

- Decentralized resolution shouldn't hurt, ie,

  http://soviet.org/ark:/12345/9876

  should be the same as

  http://russia.org/ark:/12345/9876

- Embryonic and dynamic content
  - Provider intention should be queryable

Persistence nonsense

URLs aren't long-term, instead use … (which are URLs)

URLs are locations, instead use … (which are URLs)

URL hostnames break, instead use … (which are URLs)

It's all about hosts – which of these do we trust?
- doi.org, purl.org, handle.net, n2t.net
- bnf.fr, bl.uk, loc.gov, nla.gov.au, nlm.nih.gov

ARKs breaking with tradition

- Syntax does not confer persistence
- ARK registry grows monthly, now 360+ insititutions
- ARKs appear along with DOIs in Thomson Reuters Data Citation Index
- ARK as a collection of ideas for ids

ARKs breaking with tradition

- Simple relationships shouldn't hurt
  - A/B means A logically contains B
  - A.B and A.C means A has two variant forms
- Machine-readable metadata shouldn't hurt
  - Méthode ancienne – label: value
  - "id?" requests metadata
  - "id??" requests more metadata
    - no conflict with "content negotiation"
- "Inflections" reduce need to create more ids

Breaking with centralized resolver tradition

2007: N2T (Name-to-Thing) resolver introduced

- Centralized resolution for ARKs, but also for any other identifier type
- Proposed consortium (on hold since 2008)
- Plus suffix pass-through (hierarchy shouldn't hurt)

N2T under the hood

- Built using open source components
- Replica running in Edinburgh
- Primary running in California
- Supports ARKs, DOIs, URNs
- Ids from EZID, Internet Archive, and YAMZ.net
- Experimental load of 60 million CrossRef DOIs

Summary – ARK and N2T

- Inflections (metadata shouldn't hurt)

- Structure with / and . (relations shouldn't hurt)

- Hostnames inert (decentralization shouldn't hurt)

- Id-agnostic resolution (centralization shouldn't hurt)

- Suffix pass-through (hierarchy shouldn't hurt)

- Let's build to the "Principles for open scholarly infrastructures" (Bilder, Lin, & Neylon)