Low-Risk Persistent Identification: the “Entity” (N2T) Resolver

10 October 2006

John Kunze, California Digital Library, University of California
NT2 “Entity” – overview

Establish a consortium and a small web server. Each member publishes URLs under n2t.info:

http://n2t.info/12345/foo/bar.zaf

...which redirects to the member server URL.

Why? It solves the same persistent identifier problem as URN, DOI, and Handle systems, but more fully, and at lower cost and risk.
Persistent identification

Persistent identifiers? We have them.
• But still need persistent actionable ids
Actionable “with widely available tools”
• Which really means “with URLs”
URNs, ARKs, Handles, DOIs, etc. become actionable (practically speaking) when embedded in URLs
• All these ids have similar maintenance costs, and they all break for the usual reasons
The usual reasons

Whatever the string, what matters is the *thing*

• If the thing’s unavailable, the id’s broken
  Broken, for URLs, means either
  • The hostname is broken
    – Server down, gone, or renamed *
    – Domain name lost, provider out of business
  • Or the pathname to the thing is broken
    – Thing down, gone, or renamed *

* No global fix for these, only the provider can fix.
Hostname instability

Domain name lost, provider out of business

- We can help this case

Smaller organizations most vulnerable

- The comfort of not seeing your hostname
- The comfort of seeing your hostname

- Traditional solutions: PURL, URN, Handle, DOI
- Solutions tied to special-purpose technology, sometimes complex and proprietary
N2T (Name-to-Thing)

N2T is two things at once

- A consortium of cultural memory organizations
  ... and ...

- A small, ordinary web server, mirrored in several instances globally for reliability

Basic idea: protect 200 organizations’ URLs from hostname instability with 200 rewrite rules

How: simple HTTP redirects, one per organization
N2T – user point of view

Each consortium member organization gets a unique number, such as, 12345.

Fig 1. A protected URL is forwarded to a provider currently serving things named by organization 12345. The *path part* is left alone.
N2T – system point of view

Technically, resolution (access to a thing given its name) is two simple steps.

Fig 2. N2t.info maps a protected HTTP URL into a provider URL.
N2T – consortium point of view

“Consortium-lite”
– Members have no fees or responsibilities
– One domain name for whole consortium
  Rent is $30/year, runs on 4 total web servers
Volunteer member orgs run the servers
– 1 primary + 3 mirrors
Interested bodies:  CENDI, DLF, DCC
Interested institutions:  CDL, NYU, NLA, …
N2T – global point of view

Regional (e.g., Europe, Asia, North America) clusters of mirrored resolver instances, with round-robin failover for redundancy, fault-tolerance, and load-sharing

- No browser modifications required
Org’n A’s namespace splits when B and C inherit its objects. Under the URN/Handle/DOI model, B must still forward to C.
Global resolver updating

Per-object resolver needs bulk updates
- Periodic harvest (e.g., daily) of table mappings
- From well-known provider-side web server files, e.g., tools and conventions similar to Google sitemaps
Prototype resolver

Sample identifiers at n2t.info – these work now

http://n2t.info/12345/libraries/visitor.html
http://n2t.info/13030/inside
http://n2t.info/urn:nbn:se:uu:diva-3324
http://n2t.info/ark:/13030/tf5p30086k

Incidentally, it can also redirect all URNs, DOIs, and Handles, e.g.,

The n2t.info persistent identifier resolver

Advantages

• Big reduction in architectural complexity
• No browser modification required
• Identifier scheme-agnostic
• No proprietary, special-purpose infrastructure to carry forward as a liability to persistence

John.Kunze@ucop.edu
California Digital Library (CDL)