NOID

NOID: Nice Opaque Identifier (Minter and Name Resolver)

Name:	NOID
Version:	0.424 (2006-04-21)
Status:	Beta
Specification:	NOID
Download:	http://search.cpan.org/~jak/Noid-0.424/

Have you ever noticed how some of the most "mission critical" identifiers in your daily life are numbers? How often do you use

- a driver's license number,
- a social security number, or
- a bank or credit card account number

instead of your name and address, or a photo of your honest, smiling face? We use numbers because they are short, precise, and opaque. Opaque identifiers, such as numbers or random combinations of letters, are useful as long-term descriptors for information objects because they don't contain information that is at risk of becoming untrue later.

Why Opaque Identifiers

Non-opaque descriptors represent object properties that change over time: subject classifiers, where an object "lives", the spelling of an author's name, etc. They can also be imprecise in large collections where a keyword or title search returns too many results. Moreover, unstable or impersistent identifiers, such as a web address that worked 6 months ago but not today, are a common complaint. So it is important to have precise, stable identifiers that don't include vague or changeable properties.

To help stability, an opaque identifier doesn't contain any information related to potentially changeable properties. For instance, if an identifier contains an organizational acronym and that organization is merged with another, there is often political pressure to break with the past, which means pressure not to support previously published identifiers in which the old acronym appears. Opaque identifiers also have the advantage that they can be short; for example, using combinations of letters and digits, only four characters are needed to represent as many as 1.6 million identifiers.

While opaque object identifiers have distinct advantages, they aren't always easy to use. They contain no widely recognizable words that allow people to guess what the object is, and are hard to repair because a typo doesn't create an obviously misspelled word.

Nicer Opaque Identifiers

This is where NOID (rhymes with "employed") comes in.

The NOID software tool mints (generates) opaque identifiers and tracks information to help them remain unique, stable, and closely connected to the objects that they identify. These identifiers should be opaque enough to age and travel well, but should easily resolve (connect you) to objects and to their descriptions.

Identifiers minted by NOID have long-term and short-term uses. For example, NOID can mint transaction identifiers and short-term web session keys. A more visible use of NOID is to mint identifiers for the purpose of creating long-term persistent object names (e.g., ARKs, Handles); embedded inside a URL, such an identifier can provide object access when entered into a web browser.

How NOID Works

NOID starts out by creating a small, fast database to make sure that no identifier is ever minted twice. At that time you specify the format of the identifiers you want, and you can ask for a "check character" to be added upon minting that will later allow detection of the most common transcription errors. Once it's up and running, you can mint identifiers at will until the available identifiers run out, at which point you can create a new minter. The cost to set up or take down a minter is low, so it is not uncommon for an organization to run dozens of minters (for different purposes) at once; guidelines are under preparation for running multiple minters, keeping identifiers unique between different minters, etc.

Noids (identifiers minted by NOID) can be minted remotely at a central location in your organization's internal web, or minted directly ("command-line") by a program that doesn't require network access. The CDL uses both approaches in managing its own identifiers, and also supports a minter operated remotely by the Internet Archive for its mass book digitization effort in the Open Content Alliance. The CDL is considering setting up a remote minter that will allow non-CDL users to generate unique, "preservation ready" identifiers of their own.

Noids and ARKs

Noids are not the same thing as ARKs, but can be used to form ARKs. ARKs are persistent identifiers that are actionable (work in your web browser) and will connect you to object metadata by adding a '?' to the end. A number of organizations use NOID to create a core identifier, such as

```
13030/tf5p30086k
```

and then embed that Noid in a URL to create an ARK, such as

```
http://ark.cdlib.org/ark:/13030/tf5p30086k
```

The NOID tool is not necessary to generate ARKs, but is used for that purpose by organizations such as

- the National Library of France,
- the Internet Archive,
- Portico (the permanent archive of electronic scholarly journals),
- University of California, Berkeley
- New York University.

NOID is also used to generate Handle identifiers at Cornell University and North Carolina State University. Programmers at Princeton University have developed graphical user interfaces for NOID and ARK.

NOID as a Name Resolver, and the Name-to-Thing (N2T) Resolver

What many organizations need to help make their URLs more persistent is a way to take incoming requests for those URLs and redirect ("forward") them to the present object locations. The idea is that their published (persistent) URL need never change provided that the actual location (a different URL that is not suitable for long-term reference) can change whenever they move the object. A system that redirects names this way is known as a name resolver.

NOID can be set up as a name resolver working behind a web server. There it acts as an on-the-fly translator of the permanent published "names" (URLs) into temporary locations. In this type of arrangement, NOID maintains a fast lookup table that is consulted each time a web browser requests a long-term URL from the server. The NOID table of locations can be maintained centrally and updated remotely in a manner similar to remote minting.

One persistence threat that NOID by itself cannot guard against is when an organization and its web server cease to exist. For this reason a global "Name to Thing" (N2T) resolver has been proposed that represents an entire consortium of cultural memory organizations (the proposal uses some of the advanced resolver functions of NOID).

Related Information

The above represents a simplified taste of the complex issues around opaque identifiers, persistence, and name resolution. To get started creating ARKs, first fill out the ARK request form, and to come up to speed quickly, see the ARKs FAQ.

- ARK (Archival Resource Key)
- N2T (Name-to-Thing) Resolver