ARK

ARK (Archival Resource Key) Identifiers

ARKs are URLs designed to support long-term access to information objects. In 2001 ARKs were introduced to identify objects of any type:

- digital objects – documents, databases, images, software, websites, etc.
- physical objects – books, bones, statues, etc.
- living beings and groups – people, animals, companies, orchestras, etc.
- intangible objects – places, chemicals, diseases, vocabulary terms, performances, etc.

ARKs are assigned for a variety of reasons:

- affordability – there are no fees to assign or use ARKs
- self-sufficiency – you can host ARKs on your own web server, eg, Noid (Nice Opaque Identifiers) open source software
- portability – you can move ARKs to other servers without losing their core identities
- global resolvability – you can host ARKs at a well-known server, eg, at the N2T.net (Name-to-Thing) resolver
- density – ARKs handle mixed case, permitting shorter identifiers (CD, Cd, cD, cd are all distinct)

Some advantages of ARKs:

- simplicity – access relies only on mainstream web "redirects" and ordinary "get" requests
- utility – with "inflections" (different endings), an ARK should access data, metadata, promises, and more
- compatibility – inflections don't conflict with "linked data content negotiation" (a harder and limited way to access metadata)
- versatility – ARKs support persistence statements to describe different kinds of long-term access
- transparency – no identifier can guarantee stability, and ARK inflections help users make informed judgements
- visibility – syntax rules make ARKs easy to extract from texts and to compare for variant and containment relationships
- openness – unlike other persistent identifiers, ARKs don't lock you into one specific, fee-based management and resolution infrastructure
- impact – ARKs appear in Thomson Reuters' Data Citation Index and ORCID researcher profiles

Since 2001 over 550 organizations spread across fifteen countries registered to assign ARKs. Registrants include libraries, archives, museums (Smithsonian), publishers, government agencies, academic institutions (Princeton), and technology companies (Google). Some of the major users are

- The California Digital Library
- The Internet Archive
- National Library of France (Bibliothèque nationale de France)
- Portico Digital Preservation Service
- University of California Berkeley
- University of North Texas
- University of Chicago
- University College Dublin
- The British Library

There is a discussion group for ARKs (Archival Resource Keys) at

https://groups.google.com/group/arks-forum

The group is intended as a public forum for people interested in sharing with and learning from others about how ARKs have been or could be used in identifier applications.

The forum is also intended as a mechanism for the CDL, in its role as the ARK scheme maintenance agency, to seek community feedback on a number of longer term issues and activities, including

- finalizing the ARK specification as an Internet RFC,
- clarifying local and global resolution options, and
- promoting metadata retrieval in a linked data environment.

Here is a brief summary of other resources relevant to ARKs.

- The ARK Identifier Scheme Specification PDF version TXT version
- Towards Electronic Persistence Using ARK Identifiers (July 2003)
- ARK and CDL Identifier conventions
- Archival Resource Key - Wikipedia
- Noid (Nice Opaque Identifiers) open source software for minting and resolving ARKs on your own
- ARK plugin for Omeka that creates and manages ARKs for the Omeka open source web-publishing platform
- EZID service: long term identifiers made easy, if you would rather not install and maintain those services yourself
- N2T.net resolver: Name-to-Thing, a single global resolver at n2t.net

ARK Anatomy

An ARK is represented by a sequence of characters that contains the label, "ark:". When embedded in a URL, it is preceded by the protocol ("http://" or "https://") and name of a service that provides support for that ARK. That service name, or the "Name Mapping Authority" (NMA), is mutable and replaceable, as neither the web server itself nor the current web protocols are expected to last longer than the identified objects. The immutable, globally unique identifier follows the "ark:" label. This includes a "Name Assigning Authority Number" (NAAN) identifying the naming organization, followed by the name that it assigns to the object. Please visit the NAAN request form if you are interested in generating and using ARKs for your information objects.

Here is a diagrammed example:

```
http://example.org/ark:/12025/654xz321/s3/f8.05v.tiff
   \_________________/ \__/ \___/ \_____/ \____________/
     (replaceable)     |     |      |       Qualifier
          |       ARK Label  |      |    (NMA-supported)
          |                  |      |
Name Mapping Authority       |    Name (NAA-assigned)
         (NMA)               |
                  Name Assigning Authority Number (NAAN)
```

The ARK syntax can be summarized,

```
[http://NMA/]ark:/NAAN/Name[Qualifier]
```

The NMA part, which makes the ARK actionable (clickable in a web browser), is in brackets to indicate that it is optional and replaceable. ARKs are intended to work with objects that last longer than the organizations that provide services for them, so when the provider changes it should not affect the object's identity. A different provider hosting the object would simply replace the NMA to reflect the new "home" of the object. For example,

```
http://bnf.fr/ark:/13030/tf5p30086k
```

might become

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

Note that the ark:/NAAN/Name remains the same.

NAAN: the Name Assigning Authority Number

The NAAN part, following the "ark:" label, uniquely identifies the organization that assigned the Name part of the ARK. Often the initial access provider (the first NMA) coincides with the original namer (represented by the NAAN), however, access may be provided by one or more different entities instead of or in addition to the original naming authority.

The NAAN used above, 13030, represents the California Digital Library. As of 2018, over 550 organizations have registered for ARK NAANs, including numerous universities, Google, the Internet Archive, WIPO, the British Library, and other national libraries.

Any stable memory organization may obtain a NAAN at no cost and begin assigning ARKs. Please contact the CDL if you are interested in generating and using ARKs for your information objects.

CDL maintains a complete registry of all currently assigned NAANs, which is mirrored at the (U.S.) National Library of Medicine and the Bibliothèque nationale de France.

Creating and Managing ARKs

Once your organization has a Name Assigning Authority Number (NAAN), you may begin using it immediately to assign ARKs.

In thinking about how to manage the namespace, you may find it helpful to consider the usual practice of partitioning it with reserved prefixes of,

say 1-5 characters, eg, names of the form "ark:/*NAAN*/xt3...." for each "sub-publisher" in an organization. Opaque prefixes that only have meaning to information professionals are often a good idea and have precedent in schemes such as ISBN and ISSN. The ARK specification is currently the best guide for how to create URLs that comply with ARK rules, although it is fairly technical.

You can use any system you wish to manage your identifiers. One approach is to create and assign ARKs as a side-effect of deposit into a content repository, with ARKs publicized as being hosted on your server, eg,

```
http://myrepo.example.org/ark:/12345/bcd987
```

Another option is to use the EZID service (http://ezid.cdlib.org), which means your ARKs would appear to be hosted at n2t.net, as in

```
http://n2t.net/ark:/12345/bcd987
```

As with any identifier scheme, persistence requires a redirectable reference to content in stable storage. EZID operates on a cost-recovery basis and can be used to manage your namespace, which includes minting and resolving ARKs (and other identifiers), as well as maintaining metadata. There's is also guidance on CDL Identifier Conventions available.

Because long-term identifiers often look like random strings of letters and digits, organizations typically use software to generate (or mint, in ARK parlance) and track identifiers. To mint ARKs, you may use any software that can produce identifiers conforming to the ARK specification. CDL uses the open source Noid (nice opaque identifiers, rhymes with "employed") software, which creates minters and accepts commands that operate them. The noid software documentation explains how to use noid not only to mint identifiers but also to serve as an institution's "identifier resolver".

Once minted and publicized as being associated with a specific object, the ARK becomes a stable, unique, and compact reference that can be included in metadata records, databases, redirection tables, etc. It is often useful to generate and assign ARKs well before institutional commitment has been decided because it is easier than changing the original object identifier that may have been in long established use prior to that decision.

ARKs in Action – Inflections

An ARK provides extra services above and beyond that of an ordinary URL. Instead of connecting to one thing, an ARK should connect to three things:

- the object itself,
- a brief metadata record if you append a single question mark to the ARK, and
- a maintenance commitment from the current server when you append two question marks.

This is a achieved through the use of "inflections", or different kinds of endings. With no ending, the ARK (in a URL) gives you what you expect from a web browser. If you add a single '?' to the end, for example,

```
http://texashistory.unt.edu/ark:/67531/metapth346793/?
```

it returns a brief machine- and eye-readable metadata record; in this case, an Electronic Resource Citation (ERC) using Dublin Core Kernel meta data., such as

```
erc:
who: Dallas (Tex.). Police Dept.
what: [Photographs of Identification Cards]
when: 1963
where: http://texashistory.unt.edu/ark:/67531/metapth346793/
```

Adding '??' to the end should return a policy statement. It is a side-benefit of ARKs that an object's metadata doesn't need an identifier different from that for the object, which cuts in half the number of identifiers that need to be generated and managed.

CDL Name Assignment and Support Policy Statements

The CDL assigns identifiers within the ARK domain under the NAAN 13030 and according to the following principles:

- No ARK shall be re-assigned; that is, once an ARK-to-object association has been made public, that association shall be considered

unique into the indefinite future.
- To help them age and travel well, the Name part of CDL-assigned ARKs shall contain no widely recognizable semantic information (to the extent possible).
- CDL-assigned ARKs shall be generated with a terminal check character that guarantees them against single character errors and transposition errors.

Institutions that generate ARKs may want to follow similar principles or develop their own assignment policies.

Similarly, but in the role of an NMA and not an NAA, institutions will want to develop service commitment statements for the objects themselves. These NMA commitments are different from NAA identifier assignment policies. In many cases, the NAA will operate initially as the first NMA, but for long-lived objects over time, chances are that these will become different organizations (e.g., a highly successful object may easily outlive its NAA).

In developing such statements, it is useful to recognize first, that managing a digital object may require altering it as appropriate to ensure its stability, and second, that the declared level of commitment may change as the requirements and policies for persistence become better understood over time, and as the institution implements procedures and guidelines] for maintaining the objects that it manages. The US National Library of Medicine has developed some permanence ratings that may be of interest here.

There is also information available about CDL Identifier Conventions.