



CDL

CALIFORNIA DIGITAL LIBRARY

# The Archival Resource Key – A Tool for Providing Persistent Access

A presentation prepared for the ERPANET Seminar  
on Persistent Identifiers

By Mary K. Heath

California Digital Library, University of California  
Office of the President

---

June, 2004

# Needs satisfied by ARKs

- A simple system with a low overhead for maintenance
- A non-proprietary system
- A long-term directly actionable identifier
- A policy separation between the service provider and the assigner of the identifier
- Creation of IDs that are unique in the world
- Support for attached metadata and persistence level

# An ARK is a unique actionable URL

<http://ark.cdlib.org/ark:/13030/ft4m3nb2k3>

\-----/      \-----/ \-----/

|

|

|

Name Mapping

Name

Object

Authority

Assigning

Name

Authority

# The Name Assigning Authority

- Might be a national library, a national archive, or a publisher or repository
- Establishes long-term associations between identifiers and objects
- Are designated by Name Assigning Authority Numbers, e.g.

12025	National Library of Medicine
1226	Library of Congress
12027	National Agriculture Library
13030	University of California California Digital Library
13038	World Intellectual Property Organization

# Name Mapping Authorities

- Name Mapping Authorities provide services for Name Assigning Authority Numbers
- They resolve ARKs
- They are temporary, disposable, and replaceable
- They provide identity-inert mechanisms for launching ARKs into the web
- NMAs can supply redirection for NAANs

# ARK redirection mechanisms

- Through a simple file lookup:
  - # National Library of Medicine
  - 12025: <http://www.nlm.nih.gov/xxx/naapolicy.html>  
[ark.nlm.nih.gov](http://ark.nlm.nih.gov) USNLM
  - # University of California California Digital Library
  - 12026: <http://www.cdlib.org/inside/diglib/ark>  
[ark.cdlib.org](http://ark.cdlib.org) UCCDL
  
- Through a mechanism similar to the DNS NAPTR process (planned)

# More on Name Assigning Authorities

- The NAA controls the assignment of object names
  - Names must be unique within the NAA
  - The prepending of the NAAN makes objects different between NAAs:

<http://foobar.zaf.org/ark:12025/654ab332>

<http://foobar.zaf.org/ark:13030/654/ab332>

- But the NMAH is not a part of the identity, so the following identify the same object:

<http://foobar.zaf.org/ark:/12025/654ab332>

<http://ark.cdlib.org/ark:/12025/654ab332>

# More on Name Assigning Authority

## (continued)

- Object names are constructed according to a set of rules
- Object names may be complex, e.g.
  - ark:/12025/654/xz/321
  - ark:/12025/654/xz
  - ark:/12025/654
- Object names may reveal variants
  - ark:/12025/f67k3xz.052504
  - ark:/12025/f67k3xz.eng

# ARK services

There are three ARK services:

- The digital object
- Metadata describing the digital object

|set: cdlib.org | ark:/13030/ft4m3nb2k3? | 20040525204348 |  
ark:/13030/ft4m3nb2k3

erc:

who: Jonathan Dewald

what: Aristocratic Experience and the Origins of Modern Culture:  
France, 1570=1715

when: [:] 19330318

where: <http://ark.cdlib.org/ark:/13030/ft4m3nb2k3>

# ARK services (continued)

- A persistence commitment statement

erc-support:

who: cdlib.org

what: (:tba) Commitment statement pending.

when: 20031219

where: <http://www.cdlib.org/inside/diglib/ark/>

# ARK implementation

- The scheme is independent of specific implementation; functions required are:
  - A minting process
  - A process for binding objects to names
  - An ARK server
  - A resolution mechanism
- Other useful tools include:
  - A repository ingest function that can mint and bind names for a group of objects
  - A maintenance interface

# Costs of ARK implementation

- Similar to the costs of any other persistence scheme
- Requirement for metadata creation may seem costly
- There may be unexpected decisions required:
  - Level of assignment of object names
  - What the maintenance process will allow
  - Storing and designation of multiple object versions
  - Reassignment of identifiers
  - How the object content is allowed to change
- Development of a persistence policy can be organizationally time consuming

## Examples of permanence ratings

An organization may choose to define permanence ratings in terms of availability and content stability:

- Permanent – unchanging content
- Permanent – stable content
- Permanent – dynamic content
- Permanence not guaranteed

# ARK support of persistence

Persistence is a matter of the service intent of the NMA that is providing service for the object; ARKs assist this by:

- Requiring the development of a persistence policy
- Allowing automation of probes of metadata, object, and commitment statement
- Tying the organizational track record to consumer's trust in service
- Providing a strategy for long-term resolution
- Providing easily locatable addresses
- Being directly actionable

# ARK metadata requirements

- The ARK service is agnostic in terms of metadata format
- It can return Dublin Core, METS, or other XML-formatted metadata
- The basic ARK service returns ERC (Electronic Resource Citation) format
- ERC is both eye- and machine-readable
- ERC facilitates the normalization of data by collocating similar data elements into a single field

# A bit about ERC

ERC requires:

- A definition of what the record is
- Who – statement of responsibility for item represented
- What – what the item represented is
- When – date of creation of item represented
- Where – the location of the digital representation

## A bit about ERC (continued)

erc:

who: Ibrahim Muhawi (Darwish) and Sharif Kanaana

what: Speak, Bird, Speak Again: Palestinian Arab Folktales

when: [:] 19890213

where: <http://ark.cdlib.org/ark:/13030/ft4s2005r4>

format: text/html

ERC also provides options for extension

# ARK interaction with other systems

- Works well with other systems
- Can be used as links in web pages
- Can be referenced in OpenURLs
- OAI harvesters can easily collect metadata from ARK-based repositories
- ARKs can contain other identifiers such as PURLs, ISSNs, or ISBNs

# Conclusion

- Serious providers of digital content have a more complex service burden than do maintainers of ordinary websites
- ARKs provide a low-cost method for facilitating the management of this burden
- ARKs are actionable URLs that encapsulate a globally unique identity
- ARKs work with today's browsers as well as tomorrow's tools
- ARKs both demand and reveal an organizational commitment to persistence