Integrations with ArchivesSpace Webinar Series: ARKs and ArchivesSpace

March 4, 2020 – Webinar
Presenters

John Kunze
California Digital Library

Seth Shaw
University of Nevada, Las Vegas

Christine Di Bella
ArchivesSpace
ARK design and current landscape

John Kunze, California Digital Library
March 2020

University of California
- 10 campuses, 100s of museums, archives, galleries, etc.
- plus 5 medical centers, 5 law schools, 3 national labs
- 10% of US research output
Why care about ARK identifiers?

Because persistent, reliable web links are lacking.

- The average URL lifetime is 44 days
- Maybe ok for the rest of the world, but not for archives and libraries
- URLs in vendor archiving tools (a) can break between major system releases and (b) aren’t generally portable to another vendor’s system

Wanted: flexible, low cost, vendor- and software-independent \textit{persistent} identifier
ARK (Archival Resource Key)

- ARK: a persistent link for any kind of thing
- 3.2 billion ARKs created by 550 institutions: libraries, archives, museums, publishers, educators, etc. Examples:

  Internet Archive
  Caltech Archives
  Hawaii State Archives
  French National Archives
  Rockefeller Archive Center
  Library and Archives Canada
  Archives de la Ville de Genève
  Silent Film Sound & Music Archive

  University of California Berkeley
  Smithsonian National Museum
  National Library of France
  University of Chicago
  Musée du Louvre
  Family Search
  British Library
  Google
ARK anatomy

A labelled URL with a globally unique identity inside it

https://n2t.net/ark:/12345/fk1234

makes ARK actionable (the resolver)

core globally unique identity (independent of web and hostname)
What are ARKs used for?

- genealogical records (3 billion FamilySearch)
- publisher content (100 million Portico)
- scientific datasets and records (22 million INIST)
- scanned books and texts (23 million Internet Archive)
- bibliographic records (15 million BnF main catalog)
- museum specimens (15 million Smithsonian Institution)
- public health documents (14 million UCSF IDL)
- digitized documents (21 million CDL, 5 million BnF Gallica)
- historical authors and scholars (4 million SNAC)
- vocabulary terms (9,000 Periodo, YAMZ)
Why ARKs and not DOIs (or Handles or PURLs or URNs)?

- Google for “ten persistent myths about persistent identifiers”
- Flexible resolution centrally (n2t.net) or via your own server
- University of California’s history of *open*
  - 1968 Free speech movement
  - 1982 Open source Berkeley UNIX – FreeBSD – Mac OS X
  - 2013 Groundbreaking open access research policy
  - 2019 Termination of Elsevier journal subscription
  - 2001-2020 Non-paywalled, decentralized persistent identifiers – ARKs
ARKs and DOIs

DOIs (Digital Object Identifiers) – publishing industry solution

- requires membership, per-identifier fees, rigid metadata requirements

ARKs - cultural heritage solution

- no fees or membership, and highly flexible creation and metadata policies
- like DOIs, ARKs are stable, linked to metadata, and are found in the Data Citation Index (linked to the Web of Science), Wikipedia articles, Wikidata records, Internet Archive collections, ORCID researcher profiles, etc.
The Covenant of the ARK

The ARK scheme

- will not charge fees to create or use ARKs
- will not limit the number of ARKs you assign
- will not limit the kind of content you identify
- will not require metadata, nor even persistence
- will not mandate use of any particular resolver
Getting involved in ARKs

- Learn more: “ARK Identifiers FAQ”
- Start assigning: visit n2t.net/e/naan_request

Support open infrastructure: join us at ARKsInTheOpen.org

- Community owned infrastructure
- Collaboration between CDL and LYRASIS
Finding Aid ARKs @ UNLV
Places Finding Aid URLs go...

- Blog Posts
- Collection Guides
- Aggregators
- Catalogs
- Citations
- Users' Bookmarks
ARKs for Combating Link Rot

Local DB -> PDF -> EAD -> ???

I really don't want to, and in many cases can't, go update old URLs.
Enter persistent identifiers!
Guide to the Harrah's Entertainment Corporate Archives

This finding aid was created by Hana Gutierrez and Hannah Robinson on August 25, 2017.

Persistent URL for this finding aid: http://n2t.net/ark:/62930/f1159j
Where do we put the ARK in ArchivesSpace?
### Finding Aid Data

<table>
<thead>
<tr>
<th>EAD ID</th>
<th>US::NvLN::MS00460</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EAD Location</strong></td>
<td><a href="http://n2t.net/ark:/62930/f1159j">http://n2t.net/ark:/62930/f1159j</a></td>
</tr>
</tbody>
</table>

**Finding Aid Title**: Guide to the Harrah's Entertainment Corporate Archives

**Finding Aid Filing Title**: Harrah's Entertainment Corporate Archives

**Finding Aid Date**: © 2017

**Finding Aid Author**: Hana Gutierrez and Hannah Robinson

**Description Rules**: Describing Archives: A Content Standard

**Language of Description**: English

**Finding Aid Status**: Completed
Identifier Details

Edit and add descriptive information to your identifier.

ABOUT THE IDENTIFIER

Identifier: ark:/62930/f1159j
Identifier as URL: https://n2t.net/ark:/62930/f1159j

ABOUT THE IDENTIFIED OBJECT

Location (URL): http://www.library.unlv.edu/speccol/finding-aids/MS-00460.pdf
Who: University of Nevada, Las Vegas University Libraries
What: Guide to the Harrah’s Entertainment Corporate Archives
When: 1811-2004
Manual Workflow (what we don't do)

1. Complete the Finding Aid
2. Use the EZID web interface to mint an ARK
3. Copy that ARK into ArchivesSpace
4. Generate a PDF Finding Aid
5. Copy the PDF Finding Aid to our webserver
6. Update the corresponding links in the online database
Semi-automated Workflow (what we actually do)

1. Create a list of completed Finding Aids
2. Run a script which:
   a. Mints an ARK via EZID API (if one doesn't exist)
   b. Updates the ArchivesSpace resource via ArchivesSpace API (if one is minted)
   c. Generates a PDF Finding Aid via ArchivesSpace API
   d. Copies the PDF Finding Aid to our webserver
3. Another script to update links in the online database

https://github.com/seth-shaw-unlv/small-stuff/blob/master/as-coll-mint/as_coll_mint.py
What we really want to do...

Publish All
When the time comes to migrate systems we can *batch update ARK/URL pairs* via the EZID API!

Seth Shaw
Application Developer
University Libraries, Web and Application Development Services
University of Nevada, Las Vegas
seth.shaw@unlv.edu
ARKs in ArchivesSpace

Christine Di Bella
ArchivesSpace Program Manager
March 2020
ARKs in ArchivesSpace

• Added for version 2.7.0.
• Optional feature that can be used with any resource or archival object
• More information about using ARKs in ArchivesSpace is available in the ArchivesSpace User Manual

https://archivesspace.atlassian.net/wiki/spaces/ArchivesSpaceUserManual/pages/919502964/Archival+Resource+Keys+ARKs
ARKs in ArchivesSpace

• ARKs can be generated by ArchivesSpace (internal ARKs) or can use ARKs generated by another system (external ARKs).
• External ARKs take precedence over internal ARKs in ArchivesSpace.
• Internal ARKs resolve to a record in the public user interface of ArchivesSpace.
• When enabled ARKs appear in some places in the ArchivesSpace interface and are included in exports and OAI-PMH harvests
### Configuration options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>arks_enabled</strong></td>
<td>Determines whether the options for creating internal and external ARKs are shown in the staff interface.</td>
</tr>
<tr>
<td><strong>ark_naan</strong></td>
<td>The Name Assigning Authority Number that is permanently assigned to an institution. This should be set to yr institutional NAAN. If your institution does not already have a NAAN, to use ARKs in production you should request a NAAN from the California Digital Library. A request form is linked from the information page at <a href="https://n2t.net/e/ark_ids.html">https://n2t.net/e/ark_ids.html</a>.</td>
</tr>
<tr>
<td><strong>arks_url_prefix</strong></td>
<td>The URL prefix to use for the internal ARK URLs. It is defaulted to the Public User Interface URL, since, in most cases, that is what it will be.</td>
</tr>
</tbody>
</table>
If arks_enabled has been set to true:

• ARKs will be automatically generated when new resources and archival objects are created;

• External ARKs may be input when in edit mode for resources and archival object records; and

• ARKs may be generated in bulk for all resource and archival object records in all repositories, when running the Generate ARKs in Bulk background job.

• ARKs do not delete/change your URIs (i.e. /repositories/2/resources/1) – they will still work as they did before.
Demo
Questions?

John Kunze  
California Digital Library

Seth Shaw  
University of Nevada, Las Vegas

Christine Di Bella  
ArchivesSpace

Thank you to our presenters!