Getting started with ARK persistent identifiers and their cross-domain metadata

John Kunze[1,*] and Donny Winston[2]

[1] *Drexel University, Metadata Research Center, Philadelphia, PA, USA*
[2] *Polyneme LLC, New York, NY, USA*

Abstract

This tutorial introduces ARK (Archival Resource Key) persistent identifiers and their unique approach to cross-domain metadata. As non-paywalled PIDs (persistent identifiers, permalinks) for information objects of any kind, ARKs support durable web addresses (e.g., that don't return 404 Page Not Found). Since 2001, 8.2 billion ARKs have been created by over 1300 organizations — libraries, data centers, archives, museums, publishers, government agencies, and vendors. With highly flexible metadata, both in application profile and in access, citation-friendly ARKs identify anything digital, physical, or abstract. The tutorial includes hands-on experience and is for is anyone interested in PIDs supporting cross-domain metadata and nuanced persistence policies.

Keywords

PIDs, permalinks, preservation, curation, crowdsourced, dictionary

1. Topics to be covered

In this 90-minute tutorial we will introduce you to ARKs (Archival Resource Keys), which can serve as persistent identifiers, or stable, trusted references for information objects (e.g., web addresses that don't return 404 Page Not Found errors). Without fees or centralized control, ARKs are among the most FAIR of identifiers.

In more than two decades, 8.2 billion ARKs have been created by over 1200 organizations — libraries, data centers, archives, museums, publishers, government agencies, and vendors. Highly flexible and non-paywalled, ARKs are adopted increasingly by organizations in the global South and by those that need large numbers of identifiers. ARKs are citation-friendly identifiers providing an open API to connect to FAIR metadata and persistence statements.

ARK metadata is uniquely flexible, including important persistence vocabulary terms that are being drafted and discussed in the context of YAMZ.net, a crowdsourced dictionary. YAMZ terms can be proposed by anyone and, as it happens, each term is also automatically assigned an ARK that resolves to its definition. Terms can be upvoted and downvoted and have no recognized standing unless referenced by a metadata profile or standard. This domain-agnostic dictionary provides a compelling use case for ARK persistence metadata.

We will cover:

- Why ARKs – non-paywalled, decentralized, flexible

* Corresponding author.
jakkbl@gmail.com (J. Kunze); donny@polyneme.xyz (D. Winston)
0000-0001-7604-8041 (J. Kunze); 0000-0002-8424-0604 (D. Winston)

https://doi.org/10.23106/dcmi.952467990

- Use cases – Smithsonian, French National Library, Internet Archive
- Metadata for early and ongoing object development
- Metadata for persistence
- Minting and assigning ARK identifiers
- Creating and resolving metadata vocabulary
- Resolvers, resolution, redirection
- Persistence considerations

2. Target audience and expected learning outcomes

Modifying the template — including but not limited to adjusting margins, typeface sizes, line spacing, paragraph and list definitions — is not allowed.

With guided exercises, by the end of the session, participants will know when and how to create and manage ARKs, as well as how to publish their own local metadata terms.

3. Tutorial style

Lecture with hands-on practice. The format will include presenter-led instruction, discussion questions, and hands-on experience of ARK organization registration, as well as metadata term creation and resolution.

- No prior knowledge is required.
- A laptop with wifi connection is recommended but not required.
- Familiarity with basic website management is recommended but not required.

4. Length and number of attendees

- Proposed session length: 90 mins.
- Maximum number of attendees: 60.

5. Presenter biographies

John Kunze is a pioneer in the theory and practice of digital libraries. With a background in computer science and mathematics, he wrote BSD Unix tools that come pre-installed with Mac and Linux systems. He created the ARK identifier scheme (arks.org), the N2T.net scheme-agnostic resolver, and contributed heavily to the first standards for URLs (RFC1736, RFC1625, RFC2056), library search and retrieval (Z39.50), archival transfer (BagIt - RFC8493), web archiving (WARC), and metadata (RFC2413, RFC2731, ANSI/NISO Z39.85).

Donny Winston has a bachelor's degree from the University of California at Berkeley, and master's and Ph.D. degrees from the Massachusetts Institute of Technology, all in Electrical Engineering and Computer Science. His doctoral area of specialization was nanofabrication, with both experimental and computational work. Donny has worked in a research capacity at IBM Research in Yorktown Heights, NY; at the National Institute for Standards and Technology

https://doi.org/10.23106/dcmi.952467990

(NIST) in Gaithersburg, MD; at Carl Zeiss R&D in Peabody, MA; at Hewlett-Packard Labs in Palo Alto, CA; and at Lawrence Berkeley National Laboratory (LBNL) in Berkeley, CA.

https://doi.org/10.23106/dcmi.952467990