The ARK Alliance:
20 years
850 institutions
8.2 billion persistent identifiers

John Kunze, California Digital Library
University of California Office of the President

October 2021
Digital preservation means

Long term *protection* for digital resources

- from human error, natural disaster, legal challenge, deliberate attack, social upheaval, bankruptcy, etc.

Long term *access* to those resources from unbroken links

- with *persistent identifiers (PIDs)*, also known as *permalinks*
Why persistent identifiers?

Because of “link rot” (broken references, 404 Not Found)

- Reliable, unbroken web links (URLs) are rare
- The average URL lifetime is only 100 days

But why not just search when you need a link?

- Because scholars and researchers take years to find their object references

Common types of persistent identifiers

- PURL, Handle, URN, DOI, ARK
What is an ARK (Archival Resource Key)?

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)
ARK anatomy

https://example.org/ark:/12345/x54xz321/s3/f8.05v.tiff
\_______________/  \__/  \___/  \_____/\___/\_______/

ARK Label  Sub-parts Variants

Name Mapping Authority (NMA)  Assigned Name

Name Assigning Authority Number (NAAN)
# Why ARKs?

<table>
<thead>
<tr>
<th>Major causes of broken links, and some features</th>
<th>PURL</th>
<th>Handle</th>
<th>URN</th>
<th>DOI</th>
<th>ARK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevents fire, war, flood, attack, bankruptcy, …</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Prevents human error</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Guarantees your links, or fixes them for you</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Decentralized admin plus inferenceable syntax</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Flexible metadata and persistence statements</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Identifiers extensible during resolution</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Free, non-paywalled, in unlimited numbers</strong></td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Who is using ARKs?

- Libraries, data centers, archives, museums, publishers, government agencies, and vendors
- Example institutions:

  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
What are ARKs used for?

- genealogical records (8 billion FamilySearch)
- publisher content (100 million Portico)
- scientific datasets and records (22 million INIST)
- scanned books and texts 30 million Internet Archive
- bibliographic records (15 million BnF main catalog)
- museum specimens (15 million Smithsonian Institution)
- public health documents (15 million UCSF IDL)
- historical documents (21 million CDL, 5 million BnF Gallica)
- historical authors and scholars (4 million SNAC)
- fine art museum collections (483,000 Louvre)
- vocabulary terms (9,000 Periodo, YAMZ)
ARK Alliance: 850 institutions and 8.2 billion ARKs in 20 years
The ARK Alliance

Home of the ARK Alliance

arks.org

Join one of our working groups: info@arks.org

Get started with ARKs by filling out:

n2t.net/e/naan_request

Stay in touch:

- Twitter: @arks_org
- Email forum (English): groups.google.com/group/arks-forum
- Email forum (French): framalistes.org/sympa/info/arks-forum-fr