Why care about ARK identifiers?

- Because robust web links are rare – the average URL lifetime is 100 days
- ARKs can be “persistent” identifiers (PIDs), but we prefer “persistable”
- “Ten persistent myths about persistent identifiers”
  https://n2t.net/ark:/13030/c7gb1xh09

The ARK (Archival Resource Key) identifier scheme was introduced in 2001.
How cultural heritage links break

GLAMs use DAMS,
DAMS use databases,
databases generate new URLs,
frequently.

For example, widespread URL change can occur
● when the database is reloaded,
● when the software is updated,
● when the content migrates between DAMS vendors
ARK anatomy

A labelled URL with a globally unique identity inside it

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

does not identify

makes ARK actionable (the resolver)

core globally unique identity (independent of web and hostname)
N2T.net is a global “name” to “thing” resolver

Why not “ARKresolver.net” like most other PID schemes?

Because ARKs are inclusive and resolvers generalize easily.
ARK organizations

8.2 billion ARKs created by 1100+ institutions – libraries, archives, museums, publishers, data centers, educators, etc. For example,

Internet Archive
Bodleian Libraries
Berkeley Law Library
Bibliothèque Mazarine
New York Public Library
French National Archives
National Library of Austria
Library and Archives Canada

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

https://n2t.net/ark:/53355/cl010066723
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- historical documents (21 million CDL, 5 million BnF Gallica)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 (490,000 Louvre)
- vocabulary terms (30,000 Periodo, YAMZ)
History of “persistable” id schemes

- PURL (Persistent URL) – “URLs are fine if you *redirect* from purl.org”
- URN (Uniform Resource Name), DOI (Digital Object Identifier) & Handle
  ○ “URLs and domain names are bad, except for ours, and we redirect”
- Tim Berners-Lee – “cool URIs don’t break”
- ARK (Archival Resource Key) – “URLs are fine if managed well, but do tell us which of your URLs are meant for what kind of persistence”
# PID schemes – pessimist view

<table>
<thead>
<tr>
<th>Helps with major causes of broken links?</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 or service provider 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>Best practices guard against copy/paste errors</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Global resolver downtime less than 1 day per year</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Identity independence from lost domain/server name</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

[Source: arks.org](https://arks.org)
Web access – direct

1. Click
2. URL
3. Page
4. Render page
Web access – indirect

Example: archive.example.org/photo123 → photos.example.org/vault/123

A redirect is like forwarding a (request) message to a new address
## PID schemes – optimist view

<table>
<thead>
<tr>
<th>Features and costs</th>
<th>PURL</th>
<th>Handle</th>
<th>URN</th>
<th>DOI</th>
<th>ARK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decentralized resolution</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Inferenceable syntax (variants, containment)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Flexible metadata by design, including none</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Inflections (...?info) and content negotiation</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Nuanced persistence statements by design</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Path extensions during resolution (suffix passthrough)</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>
## PID schemes – ecosystem view

<table>
<thead>
<tr>
<th>Identifiers in an Internet context</th>
<th>PURL</th>
<th>Handle</th>
<th>URN</th>
<th>DOI</th>
<th>ARK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appear in Data Citation Index, HathiTrust, Wikipedia, Wikidata, Internet Archive, ORCID profiles</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Major adoption by most academic publishers outside the global South</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Free (subsidized) account and admin interface for one-off use, e.g., purl.org, zenodo.org, archive.org</td>
<td>Yes?</td>
<td>No?</td>
<td>No?</td>
<td>Yes</td>
<td>Yes?</td>
</tr>
<tr>
<td>IETF standard URI, validated by web browsers</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Replicated global resolver architecture</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

### Notes:
- **PURL**: Appears in Data Citation Index, HathiTrust, Wikipedia, Wikidata, Internet Archive, ORCID profiles
- **Handle**: Major adoption by most academic publishers outside the global South
- **URN**: Free (subsidized) account and admin interface for one-off use, e.g., purl.org, zenodo.org, archive.org
- **DOI**: IETF standard URI, validated by web browsers
- **ARK**: Replicated global resolver architecture

---

*Source: [arks.org](https://arks.org)*
Summary: ARK benefits

ARKs can serve as persistable identifiers with metadata
- found in the Data Citation Index, HathiTrust, Wikipedia, Wikidata, Internet Archive, ORCID profiles, etc.

In contrast to other id schemes, ARKs have
- no fees, no limits, no walled gardens (decentralized)
- very flexible metadata, including none
- can be assigned to anything digital, physical, or conceptual
Smithsonian ARKs: 65665

The Smithsonian Libraries & The Smithsonian Institution
- ARKs for collection metadata & multimedia objects
- Started in 2015
- By 2020 over 15 million ARKs and counting....

“ARKs are a perfect fit for our [Smithsonian] collections”
- Project size
- Cost
- Ease of implementation
- Permanence
Smithsonian ARK record and image examples

**Scientific specimens** from the National Museum of Natural History
http://n2t.net/ark:/65665/381440f27-3f74-4eb9-ac11-b4d633a7da3d

**Cultural artifacts** from the National Museum of American History
http://n2t.net/ark:/65665/ng49ca746b2-42dc-704b-e053-15f76fa0b4fa

**Sculpture** from the Freer Gallery of Art & Arthur M. Sackler Gallery
http://n2t.net/ark:/65665/ye3080ce305-a705-49cc-a70d-99aff8cb65da

**Photographs** from the National Museum of African American History and Culture
http://n2t.net/ark:/65665/fd5ad97cb86-caaf-4209-8fde-98d70f52f072

**Paintings** from the Smithsonian American Art Museum
http://n2t.net/ark:/65665/vk7a466371d-0413-451f-bd76-ca0becc46f94
Wrap up

Questions?