

# What Mark knows about ARKs

Not very much except for some things  
to think about

With most slides stolen from John  
Kunze

[https://docs.google.com/presentation/d/1xz-swo5YGykf4lwDH-vjOV3hLxesT2IUbnOi0FMD\\_74/edit#slide=id.g1c802752a1a\\_0\\_356](https://docs.google.com/presentation/d/1xz-swo5YGykf4lwDH-vjOV3hLxesT2IUbnOi0FMD_74/edit#slide=id.g1c802752a1a_0_356)

ARK Alliance



arks.org

# Core Context

- Two kinds of identifiers — content based and authority based
- Two kinds of authority — centralized and decentralized
- Any authority-based approach requires a sustained business model
- We must consider who is responsible for all the persistences



# The Five Persistences – which one?

Need to be clear if we are talking about:

- Persistence of the object – Or mechanism to handle its non-persistence
- Persistence of the identifier as a thing
- Persistence of the binding between identifier and object
- Persistence of the service to resolve from the identifier to the object
- Persistence of a service to allow for updating of the binding between identifier and object

CC-BY, @atreloar



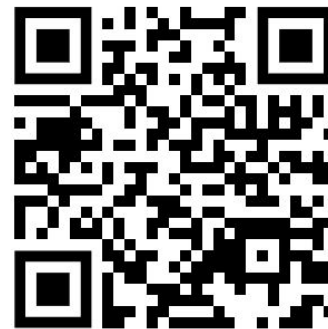
# Why care about ARK identifiers?



- 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.



# 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)



# 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.

N2T keeps identifiers persistent,  
forwarding them to the best  
known web addresses

Partners with [EZID.cdlib.org](https://ezid.cdlib.org/),  
[Identifiers.org](https://identifiers.org/), [Archive.org](https://archive.org/),  
[YAMZ.net](https://yamz.net/) metadictionary

Any kind of name – ARK, DOI,  
URN, Handle, PMID, PDB, Taxon,  
GRID, arxiv, ISSN, ...

Any kind of thing – data, web  
page, physical specimen, group,  
vocabulary term, living being, ...

**N2T is a global [ARK](#) resolver**

**Also a meta-resolver for 900+  
kinds of [compact identifiers](#)**

# ARK organizations

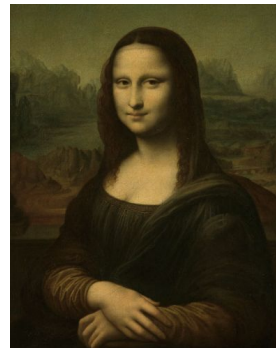
8.2 billion ARKs created by 1100+ institutions – libraries, archives, museums, publishers, data centers, educators, etc. For example,



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

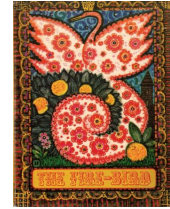
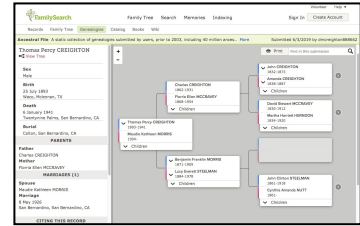
University of California Berkeley  
Smithsonian National Museum  
National Library of France  
University of Chicago  
Musée du Louvre  
Family Search  
British Library  
Google

<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](#))







# ARK metadata

- A typical ARK in a URL returns access to the thing it identifies
- To get access to its metadata, should support ARK + '?info', e.g.,

<https://n2t.net/ark:/81431/p3s39k?info> →

who: University of Pennsylvania Libraries

what: Walnut Street Theatre. Philadelphia, October 9, 1869.

when: 1869

where: ark:/81431/p3s39k (currently

<https://ezid.cdlib.org/id/ark:/81431/p3s39k>)

how: (:unav)

id created: 2017.12.06\_08:42:02

id updated: 2017.12.21\_11:16:02

persistence: (:unav)



# ARK metadata / inflections

ARK “inflection”: alter the ending to alter the request

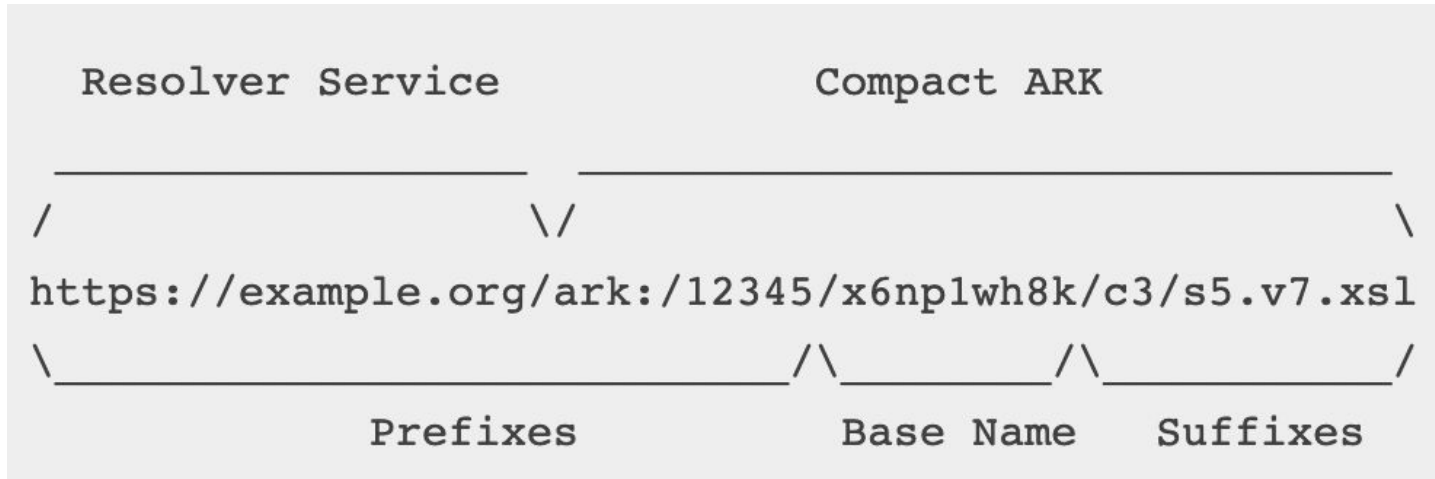
- Append ‘?info’ to request metadata (used to be ‘?’ and ‘??’)
- Earlier, append ‘?’ for metadata or ‘??’ for more (harder)
- No conflict with “content negotiation” (harder)

The inflection response format is not fixed, but it should be human- and machine-readable (JSON, YAML)

*do*  
*does*  
*done*  
*doing*  
...



# Suffixes to address image regions



# ARK + IIF example

<https://gallica.bnf.fr/iiif/ark:/12148/btv1b8449691v/f29/2131,4016,1467,948/full/0/default.jpg>



# Suffix Passthrough in Action

*Registered ARK*

<http://n2t.net/ark:/12345/x98765>



*Baseline redirection*

*Registered target URL*

<http://datazoo.example.com/carbon288>

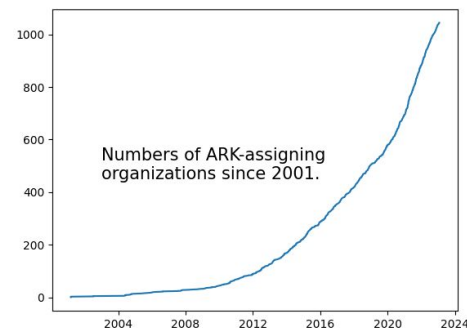
# Summary: ARK benefits

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



# What's in a NAAN record?

NAAN registry is a plain text file: [n2t.net/e/pub/naan\\_registry.txt](http://n2t.net/e/pub/naan_registry.txt)

For one NAAN, in browser enter: [n2t.net/ark:12148](http://n2t.net/ark:12148) →

```
ark:/12148:
```

```
date: 2005.07.17
```

```
name: Bibliothèque nationale de France
```

```
redirect: http://ark.bnf.fr/ark:\$id
```

NAAN as

- Resolution reference point
- Isolates assignment responsibility (autonomy, uniqueness, re-use)



# Contrasts

- Decentralized vs. centralized
- Metadata control and consistency (ARKs can use “add any kind of metadata, including no metadata”)
- Reliance on http
- Business model
- Consistency (of metadata and who plays with who)
- Allowed functionality: DOI promises wide, ARK promises specific “append extensions and query strings during resolution”



# Preservation is not binary



Persistence is not “on” or “off”. It is nuanced.

- Rapidly changing files (earth observation sensor files that grow every 6 seconds, databases that are annotated regularly)
- Similarly, this journal is preserved – why does it keep growing?
- That abstract changed – why?
- Valuable objects are often complex, volatile, human-curated clusters

# Obtain a NAAN so you can create ARKs

Fill out this form (linked, in case  
you forget, from the [arks.org](https://arks.org)  
homepage):

[n2t.net/e/naan\\_request](https://n2t.net/e/naan_request)



# ARK Alliance

## NAAN Request Form

Use this form to request a Name Assigning Authority Number (NAAN) so that you can create ARK (Archival Resource Key) identifiers. You may also use this form to request updates if you have an existing NAAN.

For a memory organization that holds content (a library, archive, data center, museum, etc.) or produces content (a laboratory, publisher, campus department, etc.), obtaining a NAAN allows it to assign ARKs. See [arks.org](https://arks.org) for more information.

When your request is verified, a unique 5-digit NAAN will be registered exclusively for the memory organization. If you have questions about this form, please use the discussion group at [groups.google.com/group/arks-forum](https://groups.google.com/group/arks-forum).

[Sign in to Google](#) to save your progress. [Learn more](#)

\* Required

I would like \*

To request a new NAAN