

# *Les identifiants pérennes*

*Archives départementales du Tarn*

4-5 octobre 2018

# Les facteurs à prendre en compte

- Des objets à identifier de types différents : image des documents, description des documents, notices d'autorité...
- Des préconisations qui peuvent paraître contradictoires : syntaxe ARK versus EAD et règles de nommage des fichiers informatiques.
- Des acteurs différents : archivistes, éditeurs de logiciels de description, éditeurs de logiciels de mise en ligne.

# Pourquoi identifier des entités ?

- Citer, référencer. Problématique ancienne.  
L'archiviste veut offrir à ses publics des moyens pour référencer une ressource repérée, consultée.
- Relier des entités. Les entités gérées sont susceptibles d'entretenir des relations sémantiques entre elles et avec d'autres entités, gérées par ailleurs, pour former, à plus ou moins long terme, un vaste réseau de données liées. On a, bien sûr, évoqué ici le web sémantique.  
L'archiviste veut s'inscrire dans la perspective du web sémantique, et donc identifier les entités qu'il gère.

# Pourquoi des identifiants pérennes ?

Si l'identifiant d'une entité change :

- Une citation ne sera plus valable
- Les liens dans le réseau sémantique seront rompus, le réseau ne sera plus fonctionnel pour les entités concernées.

La gestion de la pérennité est au cœur de notre problématique. D'ailleurs, rien n'est pérenne. Illustration sur le forum francophone ARK :

<https://framalistes.org/sympa/arc/arks-forum-fr/2018-09/msg00000.html>

# Des identifiants uniques

- Un identifiant donné ne doit correspondre qu'à une seule entité. Dit autrement, deux entités distinctes ne peuvent avoir le même identifiant.
- **Unicité locale.** Dans chaque service d'archives, on ne doit avoir qu'un article 1412 W 3.
- **Unicité globale.** Dès que l'on référence une entité sur le web, l'unicité doit être globale : le 1412 W 3 des archives du Finistère doit être identifié par une chaîne de caractères unique au monde.

# Que doit-on identifier ?

- Les documents matériels
  - Actuellement, identification locale
  - Avec la mutualisation des services, l'identification locale doit évoluer.
  - Intérêt d'une identification globale pour une citation sur le web
- Les documents numériques : ils sont variés
  - Documents nativement numériques.
  - Copies numériques de documents matériels (images, fichiers pdf). Elles peuvent être exhaustives ou partielles.
  - Images destinées à illustrer les instruments de recherche.
  - Etc.
- Les notices d'autorité.
- Les descriptions de documents d'archives. Nous allons un peu détailler ce point, le seul géré par Arkhéïa Aide au classement

# Identifier les descriptions

Les descriptions peuvent être citées, référencées. Elles doivent être identifiées. Plusieurs questions se posent.

- Question de la granularité. Description à plusieurs niveaux : quel niveau identifier ?
  - L'instrument de recherche ?
  - Chaque composant ?
  - **Tous les niveaux peuvent être cités, reliés à d'autres entités.**
  
- Question de contenu des descriptions
  - Avec le principe de description à plusieurs niveaux, un niveau peut ne porter à peu près aucune information, seulement une date, par exemple.
  - C'est l'ensemble des composants d'une branche qui porte l'information descriptive et pas un composant.
  - **Le format EAD n'est pas nécessairement pertinent pour l'identification.**

# Identifier les descriptions

- Évolution du contenu d'un instrument de recherche : un instrument de recherche n'est pas essentiellement pérenne
  - On retrouve des dossiers ou des pièces absentes lors de la rédaction initiale.
  - On accède à des informations qui permettent de rédiger des descriptions plus précises.
  - On décrit un fonds ou une collection ouvert, ce qui peut entraîner l'ajout de nouvelles notices, mais aussi, parfois, une restructuration qui peut entraîner des suppressions, des fusions, des scissions de notices...
  - Certains archivistes aiment refaire ce que leurs prédécesseurs avaient fait auparavant ;-)
  - **La description d'un même document étant susceptible de changer avec le temps, que doit-on identifier de manière pérenne ?**

# Identifier les descriptions

- Évolution des formes de l'instrument de recherche. Le format EAD a été, à l'origine et pendant plusieurs années, présenté comme pérenne. L'est-il vraiment ?
  - Arrivée de la norme ISAD(G) au milieu de la décennie 1990
  - Version 1 de la DTD EAD au tout début de la décennie 2000
  - Version 2002 l'année éponyme
  - Des déclinaisons, comme le schéma apeEAD
  - Évolution vers le schéma EAD3 qui, bien qu'officiellement adopté, n'a pas été mis en œuvre en France
  - Le SIAF préfère finalement s'orienter vers les formats du web sémantique, qui tardent aussi à s'imposer
  - **Si les différentes versions EAD ne remettaient pas en cause le principe des composants <c>, il en ira différemment avec le format RDF ! Comment assurer la pérennité de l'identification d'une description qui a changé de format et même de modèle conceptuel ?**

# Comment un identifiant peut-il être unique ?

On va emboîter plusieurs niveaux d'unicité. Exemple avec Arkhéïa.

- Le composant dans l'instrument de recherche
  - Attribut @id de chaque élément <c>. Géré automatiquement par AAC
  - Dans un même service, plusieurs id identiques dans différentes IR
- L'identifiant de l'instrument de recherche
  - Élément <eadid> géré manuellement ou par les outils d'Arkhéïa
  - Deux services d'archives distincts peuvent avoir des <eadid> identiques
- Le code du service
  - Codes forgés à l'origine selon les prescription de la DAF. Exemple FRAD029
  - À remplacer par un NAAN ARK ?
- La concaténation **code service + identifiant de l'IR + id composant** peut-être unique... si chaque élément est unique à son niveau.

# Comment un identifiant peut-il être pérenne ?

Si on ne le modifier pas !

- Le code service ne change pas... tant que l'on ne décide pas de le changer
- L'identifiant de l'instrument de recherche ne change pas... tant que les archivistes ne le modifient pas
- Pour l'identifiant d'un composant <c>, Arkhéïa Aide au classement se charge d'assurer sa pérennité

# La gestion actuelle des identifiants avec Arkhéia

- Chaque niveau de description comporte d'abord un radical **tt1** pour un titre de niveau 1, **tt2** pour un titre de niveau 2... **de** pour une description de niveau bas, **pa** pour une partie d'article. Ensuite, un numéro précédé d'un tiret. Exemple : **de-116**
- Ce numéro est calculé à partir de la base de données. Il est unique pour un IR et ne change pas :
  - Si une description est ajoutée, un nouveau numéro lui est attribué
  - Si une description est corrigée, le numéro n'est pas modifié
  - Si une description est supprimée, le numéro n'est pas réutilisé
  - Même dans le cas où l'on régénère complètement la base de données, une procédure par export et réimport balisé intègre la récupération automatique de ces numéros.

# La gestion actuelle des identifiants avec Arkhéïa

- Arkhéïa est un logiciel qui assure à un très haut niveau l'unicité et la pérennité des identifiants de ses descriptions, même si des concurrents affirment le contraire.
- Les identifiants générés par Arkhéïa sont uniques et pérennes... simplement, ils ne sont pas ARK.
- Mais, tous les logiciels de mise en ligne n'utilisent pas les identifiants d'Arkhéïa et donc font perdre ces caractéristiques.
- Il faut noter qu'à l'occasion d'un changement de système d'identifiants, la pérennité sera forcément rompue puisque l'on aura la situation avant et la situation après.

# Présentation rapide des identifiants ARK

- Bibliographie courte
  - Site de la BNF : [http://www.bnf.fr/fr/professionnels/issn\\_isbn\\_autres\\_numeros/a.ark.html](http://www.bnf.fr/fr/professionnels/issn_isbn_autres_numeros/a.ark.html)
  - Sommet international ARK :  
[http://www.bnf.fr/fr/professionnels/anx\\_journees\\_pro\\_2018/a.jp\\_180321\\_ark.html](http://www.bnf.fr/fr/professionnels/anx_journees_pro_2018/a.jp_180321_ark.html)
  - ARK pour les débutants : [http://www.bnf.fr/documents/jp\\_ark\\_2018\\_tutoriel\\_peyrard.pdf](http://www.bnf.fr/documents/jp_ark_2018_tutoriel_peyrard.pdf)
  - Transition bibliographique des catalogues vers le web des données :  
<https://www.transition-bibliographique.fr/17-10-10-numero-frbnf-identifiant-ark/>
  - On peut s'inscrire au « Forum francophone sur l'identifiant ARK » :  
<https://framalistes.org/sympa/info/arks-forum-fr>
  - Registered Name Assigning Authority Numbers (NAAN) :  
[https://www.cdlib.org/uc3/naan\\_table.html](https://www.cdlib.org/uc3/naan_table.html)

# Composition d'un identifiant ARK par l'exemple

<https://gallica.bnf.fr/ark:/12148/bpt6k107371t/f134.thumbnail>

Selon les auteurs, le sens que l'on donne au terme « identifiant ARK » varie, d'où une certaine ambiguïté.

- Pour certains auteurs, l'identifiant ARK est constitué par l'ensemble de l'URL ci-dessus.
- Pour d'autres, il s'agit de la partie centrale : **ark:/12148/bpt6k107371t/**. Il s'agit bien de l'identifiant proprement dit.

# Composition d'un identifiant ARK par l'exemple

<https://gallica.bnf.fr/ark:/12148/bpt6k107371t/f134.thumbnail>

- <https://gallica.bnf.fr/> comprend
  - [https](https://gallica.bnf.fr/) le protocole de communication (http) sécurisé (s)
  - [gallica.bnf.fr](https://gallica.bnf.fr/) « l'autorité d'adressage » (nom de domaine)  
C'est ce qui rend la ressource accessible. Cette partie n'est pas nécessairement unique, ni pérenne.
- <ark:/12148/bpt6k107371t/> est l'identifiant proprement dit, la partie pérenne. Trois parties :
  - [Ark](ark:/12148/bpt6k107371t/) : type d'identifiant
  - [12148](ark:/12148/bpt6k107371t/) : numéro d'autorité nommante (NAAN), ici NAAN de la Bnf
  - [Bpt6k107371t](ark:/12148/bpt6k107371t/) : nom ARK, c'est le nom de la ressource

<https://gallica.bnf.fr/ark:/12148/bpt6k107371t/> renvoie à un ouvrage numérisé
- [f134.thumbnail](https://gallica.bnf.fr/ark:/12148/bpt6k107371t/f134.thumbnail) est un qualificatif, optionnel. Deux parties
  - [f134](https://gallica.bnf.fr/ark:/12148/bpt6k107371t/f134.thumbnail) précise ici qu'il s'agit de la 134e image de l'ouvrage
  - [.thumbnail](https://gallica.bnf.fr/ark:/12148/bpt6k107371t/f134.thumbnail) précise que c'est une imagerie qui va être affichée. Si on le remplace par [.image](https://gallica.bnf.fr/ark:/12148/bpt6k107371t/f134.image), on affiche l'image

# ARK dans les descriptions Arkhéia ?

- Code service. Remplacer **FRAD029** par **ark:/72506**
- Identifiant de l'instrument de recherche. Respecter les règles ou recommandations ARK : identifiant opaque, éviter les voyelles et la consonne l, pas plus de 2 lettres à la suite...
- Identifiant du composant : modifier le système actuel et évitant les voyelles.

**FRAD029\_ir024\_de17** pourrait devenir, par exemple, **ark:/72506/ir024/d17**

- Problème quand même : l'identifiant d'un composant <c> est stocké dans l'attribut @id. Mais, cet attribut ne peut pas contenir le signe /. Pour cette raison, il a été dit, lors de la table ronde du *Sommet international ARK*, que la syntaxe ARK n'est pas compatible avec la syntaxe EAD.