

# AFFECTATION D'UN DOI A UNE XYLOTHEQUE

Titre élargi : « Affectation d'un identifiant numérique à une collection physique »

## Préambule

Cette note est destinée aux membres du « GT Xylothèques et collections » du GDR Bois, suite aux discussions tenues à Bordeaux le 8 novembre 2016. A cette occasion, en parlant d'identifiants numériques attribués à des objets j'ai mentionné le DOI<sup>1</sup> (*Digital Object Identifier*) car il s'agit de l'identifiant que nous allons utiliser au Cirad pour les publications et les données scientifiques, suite à une étude comparative. Le périmètre des identifiants numériques disponibles est plus large, mais l'adoption du DOI pour les objets par de multiples organismes à l'international en fait un standard partagé et reconnu ; il est même devenu l'objet de la norme ISO 26324:2012<sup>2</sup>.

## Glossaire

PID : *Persistent Identifier* (acronyme pour « identifiant numérique »)

URL : *Universal Resource Locator*, il s'agit de l'adresse d'un site web sous la forme  
« <http://www.....> »

Métadonnée : information servant à définir ou décrire une donnée (ex : date, auteur, ...)

## I. GENERALITES

### Qu'est-ce qu'un identifiant numérique ?

Un identifiant est un code alphabétique, numérique ou alphanumérique permettant d'identifier de manière **univoque** une entité de type personne physique, collectivité, publication, donnée scientifique, collection, etc. Un identifiant est **unique** : il n'est attribué qu'une seule fois. Lorsque la ressource identifiée est supprimée, son identifiant n'est pas réattribué. Un identifiant est **pérenne** : il garantit l'accès à la ressource ou à une information sur la ressource (métadonnée), quelle que soit la localisation physique de celle-ci. Il reste identique même si la ressource a été déplacée.

---

<sup>1</sup> <https://www.doi.org>

<sup>2</sup> [http://www.iso.org/iso/fr/catalogue\\_detail?csnumber=43506](http://www.iso.org/iso/fr/catalogue_detail?csnumber=43506)

L'attribution d'un identifiant unique et pérenne à une ressource permet d'en faciliter la **découverte** au travers des métadonnées, mais aussi le **partage**, la **réutilisation** et la **citation**.

## Quel identifiant numérique pour quoi(qui) ?

Ressource à identifier	Identifiant(s) approprié(s)
Un jeu de données, une ressource numérique (image, fichier PDF, ...)	<b>DOI</b> , mais aussi EPID <sup>3</sup> , Ark <sup>4</sup> , Handle, ...
Un scientifique	ORCID <sup>5</sup>
Un établissement, une institution	ISNI <sup>6</sup>
Un livre ou publication non périodique	ISBN <sup>7</sup>
Une publication périodique, une base de données	ISSN <sup>8</sup>
Un article de revue, un ouvrage, ...	DOI
<b>Un objet physique (← cas qui nous intéresse)</b>	DOI

## Présentation de l'identifiant DOI

Le système de DOI est géré par l'**International DOI Foundation** (IDF), organisation à but non lucratif qui chapeaute plusieurs *Registration Agencies*<sup>9</sup>. Ce sont ces dernières qui gèrent tout l'aspect technique du système : attribution des préfixes, enregistrement des DOI et surtout prise en charge de leur **résolution** (voir ci-contre).

Une des agences d'enregistrement spécialisée dans l'attribution de DOI à des objets de type données scientifiques, images, cartes, etc. est le **consortium DataCite**, qui propose entre autres les services suivants :

- Une interface web pour effectuer les demandes de DOI ;
- Une API pour permettre une gestion automatisée des DOI (demande, édition, suppression de métadonnées) par tout système d'information, logiciel, etc.
- Une interface de recherche de métadonnées (par auteur, institution ou autre).

L'action de « résolution » permet de mettre en correspondance, en tout temps et en tout lieu, un DOI et l'URL qui lui est associée. Concrètement cette responsabilité consiste à maintenir un annuaire répliqué sur plusieurs serveurs physiques ; cela s'apparente à la résolution de sites internet via les serveurs DNS (« qu'est ce qui fait que quand je tape google.fr dans mon navigateur je suis dirigé vers telle ou telle adresse IP ? »).

<sup>3</sup> <http://www.pidconsortium.eu>

<sup>4</sup> <http://www.cdlib.org/inside/diglib/ark/>

<sup>5</sup> <http://orcid.org>, plus d'infos : <http://guides-formadoct.ueb.eu/identifiant>



<sup>6</sup> <http://www.isni.org>

<sup>7</sup> <https://www.isbn-international.org/fr/>

<sup>8</sup> <http://www.issn.org/fr/>

<sup>9</sup> [http://www.doi.org/registration\\_agencies.html](http://www.doi.org/registration_agencies.html)

L'Inist (unité du CNRS : Institut de l'Information Scientifique et Technique) est l'institution membre de DataCite pour la France : à ce titre **l'Inist-CNRS est l'agence d'attribution de DOI pour la France.**

 **Le Cirad est un « centre de données » habilité à demander des DOI aux données scientifiques qu'il (co)produit, par l'intermédiaire de l'Inist-CNRS, via les services de DataCite<sup>10</sup>.** 

## Structure d'un DOI

Voici un exemple concret de DOI :

Version abrégée : [10.18167/archiwood/1](https://doi.org/10.18167/archiwood/1)

Version complète : <http://doi.org/10.18167/archiwood/1>

Ces 2 notations désignent la même ressource. La version complète est « autosuffisante » en termes de résolution : il suffit de la saisir dans la barre d'adresse d'un navigateur pour accéder à l'URL de la ressource. La version abrégée nécessite de passer par un service de résolution, tel celui proposé sur [doi.org](http://doi.org) (voir illustration ci-dessous).

### Resolve a DOI Name

Type or paste a DOI name, e.g., 10.1000/xyz123, into the text box below. (Be sure to enter all of the characters before and after the slash. Do not include extra characters, or sentence punctuation marks.)

*Capture d'écran du service de résolution DOI sur doi.org*

---

<sup>10</sup> Il existe d'autres agences d'attribution de DOI, notamment CrossRef qui cible les publications scientifiques (c'est grâce à CrossRef que les chercheurs ont découvert les DOI) et qui offre des services plus performants en termes de recherche de publications et de citations. Le Cirad adhère également à CrossRef pour les DOI de ses publications.

## (Dé)Composition d'un DOI

En reprenant l'exemple ci-dessus :

### **doi.org**

Service de résolution, qui concrètement assure la mise en relation entre le DOI (dans ce cas : « 10.18167/archiwood/1 ») et l'URL de la ressource identifiée (dans ce cas : <http://archiwood.cirad.fr>).

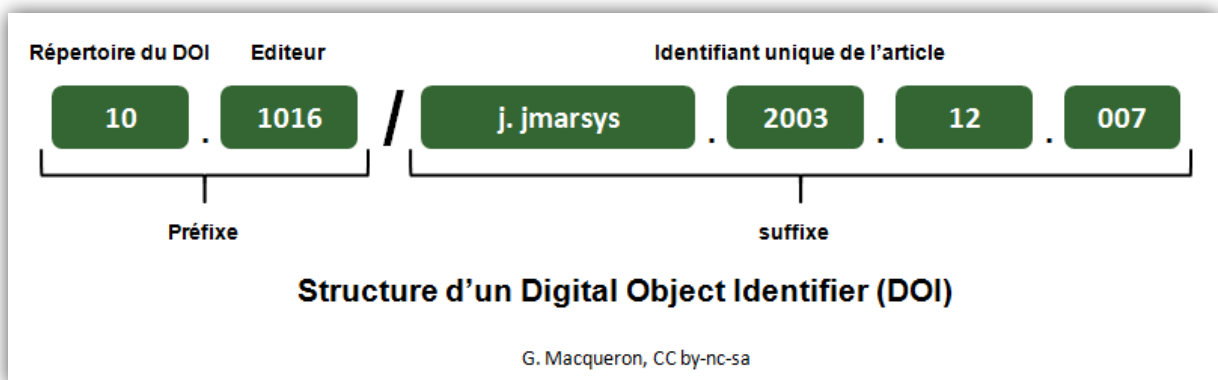
### **10.18167**

Préfixe du DOI, attribué en France par l'Inist-CNRS à chaque centre de données. Ici le préfixe attribué au Cirad.

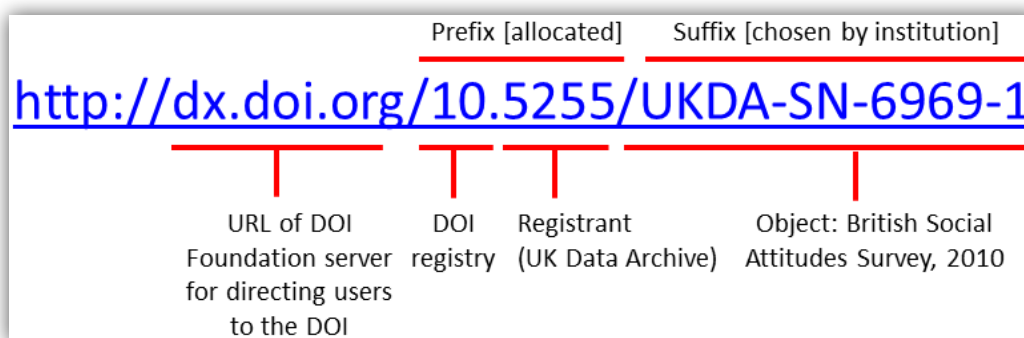
### **archiwood/1**

Suffixe du DOI, défini arbitrairement par le demandeur. Il peut être opaque ou significatif. Un identifiant opaque permet de rester indépendant de l'évolution de la ressource décrite (par exemple un nom d'équipe, de plateforme, d'UR, etc.). A l'inverse, comme dans le cas présent, rien n'interdit d'utiliser un suffixe significatif (non opaque) pour par exemple identifier les ressources issues d'une même provenance, thématique, projet ou autre.

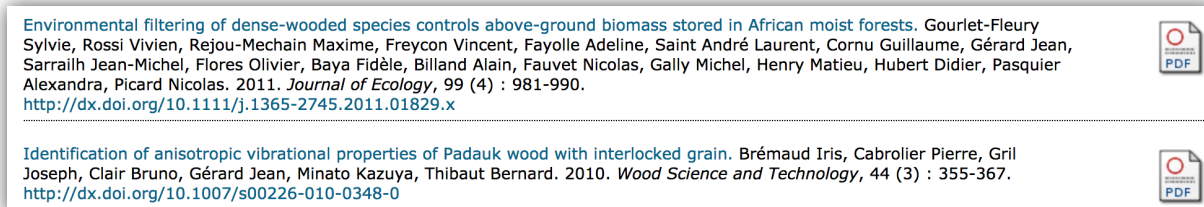
Les illustrations ci-après présentent d'autres cas réels :



*Illustration de la structure d'un DOI (récupérée sur Google Images)*



*Illustration de la structure d'un DOI (récupérée sur Google Images)*



Environmental filtering of dense-wooded species controls above-ground biomass stored in African moist forests. Gourlet-Fleury Sylvie, Rossi Vivien, Rejou-Mechain Maxime, Freycon Vincent, Fayolle Adeline, Saint André Laurent, Cornu Guillaume, Gérard Jean, Sarrailh Jean-Michel, Flores Olivier, Baya Fidèle, Billand Alain, Fauvet Nicolas, Gally Michel, Henry Matieu, Hubert Didier, Pasquier Alexandra, Picard Nicolas. 2011. *Journal of Ecology*, 99 (4) : 981-990.  
<http://dx.doi.org/10.1111/j.1365-2745.2011.01829.x>

Identification of anisotropic vibrational properties of Padauk wood with interlocked grain. Brémaud Iris, Cabrolier Pierre, Gril Joseph, Clair Bruno, Gérard Jean, Minato Kazuya, Thibaut Bernard. 2010. *Wood Science and Technology*, 44 (3) : 355-367.  
<http://dx.doi.org/10.1007/s00226-010-0348-0>

*Le DOI en action, tel que vous l'avez sûrement déjà observé au bas d'articles, publications, posters, etc. (source : Agritrop Cirad)*

## Obligations

Les seules exigences préalables à l'obtention d'un DOI sont :

- Garantir un **accès pérenne** (au minimum 10 ans) aux métadonnées, présentées sur une page descriptive internet via une URL idéalement durable<sup>11</sup>.
- Renseigner une **fiche de métadonnées** (au format XML, en suivant le schéma fourni par DataCite<sup>12</sup>) décrivant la ressource à identifier. Parmi les 18 propriétés à remplir (auteur(s) et contributeur(s), description, année de publication, sujet, date de création, format(s) de données, licence de réutilisation, etc), **seules 5 propriétés sont obligatoires**, 7 sont recommandées et 6 optionnelles.

Les points suivants ne sont pas des obligations mais font partie des recommandations et bonnes pratiques :

- Respecter les règles de propriété intellectuelle vis à vis des données concernées avant de les publier ;
- Publier les métadonnées mentionnées ci-dessus.

<sup>11</sup> « idéalement » car l'URL est modifiable à posteriori ; cependant il ne s'agit pas d'une pratique recommandée.

<sup>12</sup> Disponible à l'adresse suivante : <http://schema.datacite.org>

## Points clés

- Un DOI est un identifiant unique ;
- Un DOI pointe vers une URL, qui peut être la ressource elle-même ou bien une page descriptive de la ressource ;
- La ressource (les données) **n'est pas obligatoirement mise à disposition du public**, ses modalités de partage et réutilisation sont définies par une licence.

## II. APPLICATION

### Quel intérêt pour une xylothèque ?

D'après nous, la xylothèque du Cirad constitue un patrimoine scientifique unique. Son potentiel de valorisation scientifique est important ; elle est déjà mobilisée et constitue un élément essentiel à l'origine des bases de données de connaissance des bois dont le Cirad, via l'UR BioWooEB, tire son expertise dans ce domaine. Elle reste cependant méconnue et peu citée pour diverses raisons, la principale provenant de sa nature même : il s'agit d'une collection entièrement analogique, donc confinée physiquement au laboratoire de Montpellier.

Afin d'améliorer cet état de fait nous envisageons plusieurs actions, en mettant à profit des outils, technologies et pratiques proposés aujourd'hui dans le cadre de l'*open science* : publication des métadonnées sur internet, publication d'un *datapaper* décrivant le processus d'obtention des échantillons et la gestion de leur cycle de vie, attribution d'un identifiant numérique de type DOI afin de faciliter sa citabilité dans des articles scientifiques, etc.

L'intérêt d'un DOI réside principalement dans les points suivants :

- Il assure l'**identification univoque** de la ressource ;
- Il garantit un **accès pérenne** à la ressource ;
- Il facilite sa **citation** grâce aux métadonnées associées.

« A l'image d'une citation d'un article scientifique, il procure une visibilité au jeu de données et à son auteur. » (Ifremer)

Il en va ainsi pour une xylothèque ; j'ajouterais que le DOI peut **inciter à la citation** d'un tel objet physique, qui de prime abord n'aurait pas été un « réflexe » de la part d'un auteur, certainement davantage habitué à citer des publications scientifiques. Un cas pratique envisageable serait de voir la xylothèque citée en tant que référentiel scientifique utilisé dans le cadre d'une étude, dont découlerait la production d'un jeu de données, qui ferait lui-même éventuellement l'objet d'un *datapaper* et/ou d'une publication.

Cependant je n'ai à l'heure actuelle trouvé aucune mise en œuvre concrète d'affectation d'identifiant numérique sur une collection de bois, contrairement aux herbiers, par exemple pour lesquels la pratique est déjà répandue :

<http://data.canadensys.net/ipt/resource.do?r=qfa-vascular-specimens>  
<http://www.gbif.org/dataset/cb9beff3-a185-486f-975a-732251444158>

## Mise en œuvre

### *Information préalable*

Le Cirad est un centre de données agréé par DataCite et pouvant solliciter/attribuer des DOI. L'enjeu d'un tel dispositif est de bénéficier d'un préfixe DOI (le fameux « 10.xxxx ») propre au Cirad, qui peut être valorisant en termes de notoriété scientifique. La plupart des entrepôts de données sur internet (GBIF, FigShare, Zenodo, Dryad, etc.) sont également habilités à délivrer des DOI.

### *Processus*

Dans le cas de l'attribution d'un DOI à la xylothèque Cirad, les étapes à suivre sont les suivantes :

1. Remplir la fiche de métadonnées au format XML ;
2. Créer une page web décrivant la xylothèque, ayant pour URL par exemple xylotheque.cirad.fr (indépendante de « BioWooEB » car le nom d'une UR est susceptible de changer durant les 10 années à venir) ;
3. Demander un DOI auprès du service concerné au Cirad ;
4. Diffuser ce DOI à tous les endroits indiqués : page web de la xylothèque mais aussi de l'UR, supports de communication, etc. Inciter à sa citation.

### *Rédaction des métadonnées*

Plusieurs champs de métadonnées nécessitent réflexion et concertation, car s'ils s'appliquent de manière plutôt évidente à, par exemple, un article scientifique, il en est autrement pour un objet physique à fortiori quand il s'agit d'une collection constituée durant plusieurs décennies. Qui en sont tous les auteurs, contributeurs ? Les financeurs ? Il est utopique d'être exhaustif concernant ces données-là. A l'inverse de toutes les métadonnées descriptives, plutôt aisées à renseigner.

Un exemple de fiche de métadonnées est proposé dans la partie Annexes.

## Perspectives

L'attribution d'un DOI à la xylothèque du Cirad constituera un 1<sup>er</sup> jalon dans une démarche de valorisation et reconnaissance scientifique de notre collection. Parmi les perspectives envisagées, outre des actions de communication en 2017 à l'occasion des 80 ans de la xylothèque, apparaissent naturellement sa numérisation, la publication de la base de données d'indexation sur internet (selon l'exemple de ce que propose le MNHN), pourquoi pas la mutualisation avec d'autres collections ?

## Bibliographie

Site officiel du DOI : <http://www.doi.org>

*Factsheet DOI* : <https://www.doi.org/factsheets/DOIKeyFacts.html>

INRA datapartage : <https://www6.inra.fr/datapartage/Generer/DOI>

Inist-CNRS : <http://www.inist.fr/?Attribution-de-DOI>

Schéma de métadonnées DataCite : <http://schema.datacite.org>

Guide PDF sur les identifiants numériques édité par le GBIF : [A beginner's guide to persistent identifiers](#)



## Annexes

```
<?xml version="1.0" encoding="UTF-8"?>
<resource xsi:schemaLocation="http://datacite.org/schema/kernel-3 http://schema.dat
  <identifiant identifierType="DOI">10.18167/xy12345</identifiant>
  <creators>
    <creator>
      <creatorName>Détienne, Pierre</creatorName>
      <givenName>Pierre</givenName>
      <familyName>Détienne</familyName>
      <affiliation>CIRAD</affiliation>
    </creator>
    <creator>
      <creatorName>Langbour, Patrick</creatorName>
      <givenName>Patrick</givenName>
      <familyName>Langbour</familyName>
      <affiliation>CIRAD</affiliation>
      <affiliation>BioWooEB Research Unit</affiliation>
    </creator>
  </creators>
  <titles>
    <title xml:lang="en-us">Wood collection</title>
    <title xml:lang="fr-fr">Xtlothèque</title>
  </titles>
  <publisher>CIRAD</publisher>
  <publicationYear>2017</publicationYear>
  <subjects>
    <subject xml:lang="en-us">Biodiversity, wood, wood anatomy, tropical woods</subject>
    <subject xml:lang="fr-fr">Biodiversité, bois, anatomie du bois, bois tropic</subject>
    <subject xml:lang="en-us" schemeURI="http://id.loc.gov/authorities/subjects"></subject>
    <subject xml:lang="en-us" schemeURI="http://id.loc.gov/authorities/subjects"></subject>
    <subject xml:lang="fr-fr" schemeURI="http://www.fao.org/docrep/003/u1808f/u"></subject>
    <subject xml:lang="fr-fr" schemeURI="http://www.fao.org/docrep/003/u1808f/u"></subject>
    <subject xml:lang="fr-fr" schemeURI="http://www.fao.org/docrep/003/u1808f/u"></subject>
    <subject xml:lang="en-us" schemeURI="http://www.fao.org/docrep/003/u1808f/u"></subject>
    <subject xml:lang="en-us" schemeURI="http://www.fao.org/docrep/003/u1808f/u"></subject>
    <subject xml:lang="en-us" schemeURI="http://www.fao.org/docrep/003/u1808f/u"></subject>
  </subjects>
  <contributors>
    <contributor contributorType="HostingInstitution">
```

Fichier XML de métadonnées selon le schéma DataCite pour la xylothèque (non finalisé)